Group Theory, Linear Transformations, and Flows:
(Some) Dynamical Systems on Manifolds

Moody T. Chu
North Carolina State University
presented at
Workshop on Lie Group Methods and Control Theory
June 30, 2004
# Group Theory, Linear Transformations, and Flows: (Some) Dynamical Systems on Manifolds

1. **REPORT DATE**
   03 JAN 2005

2. **REPORT TYPE**
   N/A

3. **DATES COVERED**
   -

4. **TITLE AND SUBTITLE**
   Group Theory, Linear Transformations, and Flows: (Some) Dynamical Systems on Manifolds

5. **AUTHOR(S)**

6. **PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
   North Carolina State University

7. **SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

8. **PERFORMING ORGANIZATION REPORT NUMBER**

9. **DISTRIBUTION/AVAILABILITY STATEMENT**
   Approved for public release, distribution unlimited

10. **SUPPLEMENTARY NOTES**
    See also ADM001749, Lie Group Methods And Control Theory Workshop Held on 28 June 2004 - 1 July 2004., The original document contains color images.

11. **ABSTRACT**

12. **SUBJECT TERMS**

13. **SECURITY CLASSIFICATION OF:**
    - REPORT: unclassified
    - ABSTRACT: unclassified
    - THIS PAGE: unclassified

14. **LIMITATION OF ABSTRACT**
    UU

15. **NUMBER OF PAGES**
    47

16. **NAME OF RESPONSIBLE PERSON**

---

Standard Form 298 (Rev. 8-98)
Proscribed by ANSI Std Z39-18
Outline

- Motivation
  - Realization Process
  - A Case Study
- Basic Form
  - Similarity Property
  - Decomposition Property
  - Reversal Property
- Matrix Groups and Group Actions
- Tangent Space and Projection
- Canonical Forms
- Objective Functions and Dynamical Systems
  - Examples
    - Least Squares
- New Thoughts
- Conclusion
Motivation

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?

– V. I. Arnold


- What is the simplest form referred to here?
- What kind of continuous change can be employed?
Realization Process

- Realization process, in a sense, means any deducible procedure that we use to rationalize and solve problems.
  - The simplest form refers to the agility to think and draw conclusions.
- In mathematics, a realization process often appears in the form of an iterative procedure or a differential equation.
  - The steps taken for the realization, i.e., the changes, could be discrete or continuous.
Continuous Realization

- Two abstract problems:
  - One is a make-up and is easy.
  - The other is the real problem and is difficult.

- A bridge:
  - A continuous path connecting the two problems.
  - A path that is easy to follow.

- A numerical method:
  - A method for moving along the bridge.
  - A method that is readily available.
Build the Bridge

- Specified guidance is available.
  - The bridge is constructed by monitoring the values of certain specified functions.
  - The path is guaranteed to work.
  - Such as the projected gradient method.

- Only some general guidance is available.
  - A bridge is built in a straightforward way.
  - No guarantee the path will be complete.
  - Such as the homotopy method.

- No guidance at all.
  - A bridge is built seemingly by accident.
  - Usually deeper mathematical theory is involved.
  - Such as the isospectral flows.
A bridge, if it exists, usually is characterized by an ordinary differential equation.

The discretization of a bridge, or a numerical method in travelling along a bridge, usually produces an iterative scheme.
Two Examples

- Eigenvalue Computation
- Constrained Least Squares Approximation
The Eigenvalue Problem

- The mathematical problem:
  - A symmetric matrix $A_0$ is given.
  - Solve the equation
    $\begin{equation}
    A_0x = \lambda x
    \end{equation}$
    for a nonzero vector $x$ and a scalar $\lambda$.

