GROUP ACTIONS, LINEAR TRANSFORMATIONS, AND FLOWS

MOODY T. CHU*

Abstract. It is known that there is a close relationship between matrix groups and linear transformations. The purpose of this exposition is to explore that relationship and to bridge it to the area of applied mathematics. Some known connections between discrete algorithms and differential systems will be used to motivate a larger framework that allows embracing more general matrix groups. Different types of group actions will also be considered. The inherited topological structure of the Lie groups makes it possible to design various flows to approximate or to effect desired canonical forms of linear transformations. While the group action on a fixed matrix usually preserves a certain internal properties or structures along the orbit, the action alone often is not sufficient to direct the orbit to the desired canonical form. Various means to further control these actions will be introduced. These controlled group actions on linear transformations often can be characterized by a certain dynamical systems on a certain matrix manifolds. Wide range of applications starting from eigenvalue computation to structured low rank approximation, and to some inverse problems are demonstrated. A number of open problems are identified.

Key words. matrix group, group action, linear transformation, canonical form, least squares approximation, dynamical system.

AMS subject classifications. 65F30, 65H20, 65K10, 65L05, 20H15, 22E70, 14L35, 34C40, 37N30, 37N40

1. Introduction.

Any linear transformation over finite dimensional spaces can be represented by a matrix together with matrix-vector multiplication. Linear transformation is one of the simplest yet most profound ways to describe the relationship between two vector spaces. To further deduce the essence of a linear transformation, it is often desired to identify a matrix by its *canonical form*. For years researchers have taken great effort to describe, analyze, and modify algorithms to achieve this goal. Thus far, this identification process is mostly done via iterative procedures of which the success is evidenced by the many available discrete methods.

In his book, *Geometric Methods in the Theory of Ordinary Differential Equations*, V. I. Arnold raised a related and interesting question [1, Page 239], "What is *the simplest form* to which a family of matrices depending smoothly on the parameters can be reduced by *a change of coordinates* depending smoothly on the parameters?" In this quotation, we italicize the two phrases to further accentuate the important inquiries that

- What is qualified as the simplest form?
- What kind of continuous change can be employed?

Recently it has been observed that the use of differential equations to issues in computation can afford fundamental insights into the structure and behavior of existing discrete methods and, sometimes, can suggest new and improved numerical methods. In certain cases, there are remarkable connections between smooth flows and discrete numerical algorithms. In other cases, the flow approach seems advantageous in tackling very difficult problems.

What is being talked about thus far can be categorically characterized as a realization process. Realization process, in a sense, means any deducible procedure that we use to rationalize and solve problems. The simplest form therefore refers to the agility to think and draw conclusions out of the form. In mathematics, the steps taken in a realization process often appears in the form of an iterative procedure or a differential equation.

All together, it appears that both discrete approach and flow approach are actually following the *orbit* of a certain matrix groups action on the underlying matrix . The question to ask, therefore, is to what canonical form a matrix or a family of matrices can be linked by the orbit

*Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205. (chu@math.ncsu.edu)
This research was supported in part by the NSF under grants DMS-0073056 and CCR-0204157, and in part by the INDAM under the Visiting Professorship Program when the author visited the Department of Mathematics at the University of Bari.

of a group action. The choice of the group, the definition of the action, and the targets intended to reach will effect the various transitions that have been studied in the literature. This paper will attempt to expound the various aspects of the recent development and application in this direction. Some possible new research topics will also be proposed. The earliest discussion along these lines seem to be in [21].

A realization process usually consists of three components: First of all, we will have two abstract problems of which one is a make-up and is easy to solve while the other is the real problem and is difficult. The basic idea is to realize the solution of the difficult problem through a bridge or a path that starts from the solution of the easier problem. Once the plan is set, we then need a numerical method to move along the path to obtain the desired solution.

To build the bridge, sometimes we are given specific construction guidelines. For example, the bridge could be constructed by monitoring the values of certain specified function. In this case, the path is guaranteed to work. The projected gradient method is a typical representation of this approach. Sometimes the guidelines are not so specific and hence we are kind of constructing the bridge instinctively. We can only hope that the bridge will connect to the other end. It takes efforts to prove that the bridge actually makes the desired connection. A typical continuation method in this class is the homotopy method. Another situation is that the bridge seems to exist unnaturally. But the matter of fact is that usually a much deeper mathematics or physics is involved. When we begin to understand the theory, we are often amazed to see that these seemingly unnatural bridges exist by themselves naturally.

We shall be more specific about how a bridge could be built. For now, be it sufficient to say that a bridge, if exists, usually is characterized by an ordinary differential equation and that the discretization of a bridge, or a numerical method in travelling along a bridge, usually produces an iterative scheme. Known as the geometric integration, research of numerical integrators that respect the underlying geometric structure has been attracting considerable attention recently, but will be out of the scope of this survey.

2. A Case Study. Before we move into more details, we use two classical examples and their connections to demonstrate the main points made above. We also briefly describe their generalization with the hope that this discussion would serve as the steppingstone and shed light on the thrust throughout this paper. More details can be found in [11].

2.1. Discreteness versus Continuousness. Consider first the symmetric eigenvalue problem that has been of critical importance in many applications. Given a symmetric matrix A_0, the problem is to find all scalars λ so that the equation

$$A_0\mathbf{x} = \lambda\mathbf{x} \tag{2.1}$$

has a non-trivial solution $\mathbf{x}$. Currently, one of the most crucial techniques for eigenvalue computation is by iteration. Recall the fact that any matrix A enjoys the QR decomposition:

$$A = QR,$$

where Q is orthogonal and R is upper triangular. The basic QR algorithm defines a sequence of matrices $\{A_k\}$ via the recursion relationship [26]:

$$\left\{\begin{array}{ccc} A_k & = & Q_kR_k, \\ A_{k+1} & = & R_kQ_k. \end{array}\right. \tag{2.2}$$

Because $A_{k+1} = Q_k^T A_k Q_k$, every matrix A_k in the sequence has the same eigenvalues as A_0. More importantly, it can be proved that the sequence $\{A_k\}$ converges to a diagonal matrix and, hence, eigenvalues are found.

In contrast, let a symmetric matrix X be decomposed as

$$X = X^o + X^- + X^+,$$

where X^o, X^-, and X^+ denote the diagonal, the strictly lower triangular, and the strictly upper triangular parts of X, respectively. Define

$$\Pi_0(X) := X^- - X^{-\top}. \tag{2.3}$$

The Toda flow is an initial value problem defined by [22, 34]:

$$\left\{ \begin{array}{rcl} \frac{dX(t)}{dt} & = & [X(t), \Pi_0(X(t))], \\ X(0) & = & X_0, \end{array} \right. \tag{2.4}$$

where $[M, N] := MN - NM$ denotes the commutator bracket. It is known that when the solution of the Toda flow (2.4) is sampled at integer times, the sequence $\{X(k)\}$ gives rise to the same iterates as does the QR algorithm (2.2) applied to the matrix $A_0 = exp(X_0)$. Furthermore, it can be proved that the evolution of $X(t)$ starts from X_0, converges to the limit point of Toda flow, which is a diagonal matrix, and that $X(t)$ maintains the same spectrum as that of X_0 for all t. For tridiagonal matrices, the equation (2.4) known as the Toda lattice is a Hamiltonian system constructed from some physics settings [34]. A certain physical quantities are kept at constant, that is, the Toda flow is a *completely integrable* system [23].