- An iterative method:
  - The $QR$ decomposition:
    $A = QR$
    where $Q$ is orthogonal and $R$ is upper triangular.
  - The $QR$ algorithm (Francis’61):
    $\begin{align*}
    A_k &= Q_k R_k \\
    A_{k+1} &= R_k Q_k.
    \end{align*}$
    The sequence $\{A_k\}$ converges to a diagonal matrix.
  - Every matrix $A_k$ has the same eigenvalues of $A_0$, i.e.,
    $\begin{equation}
    (A_{k+1} = Q_k^T A_k Q_k).
    \end{equation}$
A continuous method:

- Lie algebra decomposition:
  \[ X = X^o + X^+ + X^- \]

  where \( X^o \) is the diagonal, \( X^+ \) the strictly upper triangular, and \( X^- \) the strictly lower triangular part of \( X \).

- Define \( \Pi_0(X) := X^- - X^+ \).

- The Toda lattice (Symes’82, Deift al’83):

  \[
  \begin{align*}
  \frac{dX}{dt} &= [X, \Pi_0(X)] \\
  X(0) &= X_0.
  \end{align*}
  \]

- Sampled at integer times, \( \{X(k)\} \) gives the same sequence as does the QR algorithm applied to the matrix \( A_0 = exp(X_0) \).

Evolution starts from \( X_0 \) and converges to the limit point of Toda flow, which is a diagonal matrix, maintains the spectrum.

- The construction of the Toda lattice is based on the physics.
  - This is a Hamiltonian system.
    - A certain physical quantities are kept at constant, i.e., this is a completely integrable system.
  - The convergence is guaranteed by “nature”? 
Least Squares Matrix Approximation

- The mathematical problem:
  - A symmetric matrix $N$ and a set of real values $\{\lambda_1, \ldots, \lambda_n\}$ are given.
  - Find a least squares approximation of $N$ that has the prescribed eigenvalues.

- A standard formulation:

  $$
  \text{Minimize } F(Q) := \frac{1}{2} ||Q^T \Lambda Q - N||^2
  $$

  Subject to $Q^T Q = I$.

- Equality Constrained Optimization:
  - Augmented Lagrangian methods.
  - Sequential quadratic programming methods.

- None of these techniques is easy.
  - The constraint carries lots of redundancies.
Motivation

- A continuous approach:
  - The projection of the gradient of $F$ can easily be calculated.
  - Projected gradient flow (Brockett’88, Chu&Driessel’90):
    \[
    \frac{dX}{dt} = [X, [X, N]] \\
    X(0) = \Lambda.
    \]
    \[
    X := Q^T \Lambda Q.
    \]
    \[
    \text{Flow } X(t) \text{ moves in a descent direction to reduce } ||X - N||^2.
    \]
    \[
    \text{The optimal solution } X \text{ can be fully characterized by the spectral decomposition of } N \text{ and is unique.}
    \]
- Evolution starts from an initial value 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 is a descent flow.
Equivalence

- (Bloch’90) Suppose $X$ is tridiagonal. Take $N = \text{diag}\{n, \ldots, 2, 1\}$, then $[X, N] = \Pi_0(X)$.
- A gradient flow hence becomes a Hamiltonian flow.
Basic Form

- Lax dynamics:

\[
\frac{dX(t)}{dt} := [X(t), k_1(X(t))] \\
X(0) := X_0.
\]

- Parameter dynamics:

\[
\frac{dg_1(t)}{dt} := g_1(t)k_1(X(t)) \\
g_1(0) := I.
\]

and

\[
\frac{dg_2(t)}{dt} := k_2(X(t))g_2(t) \\
g_2(0) := I.
\]

\[ k_1(X) + k_2(X) = X. \]
Basic Form

Similarity Property

\[ X(t) = g_1(t)^{-1}X_0g_1(t) = g_2(t)X_0g_2(t)^{-1} \].