In both approaches, the eigenvalue computation is cast as a realization process that starts from the “easy” matrix whose eigenvalues are to be found and ends at the “difficult” matrix which is the diagonal matrix carrying the eigenvalue information. The bridges that connect the original matrix to the final diagonal matrix are the QR algorithm in the discrete process and the Toda flow in the continuous process, respectively. Maintaining isospectrality everywhere along the bridges is the most important property inhered in both bridges. It is remarkable that the QR algorithm and the Toda flow are so closely related.

Consider next the least squares matrix approximation problem which also plays a significant role in disciplines of various areas. Given a symmetric matrix N, the problem is find a least squares approximation of N while maintaining a prescribed set of eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$. The problem can be formulated as an equality constrained minimization problem as follows:

$$\begin{array}{rl} \text{Minimize} & F(Q) := \frac{1}{2}||Q^\top \Lambda Q - N||^2 \\ \text{Subject to} & Q^T Q = I. \end{array}$$

While iterative optimization techniques such as the augmented Lagrangian methods or the sequential quadratic programming methods could readily be applied to solve this problem, none of these techniques respects the matrix structure. For example, the constraint carries lots of algebraic redundancies and actually defines only a $n(n-1)/2$-dimensional manifold.

In contrast, the projected gradient of F can easily be calculated and a projected gradient flow can be defined via the double bracket differential equation [6, 16]:

$$\left\{ \begin{array}{rcl} \frac{dX(t)}{dt} & = & [X(t), [X(t), N]], \\ X(0) & = & \Lambda, \end{array} \right. \tag{2.5}$$

The solution $X(t) := Q(t)^\top \Lambda Q(t)$ moves in a descent direction to reduce $\|X(t) - N\|^2$ as t goes to infinity. It can be proved that the optimal solution X can be fully characterized in terms of the spectral decomposition of N and is unique.

In this setting, the evolution starts from the "easy" matrix which is the diagonal matrix of the prescribed eigenvalues and converges to the limit point which solves the least squares problem. The flow is built on the basis of systematically reducing the difference between the current position and the target position. This differential system is a gradient flow.

At first glance, the eigenvalue computation and the least squares matrix approximation seem to be two unrelated problems. However, when X is tridiagonal and

$$N = \mathrm{diag}\{n, \ldots, 2, 1\},$$

it can easily be verified that [5]

$$[X, N] = \Pi_0(X).$$

In other words, the gradient flow (2.5) in fact is also a Hamiltonian flow (2.4).

2.2. Generalization. Both differential systems (2.4) and (2.5) are special cases of the more general Lax dynamical system:

$$\begin{cases} \frac{dX(t)}{dt} & := & [X(t), k_1(X(t))], \\ X(0) & := & X_0, \end{cases} \tag{2.6}$$

where $k_1 : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ is some selected matrix-valued function to be specified later. Associated with (2.6), we define two parameter dynamical systems:

$$\begin{cases} \frac{dg_1(t)}{dt} & := & g_1(t)k_1(X(t)), \\ g_1(0) & := & I, \end{cases} \tag{2.7}$$

and

$$\begin{cases} \frac{dg_2(t)}{dt} & := & k_2(X(t))g_2(t), \\ g_2(0) & := & I, \end{cases} \tag{2.8}$$

with the property that

$$k_1(X) + k_2(X) = X. \tag{2.9}$$

The following theorem has been established earlier [19].

THEOREM 2.1. *For any t within the interval of existence, the solutions $X(t)$, $g_1(t)$, and $g_2(t)$ of the systems (2.6), (2.7), and (2.8), respectively, are related to each other by the following three properties:*

1. *(Similarity Property)*

$$X(t) = g_1(t)^{-1}X_0g_1(t) = g_2(t)X_0g_2(t)^{-1}. \tag{2.10}$$

2. *(Decomposition Property)*

$$\exp(tX_0) = g_1(t)g_2(t). \tag{2.11}$$

3. *(Reversal Property)*

$$\exp(tX(t)) = g_2(t)g_1(t). \tag{2.12}$$

The consequences of Theorem 2.1 are quite broad and significant. The similarity property, for example, immediately implies that the solution $X(t)$ to (2.6) is isospectral. The theorem also suggests an abstraction of the QR algorithm in the following sense: It is known that the (Lie) group $Gl(n)$ of real-valued $n \times n$ nonsingular matrices can be decomposed as the product of two (Lie) subgroups in the neighborhood of I if and only if the corresponding tangent space (Lie algebra) $gl(n)$ of real-valued $n \times n$ matrices can be decomposed the sum of two (Lie) subalgebras. By the decomposition property and the reversal property, the Lie structure apparently is not needed in our setting. It suffices to consider a factorization of a *one-parameter semigroup* in the neighborhood of I as the product of two nonsingular matrices, that is, the decomposition as is indicated in (2.11). Then, correspondingly there is a subspace decomposition of $gl(n)$ as is indicated in (2.9). Some of the jargons will be clarified in the next section. Our point at present is to bring forth the intriguing structure that is underneath the transformations.

Analogous to the QR decomposition, the product $g_1(t)g_2(t)$ will be called the *abstract* g_1g_2 *decomposition* of $\exp(X_0 t)$. By setting $t = 1$, we see that

$$\left\{ \begin{array}{lcl} \exp(X(0)) & = & g_1(1)g_2(1), \\ \exp(X(1)) & = & g_2(1)g_1(1). \end{array} \right. \tag{2.13}$$

Since the dynamical system for $X(t)$ is autonomous, it follows that the phenomenon characterized by (2.13) will occur at every feasible integer time. Corresponding to the abstract g_1g_2 decomposition, the above iterative process for all feasible integers will be called the *abstract* g_1g_2 *algorithm*. It is thus seen that the curious iteration in the QR algorithm is completely generalized and abstracted via the mere subspace splitting (2.9).

3. General Framework. What is being manifested in the preceding section is but a special case of a more general framework for the realization process. We have already seen a resemblance between the Toda flow $X(t)$ defined by (2.4) and the isospectral flow $X(t)$ defined by (2.6). The former can be considered as a curve on the "orbit" $Q(t)^\top X_0 Q(t)$ under the orthogonal similarity transformation by all $n \times n$ orthogonal matrices Q which form a group while the latter can be considered as a curve on the orbit $g_1(t)^{-1} X_0 g_1(t)$ under the similarity transformation by whatever matrices $g_1(t)$ that can be defined but not necessarily form a group. It is natural to exploit what other transformations and sets from which a transformation is taking place can be characterized in general.

3.1. Matrix Group and Actions. We begin the generalization with the notion of matrix groups. We shall eventually drop the requirement of group structure in the later part of this discussion. Many manifolds arisen from applications are not groups themselves, but still can be "parameterized" by groups.

Recall that a subset of nonsingular matrices (over any field) which are closed under matrix multiplication and inversion is called a *matrix group* [2, 3, 20]. A smooth manifold which is also a group where the multiplication and the inversion are smooth maps is called a *Lie group*. It has been shown that every matrix group is in fact a Lie group [28]. The reason that Lie groups are interesting is because this particular entity inherits both algebraic and geometric structures. The most remarkable feature of a Lie group is that the structure is the same in the neighborhood of each of its elements. We shall elaborate more on this point later.

Matrix groups are central in many parts of mathematics and applications. Lots of realization processes used in numerical linear algebra are the results of actions of matrix groups. For the convenience of later references, we tabulate in Table 3.1 some classical matrix groups and their subgroups over $\mathbb{R}$.

Any of these matrix groups could be used to transform matrices. To be more precise, we define the action of a group according to the following rule.

Group	Subgroup	Notation	Characteristics
General linear		$\mathcal{G}l(n)$	$\{A \in \mathbb{R}^{n\times n} \mid \det(A) \neq 0\}$
	Special linear	$\mathcal{S}l(n)$	$\{A \in \mathcal{G}l(n) \mid \det(A) = 1\}$
Upper triangular		$\mathcal{U}(n)$	$\{A \in \mathcal{G}l(n) \mid A \text{ is upper triangular}\}$
	Unipotent	$\mathcal{U}nip(n)$	$\{A \in \mathcal{U}(n) \mid a_{ii} = 1 \text{ for all } i\}$
Orthogonal		$\mathcal{O}(n)$	$\{Q \in \mathcal{G}l(n) \mid Q^\top Q = I\}$
Generalized orthogonal		$\mathcal{O}_S(n)$	$\{Q \in \mathcal{G}l(n) \mid Q^\top S Q = S\}$, S is a fixed symmetric matrix
	Symplectic	$\mathcal{S}p(2n)$	$\mathcal{O}_J(2n), \quad J := \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$
	Lorentz	$\mathcal{L}or(n,k)$	$\mathcal{O}_L(n+k)$, $L := \text{diag}\{\underbrace{1,\dots,1}_{n}, \underbrace{-1,\dots-1}_{k}\}$
Affine		$\mathcal{A}ff(n)$	$\left\{\begin{bmatrix} A & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \mid A \in \mathcal{G}l(n), \mathbf{t} \in \mathbb{R}^n\right\}$
	Translation	$\mathcal{T}rans(n)$	$\left\{\begin{bmatrix} I & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \mid \mathbf{t} \in \mathbb{R}^n\right\}$
	Isometry	$\mathcal{I}som(n)$	$\left\{\begin{bmatrix} Q & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \mid Q \in \mathcal{O}(n), \mathbf{t} \in \mathbb{R}^n\right\}$
Center of G		$Z(G)$	$\{z \in G \mid zg = gz, \text{ for every } g \in G\}$, G is a given group
Product of G_1 and G_2		$G_1 \times G_2$	$\{(g_1,g_2) \mid g_1 \in G_1, g_2 \in G_2\}$, $(g_1,g_2) * (h_1,h_2) := (g_1h_1, g_2h_2)$, G_1 and G_2 are given groups
Quotient		G/N	$\{Ng \mid g \in G\}$, N is a fixed normal subgroup of G
	Hessenberg	$\mathcal{H}ess(n)$	$\mathcal{U}nip(n)/\mathcal{Z}_n$

Table 3.1
Examples of classical matrix groups over $\mathbb{R}$.

Definition 3.1. *Given a group G and a set $\mathbb{V}$, a function $\mu : G \times \mathbb{V} \longrightarrow \mathbb{V}$ is said to be a group action of G on $\mathbb{V}$ if and only if*

1. $\mu(gh, \mathbf{x}) = \mu(g, \mu(h, \mathbf{x}))$ for all $g, h \in G$ and $\mathbf{x} \in \mathbb{V}$.

2. $\mu(e, \mathbf{x}) = \mathbf{x}$, *if e is the identity element in G.*

In our applications, the set $\mathbb{V}$ is used to specify the principal characteristics of the matrices that we are interested. The action, in turn, is used to specify the transformations that are allowed. Given an arbitrary $\mathbf{x} \in \mathbb{V}$, two important sets are associated with a group action. These are the *stabilizer* of $\mathbf{x}$,

$$Stab_G(\mathbf{x}) := \{g \in G | \mu(g, \mathbf{x}) = \mathbf{x}\}, \tag{3.1}$$

which is a subgroup of G, and the *orbit* of $\mathbf{x}$,

$$Orb_G(\mathbf{x}) := \{\mu(g, \mathbf{x}) | g \in G\}. \tag{3.2}$$

The orbit of a matrix under a certain group action is particularly relevant to the central theme of this paper. To demonstrate how some of the conventional transformations used in practice can be cast as group actions, we list in Table 3.2 some of the useful actions by matrix groups.

Set $\mathbb{V}$	Group G	Action $\mu(g, A)$	Application
$\mathbb{R}^{n\times n}$	Any subgroup	$g^{-1}Ag$	conjugation
$\mathbb{R}^{n\times n}$	$\mathcal{O}(n)$	$g^\top Ag$	orthogonal similarity
$\underbrace{\mathbb{R}^{n\times n} \times \ldots \times \mathbb{R}^{n\times n}}_{k}$	Any subgroup	$(g^{-1}A_1g, \ldots, g^{-1}A_kg)$	simultaneous reduction
$\mathbb{S}(n) \times \mathbb{S}_{PD}(n)$	Any subgroup	$(g^\top Ag, g^\top Bg)$	symm. positive definite pencil reduction
$\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$	$\mathcal{O}(n) \times \mathcal{O}(n)$	$(g_1^\top Ag_2, g_1^\top Bg_2)$	QZ decomposition
$\mathbb{R}^{m\times n}$	$\mathcal{O}(m) \times \mathcal{O}(n)$	$g_1^\top Ag_2$	singular value decomp.
$\mathbb{R}^{m\times n} \times \mathbb{R}^{p\times n}$	$\mathcal{O}(m) \times \mathcal{O}(p) \times \mathcal{G}l(n)$	$(g_1^\top Ag_3, g_2^\top Bg_3)$	generalized singular value decomp.

TABLE 3.2
Examples of group actions and their applications.

Groups listed in Table 3.1 and actions listed in Table 3.2 represent only a small collection of wild possibilities. They already indicate a wide open area for further research because, except for the orthogonal group $\mathcal{O}(n)$, not many other groups nor actions have been employed in practice. In numerical analysis, it is customary to use actions of the orthogonal group to perform the change of coordinates for the sake of cost efficiency and numerical stability. It becomes interesting to ask, for example, what conclusion could be drawn if actions of the isometry group $\mathcal{I}som(n)$ are used instead. The isometry group is appealing for at least three reasons: that the inverse of an isometry matrix is easy to compute since

$$\begin{bmatrix} Q & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} Q^\top & -Q^\top\mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix};$$

that the numerical stability by isometric transformation is guaranteed; and that the isometry group contains the orthogonal group as its subgroup and hence offers more flexibilities. To our knowledge, however, we have seen the usage of such an isometric transformation only in the discipline of physics, but rarely in numerical computation.

Some more exotic actions include, for example, the orthogonal conjugation plus shift by the product group of the orthogonal group $\mathcal{O}(n)$ and the additive group $\mathbb{R}_+$ of real numbers,

$$\mu((Q,s),A) := Q^\top AQ + sI, \quad Q \in \mathcal{O}(n), s \in \mathbb{R}_+;$$

the orthogonal group with scaling by the product group of $\mathcal{O}(n)$ and the multiplicative group $\mathbb{R}_\times$ of nonzero real numbers,

$$\mu((Q,s),A) := sQ^\top AQ, \quad Q \in \mathcal{O}(n), s \in \mathbb{R}_\times;$$

or even the orthogonal conjugation with multiple scalings,

$$\mu((Q,\mathbf{s},\mathbf{t}),A) := \mathrm{diag}\{\mathbf{s}\}Q^\top AQ\mathrm{diag}\{\mathbf{t}\}, \quad Q \in \mathcal{O}(n), \mathbf{s},\mathbf{t} \in \mathbb{R}^n_\times.$$

It is important to note that, by using the group inverse, the information of A can be retrieved at any point along any of the orbits defined above. The issue is how to define curves on these orbits so that the limit points would be useful for us to retrieve information about A. We do not think preposterous group actions by arbitrary groups are all useful. We do want to point out, however, that there are many unanswered questions that deserve further study.