- Define \( Z(t) = g_1(t)X(t)g_1(t)^{-1} \).
- Check

\[
\frac{dZ}{dt} = \frac{dg_1}{dt}Xg_1^{-1} + g_1\frac{dX}{dt}g_1^{-1} + g_1X\frac{dg_1}{dt}^{-1}
\]

\[
= (g_1k_1(X))Xg_1^{-1} + g_1(Xk_1(X) - k_1(X)X)g_1^{-1} + g_1X(-k_1(X)g_1^{-1})
\]

\[
= 0.
\]

- Thus \( Z(t) = Z(0) = X(0) = X_0 \).
Decomposition Property

\[ \exp(tX_0) = g_1(t)g_2(t). \]

- Trivially \( \exp(X_0t) \) satisfies the IVP

\[
\frac{dY}{dt} = X_0Y, \; Y(0) = I.
\]

- Define \( Z(t) = g_1(t)g_2(t) \).

- Then \( Z(0) = I \) and

\[
\frac{dZ}{dt} = \frac{dg_1}{dt}g_2 + g_1\frac{dg_2}{dt} = (g_1k_1(X))g_2 + g_1(k_2(X)g_2) = g_1Xg_2 = X_0Z \quad \text{(by Similarity Property)}.
\]

- By the uniqueness theorem in the theory of ordinary differential equations, \( Z(t) = \exp(X_0t) \).
Reversal Property

\[ \exp(tX(t)) = g_2(t)g_1(t). \]

- By Decomposition Property,

\[
\begin{align*}
  g_2(t)g_1(t) & = g_1(t)^{-1}\exp(X_0t)g_1(t) \\
  & = \exp(g_1(t)^{-1}X_0g_1(t)t) \\
  & = \exp(X(t)t).
\end{align*}
\]
Abstraction

- **QR-type Decomposition:**
  - Lie algebra decomposition of $gl(n) \iff$ Lie group decomposition of $Gl(n)$ in the neighborhood of $I$.
  - Arbitrary subspace decomposition $gl(n) \iff$ Factorization of a one-parameter semigroup in the neighborhood of $I$ as the product of two nonsingular matrices, i.e.,
    \[
    \exp(X_0t) = g_1(t)g_2(t).
    \]
  - The product $g_1(t)g_2(t)$ will be called the abstract $g_1g_2$ decomposition of $\exp(X_0t)$.

- **QR-type Algorithm:**
  - By setting $t = 1$, we have
    \[
    \exp(X(0)) = g_1(1)g_2(1) \\
    \exp(X(1)) = g_2(1)g_1(1).
    \]
  - The dynamical system for $X(t)$ is autonomous $\iff$ The above phenomenon 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.
Matrix Groups