3.2. Tangent Space and Projection. Given a group G and its action μ on a set $\mathbb{V}$, the associated orbit $Orb_G(\mathbf{x})$ characterizes the rule by which $\mathbf{x}$ is to be changed in $\mathbb{V}$. Depending on the group G, an orbit is often a high dimensional manifold that is too "wild" to be readily traced for finding the "simplest form" of $\mathbf{x}$. Therefore, depending on the applications, it is desired to build a path, bridge, curve, or differential equation on the orbit so as to connect $\mathbf{x}$ to its simplest form. We have yet to define what is meant by the canonical form, but we first understand the notion of a vector field on a manifold.

A differential equation on the orbit $Orb_G(\mathbf{x})$ is equivalent to a differential equation on the group G. We have seen this relationship in Section 2. The Lax dynamics (2.6) describes a bridge for the isospectral curve $X(t)$ on the orbit under orthogonal conjugation. Correspondingly, the parameter dynamics (2.7) and (2.8) characterize the flows for $g_1(t)$ in $\mathcal{O}(n)$ and $g_2(t)$ in $\mathcal{U}(n)$, respectively.

To stay in either the orbit or the group, the vector field of the dynamical system must be distributed over the tangent space of the corresponding manifold. Most of the tangent spaces for matrix groups can be calculated explicitly. If some kind of objective function has been used to control the connecting bridge, the gradient of the objective function should be projected to the tangent space.

Again, for completeness, we briefly review the notion of tangent space for a matrix group below. Given a matrix group G which is a subgroup in $\mathcal{G}l(n)$, the *tangent space* to G at an element $A \in G$ can be defined as the set

$$\mathcal{T}_A G := \{\gamma'(0) | \gamma \text{ is a differentiable curve in } G \text{ with } \gamma(0) = A\}. \tag{3.3}$$

The tangent space of a matrix group G at the identity I is so critical that it deserves a special notation. We shall state the following facts without the proof.

THEOREM 3.2. *Let $\mathfrak{g}$ denote the tangent space $\mathcal{T}_I G$ of a matrix group G at the identity I. Then*

1. *The set $\mathfrak{g}$ is a Lie subalgebra in $\mathbb{R}^{n\times n}$. That is, if $\alpha'(0)$ and $\beta'(0)$ are two elements in $\mathfrak{g}$, then the Lie bracket $[\alpha'(0), \beta'(0)]$ is also an element $\mathfrak{g}$.*

2. *The tangent space of a matrix group has the same structure everywhere in the sense that the tangent space at any elements A in G is a translation of $\mathfrak{g}$ via*

$$\mathcal{T}_A G = A\mathfrak{g}. \tag{3.4}$$

3. *The Lie subalgebra $\mathfrak{g}$ can be characterized as the logarithm of G in the sense that*

$$\mathfrak{g} = \{M \in \mathbb{R}^{n\times n} | \exp(tM) \in G, \text{ for all } t \in \mathbb{R}\}. \tag{3.5}$$

Tangent spaces for some of the groups mentioned earlier are listed in Table 3.3.

Group G	Algebra $\mathfrak{g}$	Characteristics
$\mathcal{G}l(n)$	$gl(n)$	$\mathbb{R}^{n\times n}$
$\mathcal{S}l(n)$	$sl(n)$	$\{M \in gl(n) \vert \mathrm{trace}(M) = 0\}$
$\mathcal{A}ff(n)$	$aff(n)$	$\{\begin{bmatrix} M & \mathbf{t} \\ \mathbf{0} & 0 \end{bmatrix} \vert M \in gl(n), \mathbf{t} \in \mathbb{R}^n\}$
$\mathcal{O}(n)$	$o(n)$	$\{K \in gl(n) \vert \mathrm{K}$ is skew-symmetric$\}$
$\mathcal{I}som(n)$	$isom(n)$	$\{\begin{bmatrix} K & \mathbf{t} \\ \mathbf{0} & 0 \end{bmatrix} \vert K \in o(n), \mathbf{t} \in \mathbb{R}^n\}$
$G_1 \times G_2$	$\mathcal{T}_{(e_1,e_2)} G_1 \times G_2$	$\mathfrak{g}_1 \times \mathfrak{g}_2$

TABLE 3.3
Example of tangent spaces.

The exponential map $\exp : \mathfrak{g} \to G$, as we have seen in (2.11), is a central step from a Lie algebra $\mathfrak{g}$ to the associated Lie group G [8]. Since exp is a local homeomorphism, mapping a neighborhood of the zero matrix O in $\mathfrak{g}$ onto a neighborhood of the identity matrix I in G, any dynamical system in G in the neighborhood of I therefore would have a corresponding dynamical system in $\mathfrak{g}$ in the neighborhood of O. It is easy to see that for any fixed $M \in \mathfrak{g}$, the map

$$\gamma(t) := \exp(tM) \tag{3.6}$$

defines a one-parameter subgroup in G.

Thus far we have drawn only an outline suggesting that different transformations can be unified under the same framework of tracing orbits associated with corresponding group actions. Under this framework it is plausible that more sophisticated actions could be composed that, in turn, might offer the design of new numerical algorithms. We point out specifically that, because of the Lie structure, the tangent space structure of a matrix group is the same at every of its element. It is yet to be determined how a dynamical system should be defined over an orbit or, equivalently, over a group so as to locate a certain canonical matrix on the orbit. Toward that end, the notion of "simplicity" of a matrix must be addressed. We shall demonstrate that such a conception varies according to the applications in the next section. Once the goal is set, we shall also discuss how various objective functions could be used to control the dynamical systems that drive curves on the orbits to a desired limit point. The vector field of such a dynamical system, if can be defined, must be lying in the tangent space. This requisite usually can be assured by a proper projection onto the tangent space. If the tangent space of a matrix group is known, then the projection usually can be calculated explicitly.

We illustrate the idea of projection as follows. By Theorem 3.2, the tangent space of $\mathcal{O}(n)$ at any orthogonal matrix Q is

$$\mathcal{T}_Q\mathcal{O}(n) = Qo(n).$$

It is not difficult to see that the normal space of $\mathcal{O}(n)$ at any orthogonal matrix Q is

$$\mathcal{N}_Q\mathcal{O}(n) = Qo(n)^{\perp},$$

where the orthogonal complement $o(n)^{\perp}$ is precisely the subspace of all symmetric matrices. The space $\mathbb{R}^{n\times n}$ can be split as the direct sum of

$$\mathbb{R}^{n\times n} = Qo(n) \oplus Qo(n)^{\perp}.$$

Any $X \in \mathbb{R}^{n\times n}$ therefore has a unique orthogonal splitting as

$$X = Q(Q^TX) = Q\left\{\frac{1}{2}(Q^TX - X^TQ)\}+Q\{\frac{1}{2}(Q^TX + X^TQ)\right\}.$$

The projection of X onto the tangent space $\mathcal{T}_Q\mathcal{O}(n)$ is given by the formula

$$\mathrm{Proj}_{\mathcal{T}_Q\mathcal{O}(n)}X = Q\left\{\frac{1}{2}(Q^TX - X^TQ)\right\}. \tag{3.7}$$

4. Canonical Form. We point out earlier that one of the main purposes in a realization process is to identify the simplest representation of a given linear transformation. The superlative adjective "simplest" is a relative term which should be interpreted broadly. Roughly speaking, canonical form refers to a "specific structure" by which a certain conclusion can be drawn or a certain goal can be achieved. Thus, depending on the applications, the specific structure could refer to a matrix with a specified pattern of zeros, such as a diagonal, tridiagonal, or triangular matrix. It could also refer to a matrix with a specified construct, such Toeplitz, Hamiltonian, stochastic, or other linear varieties. It could even refer to a matrix with a specified algebraic constraint, such as low rank or nonnegativity. To reach each of the different canonical forms, different group actions and different curves on the orbits must be taken. We find that continuous group actions often enable to tackle existence problems that are seemingly impossible to be solved by conventional discrete methods.