- A subset of nonsingular matrices (over any field) which are closed under matrix multiplication and inversion is called a matrix group.
  - Matrix groups are central in many parts of mathematics and applications.
- A smooth manifold which is also a group where the multiplication and the inversion are smooth maps is called a Lie group.
  - The most remarkable feature of a Lie group is that the structure is the same in the neighborhood of each of its elements.
- (Howe’83) Every (non-discrete) matrix group is in fact a Lie group.
  - Algebra and geometry are intertwined in the study of matrix groups.
- Lots of realization processes used in numerical linear algebra are the results of group actions.
<table>
<thead>
<tr>
<th>Group</th>
<th>Subgroup</th>
<th>Notation</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>General linear</td>
<td>$\mathcal{G}l(n)$</td>
<td>${A \in \mathbb{R}^{n \times n}</td>
<td>\det(A) \neq 0}$</td>
</tr>
<tr>
<td>Special linear</td>
<td>$\mathcal{S}l(n)$</td>
<td>${A \in \mathcal{G}l(n)</td>
<td>\det(A) = 1}$</td>
</tr>
<tr>
<td>Upper triangular</td>
<td>$\mathcal{U}(n)$</td>
<td>${A \in \mathcal{G}l(n)</td>
<td>A$ is upper triangular$}$</td>
</tr>
<tr>
<td>Unipotent</td>
<td>$\mathcal{U}nip(n)$</td>
<td>${A \in \mathcal{U}(n)</td>
<td>a_{ii} = 1$ for all $i}$</td>
</tr>
<tr>
<td>Orthogonal</td>
<td>$\mathcal{O}(n)$</td>
<td>${Q \in \mathcal{G}l(n)</td>
<td>Q^\top Q = I}$</td>
</tr>
<tr>
<td>Generalized orthogonal</td>
<td>$\mathcal{O}_S(n)$</td>
<td>${Q \in \mathcal{G}l(n)</td>
<td>Q^\top SQ = S}; S$ is a fixed matrix $</td>
</tr>
<tr>
<td>Symplectic</td>
<td>$\mathcal{S}p(2n)$</td>
<td>$\mathcal{O}_J(2n); J := \begin{bmatrix} 0 &amp; I \ -I &amp; 0 \end{bmatrix}$</td>
<td></td>
</tr>
<tr>
<td>Lorentz</td>
<td>$\mathcal{L}or(n,k)$</td>
<td>$\mathcal{O}_L(n+k); L := \text{diag}{1,\ldots,1,-1,\ldots,-1}$</td>
<td></td>
</tr>
<tr>
<td>Affine</td>
<td>$\mathcal{A}ff(n)$</td>
<td>$\left{ \begin{bmatrix} A &amp; t \ 0 &amp; 1 \end{bmatrix}</td>
<td>A \in \mathcal{G}l(n), t \in \mathbb{R}^n \right}$</td>
</tr>
<tr>
<td>Translation</td>
<td>$\mathcal{T}rans(n)$</td>
<td>$\left{ \begin{bmatrix} I &amp; t \ 0 &amp; 1 \end{bmatrix}</td>
<td>t \in \mathbb{R}^n \right}$</td>
</tr>
<tr>
<td>Isometry</td>
<td>$\mathcal{I}som(n)$</td>
<td>$\left{ \begin{bmatrix} Q &amp; t \ 0 &amp; 1 \end{bmatrix}</td>
<td>Q \in \mathcal{O}(n), t \in \mathbb{R}^n \right}$</td>
</tr>
<tr>
<td>Center of $G$</td>
<td>$Z(G)$</td>
<td>${z \in G</td>
<td>zg = gz, \text{ for every } g \in G}; G$ is a given group $</td>
</tr>
<tr>
<td>Product of $G_1$ and $G_2$</td>
<td>$G_1 \times G_2$</td>
<td>${(g_1, g_2)</td>
<td>g_1 \in G_1, g_2 \in G_2}; (g_1, g_2) * (h_1, h_2) := (g_1 h_1, g_2 h_2); G_1$ and $G_2$ are given groups $</td>
</tr>
<tr>
<td>Quotient</td>
<td>$G/N$</td>
<td>${Ng</td>
<td>g \in G}; N$ is a fixed normal subgroup of $G$ $</td>
</tr>
<tr>
<td>Hessenberg</td>
<td>$\mathcal{H}ess(n)$</td>
<td>$\mathcal{U}nip(n)/\mathcal{Z}_n$</td>
<td></td>
</tr>
</tbody>
</table>
A function $\mu : G \times \mathbb{V} \rightarrow \mathbb{V}$ is said to be a *group action* of $G$ on a set $\mathbb{V}$ if and only if

- $\mu(gh, x) = \mu(g, \mu(h, x))$ for all $g, h \in G$ and $x \in \mathbb{V}$.
- $\mu(e, x) = x$, if $e$ is the identity element in $G$.

Given $x \in \mathbb{V}$, two important notions associated with a group action $\mu$:

- The *stabilizer* of $x$ is
  \[
  Stab_G(x) := \{ g \in G | \mu(g, x) = x \}.
  \]
- The *orbit* of $x$ is
  \[
  Orb_G(x) := \{ \mu(g, x) | g \in G \}.
  \]
<table>
<thead>
<tr>
<th>Set ( V )</th>
<th>Group ( G )</th>
<th>Action ( \mu(g, A) )</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbb{R}^{n \times n} )</td>
<td>Any subgroup</td>
<td>( g^{-1}Ag )</td>
<td>conjugation</td>
</tr>
<tr>
<td>( \mathbb{R}^{n \times n} )</td>
<td>( O(n) )</td>
<td>( g^\top Ag )</td>
<td>orthogonal similarity</td>
</tr>
<tr>
<td>( \mathbb{R}^{n \times n} \times \cdots \times \mathbb{R}^{n \times n} )</td>
<td>( k \times k )</td>
<td>Any subgroup</td>
<td>( (g^{-1}A_1g, \ldots, g^{-1}A_kg) )</td>
</tr>
<tr>
<td>( \mathbb{S}(n) \times \mathbb{S}_{PD}(n) )</td>
<td>Any subgroup</td>
<td>( (g^\top Ag, g^\top Bg) )</td>
<td>symm. positive definite pencil reduction</td>
</tr>
<tr>
<td>( \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n} )</td>
<td>( O(n) \times O(n) )</td>
<td>( (g_1^\top Ag_2, g_1^\top Bg_2) )</td>
<td>QZ decomposition</td>
</tr>
<tr>
<td>( \mathbb{R}^{m \times n} )</td>
<td>( O(m) \times O(n) )</td>
<td>( g_1^\top Ag_2 )</td>
<td>singular value decomp.</td>
</tr>
<tr>
<td>( \mathbb{R}^{m \times n} \times \mathbb{R}^{p \times n} )</td>
<td>( O(m) \times O(p) \times Gl(n) )</td>
<td>( (g_1^\top Ag_3, g_2^\top Bg_3) )</td>
<td>generalized singular value decomp.</td>
</tr>
</tbody>
</table>
Some Exotic Group Actions (yet to be studied!)

- 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.

  - What could be said if actions of the isometry group are used?

    - Being isometric, stability is guaranteed.
    - The inverse of an isometry matrix is easy.

      \[
      \begin{bmatrix}
      Q & t \\
      0 & 1
      \end{bmatrix}^{-1} = \begin{bmatrix}
      Q^\top & -Q^\top t \\
      0 & 1
      \end{bmatrix}.
      \]

    - The isometry group is larger than the orthogonal group.

- What could be said if actions of the orthogonal group plus shift are used?

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

- What could be said if action of the orthogonal group with scaling are used?

  \[\mu((Q, s), A) := sQ^\top AQ, \quad Q \in O(n), s \in \mathbb{R}_+,\]

  or

  \[\mu((Q, s, t), A) := \text{diag}(s)Q^\top AQ\text{diag}(t), \quad Q \in O, s, t \in \mathbb{R}_n.\]
Tangent Space and Project Gradient

- Given a group $G$ and its action $\mu$ on a set $V$, the associated orbit $Orb_G(x)$ characterizes the rule by which $x$ is to be changed in $V$.
  - Depending on the group $G$, an orbit is often too “wild” to be readily traced for finding the “simplest form” of $x$.
  - Depending on the applications, a path/bridge/highway/differential equation needs to be built on the orbit to connect $x$ to its simplest form.
- A differential equation on the orbit $Orb_G(x)$ is equivalent to a differential equation on the group $G$.
  - Lax dynamics on $X(t)$.
  - Parameter dynamics on $g_1(t)$ or $g_2(t)$.
- To stay in either the orbit or the group, the vector field of the dynamical system must be distributed in the tangent space of the corresponding manifold.
- Most of the tangent spaces for the matrix groups can be calculated explicitly.
- If some kind of objective function has been used to control the connecting bridge, its gradient should be projected to the tangent space.
Tangent Space in General

- Given a matrix group $G \leq \mathcal{G}l(n)$, the tangent space to $G$ at $A \in G$ can be defined as
  
  $T_A G := \{ \gamma'(0) | \gamma \text{ is a differentiable curve in } G \text{ with } \gamma(0) = A \}$.

- The tangent space $\mathfrak{g} = T_I G$ at the identity $I$ is critical.
  