Listed in Table 4.1 are some commonly used canonical forms together with desired actions to reach these forms.

It should be noted that in the last example above we have slightly extended the meaning of an orbit. The map $(U,S,V) \hookrightarrow \left(\mathrm{diag}\left(USS^\top U^\top\right)\right)^{-1/2} USV^\top$ is not a group action. However, it is a natural way to define the set of rank k matrices whose rows are normalized to unity. The manifold of images under such a nonlinear map certainly is nontrivial but, analogous to a group orbit, is still computational traceable if we know how to manage the changes of U, S, and V in their respective groups. Generalizations such as this offer considerably more flexibility in applications. We shall see few more similar generalizations in the sequel.

5. Objective Functions. The orbit of a selected group action only defines the rule by which a transformation is to take place. To connect the current point and the desired canonical form on a given orbit, a bridge in the form of either an iterative scheme or a differential equation must be properly formulated. With the help of some objective functions, it is easy to control the construction of such a bridge. In this section, we outline some ideas on how objection functions could be designed.

Canonical form	Also know as	Action
Bidiagonal J	Quasi-Jordan Decomp., $A \in \mathbb{R}^{n\times n}$	$P^{-1}AP = J$, $P \in \mathcal{G}l(n)$
Diagonal Σ	Sing. Value Decomp., $A \in \mathbb{R}^{m\times n}$	$U^\top AV = \Sigma$, $(U,V) \in \mathcal{O}(m) \times \mathcal{O}(n)$
Diagonal pair (Σ_1, Σ_2)	Gen. Sing. Value Decomp., $(A,B) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{p\times n}$	$(U^\top AX, V^\top BX) = (\Sigma_1, \Sigma_2)$, $(U,V,X) \in \mathcal{O}(m) \times \mathcal{O}(p) \times \mathcal{G}l(n)$
Upper quasi-triangular H	Real Schur Decomp., $A \in \mathbb{R}^{n\times n}$	$Q^\top AQ = H$, $Q \in \mathcal{O}(n)$
Upper quasi-triangular H Upper triangular U	Gen. Real Schur Decomp., $A, B \in \mathbf{R}^{n\times n}$	$(Q^\top AZ, Q^\top BZ) = (H, U)$, $Q, Z \in \mathcal{O}(n)$
Symmetric Toeplitz T	Toeplitz Inv. Eigenv. Prob., $\{\lambda_1, \ldots, \lambda_n\} \subset \mathbb{R}$ is given	$Q^\top \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\} Q = T$, $Q \in \mathcal{O}(n)$
Nonnegative $N \geq 0$	Nonneg. inv. Eigenv. Prob., $\{\lambda_1, \ldots, \lambda_n\} \subset \mathbb{C}$ is given	$P^{-1}\mathrm{diag}\{\lambda_1, \ldots, \lambda_n\} P = N$, $P \in \mathcal{G}l(n)$
Linear variety X with fixed entries at fixed locations	Matrix Completion Prob., $\{\lambda_1, \ldots, \lambda_n\} \subset \mathbb{C}$ is given $X_{i_\nu, j_\nu} = a_\nu, \nu = 1, \ldots, \ell$	$P^{-1}\{\lambda_1, \ldots, \lambda_n\} P = X$, $P \in \mathcal{G}l(n)$
Nonlinear variety with fixed singular values and eigenvalues	Test Matrix Construction, $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$ and $\Sigma = \mathrm{diag}\{\sigma_1, \ldots \sigma_n\}$ are given	$P^{-1}\Lambda P = U^\top \Sigma V$ $P \in \mathcal{G}l(n), \quad U, V \in \mathcal{O}(n)$
Maximal fidelity	Structured Low Rank Approx. $A \in \mathbb{R}^{m\times n}$	$\left(\mathrm{diag}\left(USS^\top U^\top\right)\right)^{-1/2} USV^\top$, $(U,S,V) \in \mathcal{O}(m) \times \mathbb{R}^k_\times \times \mathcal{O}(n)$

TABLE 4.1
Examples of canonical forms used in practice.

5.1. Least Squares and Projected Gradient. We begin with the most general setting of projected gradient flows on a given continuous matrix group $G \subset \mathcal{G}l(n)$. Let X be a fixed matrix in a subset $\mathbb{V} \subset \mathbb{R}^{n\times n}$. Let $f : \mathbb{V} \longrightarrow \mathbb{R}^{n\times n}$ be a differentiable map with a certain "inherent" properties such as symmetry, isospectrality, low rank, or other algebraic constraints. Suppose that a group action $\mu : G \times \mathbb{V} \longrightarrow \mathbb{V}$ has been specified. Suppose also that a projection map P from $\mathbb{R}^{n\times n}$ onto a singleton, a linear subspace, or an affine subspace $\mathbb{P}$ in $\mathbb{R}^{n\times n}$ is available. Elements in the set $\mathbb{P}$ carry a certain desired structure, for example, the canonical form. Consider the functional $F : G \longrightarrow \mathbb{R}$

$$F(g) := \frac{1}{2}\|f(\mu(g,X)) - P(\mu(g,X))\|_F^2. \tag{5.1}$$

The goal is to minimize F over G. The meaning of this constrained minimization is that, while staying in the orbit of X under the action of μ and maintaining the inherent property guaranteed by the function f, we look for the element $g \in G$ so that the matrix $f(\mu(g, X))$ best realizes the desired canonical structure in the sense of least squares.

Though iterative methods based on conventional optimization techniques for (5.1) are possible, we find that the projected gradient flow approach can conveniently be formulated as follows:

ALGORITHM 5.1. *(Projected Gradient Flow) A generic projected gradient flow can be constructed via three steps:*

1. *Compute the gradient* $\nabla F(g)$.
2. *Project* $\nabla F(g)$ *onto the tangent space* $\mathcal{T}_g G$ *of* G *at* g.
3. *Follow the projected gradient numerically until convergence.*

We note that there are many new techniques developed recently for dynamical systems on Lie groups, including the RK-MK methods [32], Magnus and Fer expansions [4] and so on. A partial list of references for Lie structure preserving algorithms can be found in the seminal review paper by Iserles el al [30] and the book by Hairer el al [29]. These new ODE techniques certainly can benefit the computations described in this paper, but we shall avoid discussing computational specifics in the current presentation.

Applications embracing the framework of projected gradient approach are plenty. The following differential equations exemplify various dynamical systems on some appropriate orbits after computing gradients of some appropriate objective functions and performing projections onto some appropriate tangent spaces. Many other applications can be found in the excellent book by Helmke and Moore [27]. To keep this paper concise, we shall not give out great details for each of the cases. Readers are referred to the references for more careful layout of the settings.

Example 1. Given a symmetric matrix Λ and a desirable structure $\mathbb{V}$, find a symmetric matrix X in $\mathbb{V}$ or closest to $\mathbb{V}$ and has the same spectrum as Λ [16].

$$\left\{ \begin{array}{rcl} \dot{X} & = & [X, [X, P(X)]], \\ X(0) & = & \Lambda. \end{array} \right. \tag{5.2}$$

There are many important applications of (5.2). We point out specifically that, by taking $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$ and by choosing the structure retained in $\mathbb{V}$, the flow (5.2) can be used to tackle various kinds of very difficult structured inverse eigenvalue problems [17].

Example 2. Given a sequence of matrices $A_1, \ldots, A_k \in \mathbb{R}^{n \times n}$ and desirable structures in each individual $\mathbb{V}_1, \ldots, \mathbb{V}_k$, reduce each A_i to X_i that is closest to the structure $\mathbb{V}_i$ simultaneously by orthogonal similarity transformations [12].

$$\left\{ \begin{array}{rcl} \dot{X}_i & = & \left[X_i, \sum_{j=1}^{p} \frac{[X_j, P_j^T(X_j)] - [X_j, P_j^T(X_j)]^T}{2} \right], \\ X_i(0) & = & A_i, \quad i = 1, \ldots k. \end{array} \right. \tag{5.3}$$

A similar dynamical system can be derived if orthogonal equivalence transformations are used. Simultaneous reduction problems arise from areas such system identification where it is desirable to model complicated phenomena by as fewer variables as possible.

Example 3. Given a matrix $A \in \mathbb{C}^{n \times n}$, find its nearest normal matrix approximation $W \in \mathbb{C}^{n \times n}$ [12, 33].

$$\left\{ \begin{array}{rcl} \dot{W} & = & \left[W, \frac{[W, diag(W^*)] - [W, diag(W^*)]^*}{2} \right], \\ W(0) & = & A. \end{array} \right. \tag{5.4}$$

We note that this interesting approximation problem can be solved by using the unitary group over the complex domain.

Example 4. Construct a symmetric matrix with prescribed diagonal entries $\mathbf{a}$ and spectrum $\{\lambda_1, \ldots, \lambda_n\}$ [13].

$$\left\{ \begin{array}{rcl} \dot{X} & = & [X, [\mathrm{diag}(X) - \mathrm{diag}(\mathbf{a}), X]], \\ X(0) & = & \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}. \end{array} \right. \tag{5.5}$$

This inverse problem is the harder part of the Schur-Horn Theorem [25].

Example 5. Given a matrix pencil $A - \lambda B$ where A is symmetric and B is symmetric and positive definite, and two desirable structures $\mathbb{V}_1$ and $\mathbb{V}_2$, find a matrix pencil $X - \lambda Y$ in or closest to the structure under the congruence transformation [18].

$$\left\{ \begin{array}{rcl} \dot{X} & = & -\left((XW)^T + XW\right), \\ \dot{Y} & = & -\left((YW)^T + YW\right), \end{array} \right. \quad W := X(X - P_1(X)) + Y(Y - P_2(Y)). \tag{5.6}$$

This problem is a generalization of Example 5.1a to matrix pencils. The congruence transformation guarantees that the spectrum of the pencil $X - \lambda Y$ is the same as that of $A - \lambda B$.

5.2. Systems for Other Objectives. The least squares formulation (5.1) is but one way to define a dynamical system on an orbit. In this section, we briefly review some systems that are impelled by other objectives in mind. Some are defined even under no palpable objectives.

Example 6. The solution $X(t)$ to the Toda flow (2.4) starting with a symmetric and tridiagonal initial value $X(0)$ will remain to be tridiagonal and symmetric for all $t \in \mathbb{R}$. The lattice actually arises from the dynamics of a special mass-spring system [22, 34]. It is the setup of the physical configuration that induces the tridiagonal structure. It is the law of the Hamiltonian equations that governs the definition of the vector field. Other than these, there was no specific objective function used. The extension of (2.4) to general matrices,

$$\dot{X} = [X, \Pi_0(G(X))], \tag{5.7}$$

where $G(z)$ is an analytic function over spectrum of $X(0)$, appears to be done totally by brutal force and blind reasoning [11, 35]. Nevertheless, the differential equation approach nicely explains the pseudo-convergence and convergence behavior of the classical QR algorithm for general and normal matrices, respectively. The sorting of eigenvalues at the limit point had been observed, but did not seem to be clearly understood until Brockett's double bracket flow [6],

$$\dot{X} = [X, [X, N]], \tag{5.8}$$

where N is a fixed and symmetric matrix, was developed. The differential system (5.8) is a special case of the projected gradient flow (5.1) where $G = \mathcal{O}(n)$, $\mu(Q, X) = Q^\top \Lambda Q$, $P(\mu(Q, X)) \equiv N$, and f is the identity map. It follows that sorting is necessary in the first order optimality condition [11]. Furthermore, by taking the special diagonal matrix $N = \mathrm{diag}\{n, n-1, \ldots, 2, 1\}$ in the symmetric and tridiagonal case, it can be checked that the double bracket flow is exactly the same as the Toda lattice [5]. Thus, here comes the sudden understanding that the classical Toda lattice does have an objective function in mind.

The componentwise scaled Toda flow [14],

$$\dot{X} = [X, K \circ X], \tag{5.9}$$

where K is fixed and skew-symmetric and $\circ$ denotes the Hadamand product, is yet another venture into abstract mathematics where no explicit objective function in sight. By varying K, it offers considerable flexibilities in componentwise scaling and enjoy very general convergence behavior.

Example 7. Consider further some flows on the orbit $Orb_{\mathcal{O}(m)\times\mathcal{O}(n)}(X)$ under equivalence actions. Any flow on the orbit $Orb_{\mathcal{O}(m)\times\mathcal{O}(n)}(X)$ under equivalence must be of the form

$$\dot{X} = X(t)h(t) - k(t)X(t),$$

where $h(t) \in o(n)$ and $k(t) \in o(m)$ are functions of skew-symmetric matrices.

Mimicking the way that the Toda flow is related to the QR algorithm, the differential system

$$\left\{ \begin{array}{rcl} \dot{X}_1 &=& X_1\Pi_0(X_2^{-1}X_1) - \Pi_0(X_1X_2^{-1})X_1, \\ \dot{X}_2 &=& X_2\Pi_0(X_2^{-1}X_1) - \Pi_0(X_1X_2^{-1})X_2, \end{array} \right. \tag{5.10}$$

is a continuous realization of the discrete QZ algorithm that has used for solving the generalized eigenvalue problem [9]. The original objective in deriving (5.10) was simply to fill in the natural generalization of the Toda flow in the same way as the QZ algorithm generalizes the QR algorithm.

In contrast, the SVD flow [10],

$$\left\{ \begin{array}{rcl} \dot{Y} &=& Y\Pi_0\left(Y(t)^\top Y(t)\right) - \Pi_0\left(Y(t)Y(t)^\top\right)Y, \\ Y(0) &=& Y_0, \end{array} \right. \tag{5.11}$$

where Y_0 is a bidiagonal matrix was designed with two specific objectives in mind: one is that both skew-symmetric matrices $h(t)$ and $k(t)$ are kept tridiagonal and the other is that $Y(t)$ is kept bidiagonal throughout the continuous transition. Componentwise, the resulting dynamical system can be represented simply by

$$\dot{x}_{i,i} = x_{i,i}\left(x_{i,i+1}^2 - x_{i-1,i}^2\right), \quad 1 \le i \le n,$$
$$\dot{x}_{j,j+1} = x_{j,j+1}\left(x_{j+1,j+1}^2 - x_{j,j}^2\right), \quad 1 \le j \le n-1,$$

but we think that the matrix equation (5.11) clearly manifests the group actions. It turns out that the flow (5.11) gives rise to the Toda flows for the tridiagonal matrices $Y^\top Y$ and $YY^\top$. We mention that a related flow in fact interpolates the SVD algorithm and has been used to prove certain stability properties of a hybrid-SVD algorithm introduced by Demmel and Kahan. See [24].