  - $\mathfrak{g}$ is a Lie subalgebra in $\mathbb{R}^{n \times n}$, i.e.,
    
    $\text{If } \alpha'(0), \beta'(0) \in \mathfrak{g}, \text{ then } [\alpha'(0), \beta'(0)] \in \mathfrak{g}$

  - The tangent space of a matrix group has the same structure everywhere, i.e.,
    
    $T_A G = A \mathfrak{g}$.

  - $T_I G$ can be characterized as the logarithm of $G$, i.e.,
    
    $\mathfrak{g} = \{ M \in \mathbb{R}^{n \times n} | \exp(tM) \in G, \text{ for all } t \in \mathbb{R} \}$. 
<table>
<thead>
<tr>
<th>Group $G$</th>
<th>Algebra $\mathfrak{g}$</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Gl}(n)$</td>
<td>$\text{gl}(n)$</td>
<td>$\mathbb{R}^{n \times n}$</td>
</tr>
<tr>
<td>$\text{Sl}(n)$</td>
<td>$\text{sl}(n)$</td>
<td>${ M \in \text{gl}(n)</td>
</tr>
<tr>
<td>$\text{Aff}(n)$</td>
<td>$\text{aff}(n)$</td>
<td>${ \begin{bmatrix} M &amp; t \ 0 &amp; 0 \end{bmatrix}</td>
</tr>
<tr>
<td>$\mathcal{O}(n)$</td>
<td>$\mathcal{o}(n)$</td>
<td>${ K \in \text{gl}(n)</td>
</tr>
<tr>
<td>$\text{Isom}(n)$</td>
<td>$\text{isom}(n)$</td>
<td>${ \begin{bmatrix} K &amp; t \ 0 &amp; 0 \end{bmatrix}</td>
</tr>
<tr>
<td>$G_1 \times G_2$</td>
<td>$\mathcal{T}_{(e_1,e_2)}G_1 \times G_2$</td>
<td>$\mathfrak{g}_1 \times \mathfrak{g}_2$</td>
</tr>
</tbody>
</table>
The tangent space of $\mathcal{O}(n)$ at any orthogonal matrix $Q$ is

$$\mathcal{T}_Q \mathcal{O}(n) = Q \mathcal{K}(n)$$

where

$$\mathcal{K}(n) = \{ \text{All skew-symmetric matrices} \}.$$

The normal space of $\mathcal{O}(n)$ at any orthogonal matrix $Q$ is

$$\mathcal{N}_Q \mathcal{O}(n) = QS(n).$$

The space $\mathbb{R}^{n \times n}$ is split as

$$\mathbb{R}^{n \times n} = QS(n) \oplus Q\mathcal{K}(n).$$

A unique orthogonal splitting of $X \in \mathbb{R}^{n \times n}$:

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

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

$$\text{Proj}_{\mathcal{T}_Q \mathcal{O}(n)} X = Q \left\{ \frac{1}{2} (Q^T X - X^T Q) \right\}.$$
A canonical form refers to a “specific structure” by which a certain conclusion can be drawn or a certain goal can be achieved.

The superlative adjective “simplest” is a relative term which should be interpreted broadly.

- A matrix with a specified pattern of zeros, such as a diagonal, tridiagonal, or triangular matrix.
- A matrix with a specified construct, such as Toeplitz, Hamiltonian, stochastic, or other linear varieties.
- A matrix with a specified algebraic constraint, such as low rank or nonnegativity.
<table>
<thead>
<tr>
<th>Canonical form</th>
<th>Also know as</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bidiagonal $J$</td>
<td>Quasi-Jordan Decomp., $A \in \mathbb{R}^{n\times n}$</td>
<td>$P^{-1}AP = J,$ $P \in \mathfrak{gl}(n)$</td>
</tr>
<tr>
<td>Diagonal $\Sigma$</td>
<td>Sing. Value Decomp., $A \in \mathbb{R}^{m\times n}$</td>
<td>$U^\top AV = \Sigma,$ $(U,V) \in \mathcal{O}(m) \times \mathcal{O}(n)$</td>
</tr>
<tr>
<td>Diagonal pair $(\Sigma_1, \Sigma_2)$</td>
<td>Gen. Sing. Value Decomp., $(A,B) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{p\times n}$</td>
<td>$(U^\top AX, V^\top BX) = (\Sigma_1, \Sigma_2),$ $(U,V,X) \in \mathcal{O}(m) \times \mathcal{O}(p) \times \mathfrak{gl}(n)$</td>
</tr>
<tr>
<td>Upper quasi-triangular $H$</td>
<td>Real Schur Decomp., $A \in \mathbb{R}^{n\times n}$</td>
<td>$Q^\top AQ = H,$ $Q \in \mathcal{O}(n)$</td>
</tr>
<tr>
<td>Upper quasi-triangular $H$</td>
<td>Gen. Real Schur Decomp., $A,B \in \mathbb{R}^{n\times n}$</td>
<td>$(Q^\top AZ, Q^\top BZ) = (H,U),$ $Q,Z \in \mathcal{O}(n)$</td>
</tr>
<tr>
<td>Upper triangular $U$</td>
<td>Real Schur Decomp., $A \in \mathbb{R}^{n\times n}$</td>
<td>$Q^\top AQ = H,$ $Q \in \mathcal{O}(n)$</td>
</tr>
<tr>
<td>Symmetric Toeplitz $T$</td>
<td>Toeplitz Inv. Eigenv. Prob., {$\lambda_1, \ldots, \lambda_n$} $\subset \mathbb{R}$ is given</td>
<td>$Q^\top \text{diag}{\lambda_1,\ldots,\lambda_n}Q = T,$ $Q \in \mathcal{O}(n)$</td>
</tr>
<tr>
<td>Nonnegative $N \geq 0$</td>
<td>Nonneg. inv. Eigenv. Prob., {$\lambda_1, \ldots, \lambda_n$} $\subset \mathbb{C}$ is given</td>
<td>$P^{-1}\text{diag}{\lambda_1,\ldots,\lambda_n}P = N,$ $P \in \mathfrak{gl}(n)$</td>
</tr>
<tr>
<td>Linear variety $X$</td>
<td>Matrix Completion Prob., {$\lambda_1, \ldots, \lambda_n$} $\subset \mathbb{C}$ is given</td>
<td>$P^{-1}\text{diag}{\lambda_1,\ldots,\lambda_n}P = X,$ $P \in \mathfrak{gl}(n)$</td>
</tr>
<tr>
<td>Nonlinear variety</td>
<td>Test Matrix Construction, $\Lambda = \text{diag}{\lambda_1,\ldots,\lambda_n}$ and $\Sigma = \text{diag}{\sigma_1,\ldots,\sigma_n}$ are given</td>
<td>$P^{-1}\Lambda P = U^\top \Sigma V,$ $P \in \mathfrak{gl}(n),$ $U,V \in \mathcal{O}(n)$</td>
</tr>
<tr>
<td>Maximal fidelity</td>
<td>Structured Low Rank Approx., $A \in \mathbb{R}^{m\times n}$</td>
<td>$(\text{diag } (USS^\top U^\top))^{-1/2} USV^\top,$ $(U,S,V) \in \mathcal{O}(m) \times \mathbb{R}^k \times \mathcal{O}(n)$</td>
</tr>
</tbody>
</table>
Objective Functions

- The orbit of a selected group action only defines the rule by which a transformation is to take place.
- Properly formulated objective functions helps to control the construction of a bridge between the current point and the desired canonical form on a given orbit.
  - The bridge often assumes the form of a differential equation on the manifold.
  - The vector field of the differential equation must distributed over the tangent space of the manifold.
  - Corresponding to each differential equation on the orbit of a group action is a differential equation on the group, and vice versa.
- How to choose appropriate objective functions?
Some Flows on $\text{Orb}_{\mathcal{O}(n)}(X)$ under Conjugation

- Toda lattice arises from a special mass-spring system (Symes’82, Deift et al.’83),

$$\frac{dX}{dt} = [X, \Pi_0(X)], \quad \Pi_0(X) = X^{-} - X^{-T},$$

$$X(0) = \text{tridiagonal and symmetric.}$$

- No specific objective function is used.
  - Physics law governs the definition of the vector field.