6. Generalization to Non-group Structures. Thus far, it appears that the orthogonal group $\mathcal{O}(n)$ is the most frequently employed group for actions. Though orthogonal group actions seem to be significant enough, it is at least of theoretical interest to ask the following questions:

- Can any advantages or applications be taken out of actions by other matrix groups?
- What can be said of the limiting behavior of dynamical systems defined on orbits by using different groups and actions?
- Does the isometry group, as a larger group, offer any serious alternatives to the orthogonal group?

As many as there are unexplored groups or actions, so are there many of these questions yet to be answered. Not all these groups or actions will be useful, but we think that there is a wide open area in this direction that deserves further study.

Furthermore, we think that the idea of group actions, least squares, and the corresponding gradient flows can be extended to other structures that have the construct of groups but are not groups. Some of the structures that have found applications in practice include, for example, the Stiefel manifold,

$$\mathcal{O}(p,q) := \{Q \in \mathbb{R}^{p\times q} | Q^\top Q = I_q\}; \tag{6.1}$$

the manifold of oblique matrices,

$$\mathcal{OB}(m,n) := \{Q \in \mathbb{R}^{m\times n} | \mathrm{diag}(QQ^\top) = I_m\}; \tag{6.2}$$

the cone of nonnegative matrices,

$$\mathcal{C}(n) := \{R \circ R | R \in \mathcal{R}^{n \times n}\}; \tag{6.3}$$

the manifold of low rank matrices,

$$\mathcal{L}(m, n, k) := \mathcal{O}(m, k) \times \mathbb{R}^k_\times \times \mathcal{O}(n, k); \tag{6.4}$$

and semigroups such as the standard symplectic matrices,

$$\mathcal{Z}(n) := \left\{ \begin{bmatrix} \alpha & \alpha t \\ s\alpha & \alpha^{-\top} + s\alpha t \end{bmatrix} \in \mathbb{R}^{2n \times 2n} \, | s \text{ and } t \text{ symmetric and positive definite} \right\}. \tag{6.5}$$

These sets do not meet the criteria of being a Lie group, but can be characterized by using the product topology of some separate groups. In other words, groups can be used as some superficial coordinate systems to describe the motions on these manifolds. The framework discussed earlier thus can still be applied.

We illustrate how to use the above ideas to tackle some very challenging problems. So far as we know, no effective iterative methods are available yet that can address these difficulties.

Example 8. The stochastic inverse eigenvalue problem is to construct a stochastic matrix with prescribed spectrum. The best known result in this regard is the Karpelevic theorem which characterizes completely the set Θ_n of eigenvalues of *all* $n \times n$ stochastic matrices [31]. The forward problem of eigenvalue locations is complicated enough. As is illustrated in Figure 6.1, the boundaries of Θ_n involve highly nonlinear curves. The inverse problem is even harder because,

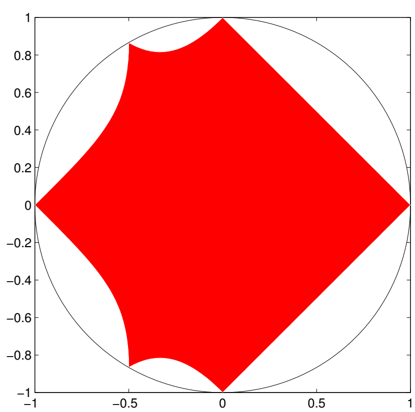

FIG. 6.1. Θ_4 *by the Karpelevič theorem.*

even if all n points in a given set are in Θ_n, there is no obvious way to qualify whether these given n points would belong to the same spectrum of a certain stochastic matrix.

Actually, it is generically true that the stochastic inverse eigenvalue problem would have been solved if the nonnegative inverse eigenvalue problem were solved. But even for the latter, the inverse eigenvalue problem has been a long standing open question.

Combining the ideas of least squares, actions and the generalizations, we can recast the inverse problem through a continuous realization process as the following minimization problem:

$$\begin{aligned} \text{Minimize} \quad & F(g, R) := \frac{1}{2} ||gJg^{-1} - R \circ R||^2, \\ \text{Subject to} \quad & g \in \mathcal{G}l(n), \; R \in gl(n), \end{aligned}$$

where J is a real-valued matrix carrying the prescribed spectral information (since the spectrum could be complex-valued, but is closed under complex conjugation). Note that the constraints literally are immaterial because both $\mathcal{G}l(n)$ and $gl(n)$ are open sets. No projection onto the constraints is needed. Thus

$$\begin{cases} \dot{g} &=& [(gJg^{-1})^\top, \alpha(g,R)]g^{-\top} \\ \dot{R} &=& 2\alpha(g,R)\circ R, \end{cases} \tag{6.6}$$

with $\alpha(g,R) := gJg^{-1} - R\circ R$, represents the steepest descent flow in $\mathcal{G}l(n)\times gl(n)$. To further stabilize the computation and avoid g^{-1}, we introduce a "parametrization " of g by its analytic singular value decomposition [7, 36] in the product group $\mathcal{O}(n)\times\mathbb{R}^n_\times\times\mathcal{O}(n)$. Suppose $g(t) = X(t)S(t)Y(t)^\top$ is the singular value decomposition of $g(t)$ where $S(t)$ is a diagonal matrix with elements from $\mathbb{R}^n_\times$ and $X(t)$ and $Y(t)$ are elements from $\mathcal{O}(n)$. Then the relationship of derivatives

$$X^T\dot{g}Y = \underbrace{X^T\dot{X}}_{Z}S + \dot{S} + S\underbrace{\dot{Y}^TY}_{W}, \tag{6.7}$$

is clearly true. Define $\Upsilon := X^\top\dot{g}Y$ with $\dot{g}$ given by (6.6). Then we now has a differential system,

$$\begin{cases} \dot{S} &=& \operatorname{diag}(\Upsilon), \\ \dot{X} &=& XZ, \\ \dot{Y} &=& YW, \end{cases} \tag{6.8}$$

that governs how the triplet $(X(t),S(t),Y(t))$ should be varied in the group $\mathcal{O}(n)\times\mathbb{R}^n_\times\times\mathcal{O}(n)$. Note that in the above Z and W are skew-symmetric matrices obtainable from off-diagonal elements of Υ and S. Total together, the objective function F now is a function of the four variables (X,S,Y,R) in $\mathcal{O}(n)\times\mathbb{R}^n_\times\times\mathcal{O}(n)\times gl(n)$

Example 9. Given a matrix $A\in\mathbb{R}^{n\times m}$ whose rows are of unit length, consider the problem of finding an approximation matrix $Z\in\mathbb{R}^{n\times m}$ of A whose rows are also of unit length, but $\operatorname{rank}(Z) = k$ where $k<\min\{m,n\}$. This low rank approximation problem in the oblique space $\mathcal{OB}(n,m)$ arises from many areas of important applications including factor analysis and data mining [15]. The tangling of two constraints, unit row length and low rank, makes the approximation difficult to find. However, we can recast the problem as minimizing the functional,

$$E(U,S,V) := \frac{1}{2}\|\left(\operatorname{diag}(USSU^\top)\right)^{-1/2}USV^\top - A\|_F^2, \tag{6.9}$$

with $U\in\mathcal{O}(n,k)$, $S\in\mathbb{R}^k_\times$, and $V\in\mathcal{O}(m,k)$, where $\mathcal{O}(p,q)$ denotes the Stiefel manifold. By construction, the product $Z = \left(\operatorname{diag}(USSU^\top)\right)^{-1/2}USV^\top$ is guaranteed to be of rank k and in $\mathcal{OB}(n,m)$. We thus can control the feasible matrix Z by controlling the variables (U,S,V). The Stiefel manifold is not a group, but its tangent space can be explicitly calculated. It can be shown that the projection $P_{\mathcal{O}(p,q)}(M)$ of any matrix $M\in\mathbb{R}^{p\times q}$ onto the tangent space $\mathcal{T}_Q\mathcal{O}(p,q)$ is given by

$$\mathcal{P}_{\mathcal{O}(p,q)}(M) = Q\frac{Q^\top M - M^\top Q}{2} + (I - QQ^\top)M. \tag{6.10}$$

Replacing M in (6.10) by the appropriate partial derivatives of E with respect to U and V, respectively, we have established the projected gradient flow. Detailed calculations and numerical evidence are given in [15].

7. Conclusion. There is a close relationship between matrix groups and linear transformations. Group actions together with properly formulated objective functions can offer a channel to tackle various classical or new and challenging problems arisen from applied linear algebra. This note outlines some basic ideas and examples with the hope to bring together the notions of group theory, linear transformations, and dynamical systems as a tool to undertake the task of system identification by canonical forms. More sophisticated actions can be composed that might offer the design of new numerical algorithms. The list of applications continues to grow. New computational techniques for structured dynamical systems on matrix group will further extend and benefit the scope of this interesting topic.

REFERENCES

[1] V. I. Arnold, Geometrical Methods in the Theory of Ordinary Differential Equations, 2ed., Springer-Verlag, New York, 1988.

[2] A. Baker, Matrix Groups, An Introduction to Lie Group Theory, Springer-Verlag, London, 2002.

[3] J. G. F. Belinfante and B. Kolman, A Survey of Lie Groups and Lie Algebras with Applications and Computational Methods, Classics in Applied Mathematics 2, SIAM, Philadelphia, 1989.

[4] S. Blanes, F. Casas. J.A. Oteo and J. Ros, Magnus and Fer expansions for matrix differential equations: The convergence problem, J. Phys. A31(1998), 259-268.

[5] A. M. Bloch, Steepest descent, linear programming, and Hamiltonian flows, Contemporary Math., 114(1990), 77-88.

[6] R. W. Brockett, Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems, in Proceedings of the 27th IEEE Conference on Decision and Control, IEEE, (1988), 799-803, and Linear Alg. Appl., 146(1991), 79-91.

[7] A. Bunse-Gerstner, R. Byers, V. Mehrmann and N. K. Nichols, Numerical computation of an analytic singular value decomposition of a matrix valued function, Numer. Math. 60(1991), 1-40.

[8] E. Celledoni and A. Iserles, Approximating the exponential from a Lie algebra to a Lie group, Math. Compu., 69(2000), 1457-1480.

[9] M. T. Chu, A continuous approximation to the generalized Schur decomposition, Linear Alg. Appl., 78(1986), 119-132.

[10] M. T. Chu, On a differential equation approach to the singular value decomposition of bi-diagonal matrices, Linear Alg. Appl., 80(1986), 71-79.

[11] M. T. Chu, On the continuous realization of iterative processes, SIAM Review, 30(1988), 375-387.

[12] M. T. Chu, A continuous Jacobi-like approach to the simultaneous reduction of real matrices, Linear Alg. Appl., 147(1991), 75-96.

[13] M. T. Chu, Constructing a Hermitian matrix from its diagonal entries and eigenvalues, SIAM J. Matrix Appl., 16(1995), 207-217.

[14] M. T. Chu, Scaled Toda-like flows, Linear Alg. Appl., 215(1995), 261-273.

[15] M. T. Chu, N. Del Buono, L. Lopez, and T. Politi, On the low rank approximation of data on the unit sphere, preprint, 2003.

[16] M. T. Chu and K. R. Driessel, The projected gradient method for least squares matrix approximations with spectral constraints, SIAM J. Numer. Anal., 27(1990), 1050-1060.

[17] M. T. Chu and G. Golub, Structured inverse eigenvalue problems, Acta Numerica, invited paper, 2002, 1-71.

[18] M. T. Chu and Q. Guo, On the least squares approximation of symmetric-definite pencils subject to generalized spectral constraints, SIAM J. Matrix Anal. Appl., 19(1998), 1-20.

[19] M. T. Chu and L. K. Norris, Isospectral flows and abstract matrix factorizations, SIAM J. Numer. Anal., 25(1988), 1383-1391.

[20] M. L. Curtis, Matrix Groups, 2nd ed., Springer-Verlag, New York, 1984.

[21] J. Della-Dora, Numerical Linear Algorithms and Group Theory, Linear Alg. Appl., 10(1975), 267-283.

[22] P. Deift, T. Nanda and C. Tomei, Differential equations for the symmetric eigenvalue problem, SIAM J. Numer. Anal., 20(1983), 1-22.

[23] P. Deift, L. C. Li and C. Tomei, Matrix factorizations and integrable systems, Comm. Pure Appl. Math. 42(1989), 443–521.

[24] P. Deift, J. W. Demmel, L.-C. Li, and C. Tomei, The bidiagonal singular value decomposition and Hamiltonian mechanics, SIAM J. Num. Anal., 28 (1991), pp. 1463-1516. (LAPACK Working Note 11).

[25] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, New York, 1991.

[26] G. H. Golub and C. F. Van Loan, Matrix Computations, 3rd., Johns Hopkins, Baltimore, 1996.

[27] U. Helmke and J. Moore, Optimization and Dynamical Systems, Communications and Control Engineering Series. Springer-Verlag, London, 1994.

[28] R. Howe, Very basic Lie theory, Amer. Math. Monthly, 90(1983), 600-623; Corr., Amer. Math. Monthly, 91(1984), 247.

[29] E. Hairer, C. Lubich, and G. Wanner, Geometric Numerical Integration, Structure-Preserving Algorithms for Ordinary Differential Equations, Springer Series in Comput. Math., Vol. 31, Springer-Verlag, 2002.

[30] A. Iserles, H. Munthe-Kaas, S. P. Nørsett, and A. Zanna, Lie-group methods, Acta Numerica, 9(2000), 215-365.

[31] H. Minc, Nonnegative matrices, Wiley, New York, 1988.

[32] H. Munthe-Kaas, Runge-Kutta methods on Lie groups, BIT, 38(1998), 92-111.

[33] A. Ruhe, Closest normal matrix finally found!, BIT, 27(1987), 585-598.

[34] W. W. Symes, The QR algorithm and scattering for the finite non-periodic Toda lattice, Physica, 4D(1982), 275-280.

[35] D. S. Watkins, Isospectral flows, SIAM Rev., 26(1984), 379-391.

[36] K. Wright, Differential equations for the analytic singular value decomposition of a matrix, Numer. Math., 63(1992), 283-295.