- Generalization to general matrices is totally by brutal force and blindness (and by the then young and desperate researchers) (Chu’84, Watkins’84).

$$\frac{dX}{dt} = [X, \Pi_0(G(X))], \quad G(z) \text{ is analytic over spectrum of } X(0).$$

  - But nicely explains the pseudo-convergence and convergence behavior of the classical QR algorithm for general and normal matrices, respectively.
  - Sorting of eigenvalues at the limit point is observed, but not quite clearly understood.
- Double bracket flow (Brockett’88),
  \[
  \frac{dX}{dt} = [X, [X, N]], \quad N = \text{fixed and symmetric.}
  \]
  
  ◦ This is the projected gradient flow of the objective function

  Minimize \( F(Q) := \frac{1}{2} \|Q^T \Lambda Q - N\|^2 \),
  
  Subject to \( Q^T Q = I \).

  ▶ Sorting is necessary in the first order optimality condition (Wielandt&Hoffman’53).

- Take a special \( N = \text{diag}\{n, n-1, \ldots, 2, 1\} \),
  
  ◦ \( X \) is tridiagonal and symmetric \( \implies \) Double bracket flow \( \equiv \) Toda lattice (Bloch’90).
    
    ▶ Bingo! The classical Toda lattice does have an objective function in mind.
  
  ◦ \( X \) is a general symmetric matrix \( \implies \) Double bracket = A specially scaled Toda lattice.

- Scaled Toda lattice (Chu’95),
  
  \[
  \frac{dX}{dt} = [X, K \circ X], \quad K = \text{fixed and skew-symmetric.}
  \]
  
  ◦ Flexible in componentwise scaling.
  
  ◦ Enjoy very general convergence behavior.
  
  ◦ But still no explicit objective function in sight.
Objective Functions

Some Flows on $\text{Orb}_{\mathcal{O}(m) \times \mathcal{O}(n)}(X)$ under Equivalence

- Any flow on the orbit $\text{Orb}_{\mathcal{O}(m) \times \mathcal{O}(n)}(X)$ under equivalence must be of the form

$$\frac{dX}{dt} = X(t)h(t) - k(t)X(t), \quad h(t) \in \mathbb{K}(n), \quad k(t) \in \mathbb{K}(m).$$

- QZ flow (Chu’86),

$$\frac{dX_1}{dt} = X_1 \Pi_0 (X_2^{-1}X_1) - \Pi_0 (X_1X_2^{-1})X_1,$$
$$\frac{dX_2}{dt} = X_2 \Pi_0 (X_2^{-1}X_1) - \Pi_0 (X_1X_2^{-1})X_2.$$  

- SVD flow (Chu’86),

$$\frac{dY}{dt} = Y \Pi_0 (Y(t)^t Y(t)) - \Pi_0 (Y(t)Y(t)^t) Y,$$
$$Y(0) = \text{bidiagonal}.$$

- The “objective” in the design of this flow was to maintain the bidiagonal structure of $Y(t)$ for all $t$.
- The flow gives rise to the Toda flows for $Y^t Y$ and $YY^t$. 


Projected Gradient Flows

- Given
  - A continuous matrix group $G \subset \mathcal{G}l(n)$.
  - A fixed $X \in \mathcal{V}$ where $\mathcal{V} \subset \mathbb{R}^{n \times n}$ be a subset of matrices.
  - A differentiable map $f: \mathcal{V} \rightarrow \mathbb{R}^{n\times n}$ with a certain “inherent” properties, e.g., symmetry, isospectrum, low rank, or other algebraic constraints.
  - A group action $\mu: G \times \mathcal{V} \rightarrow \mathcal{V}$.
  - A projection map $P$ from $\mathbb{R}^{n\times n}$ onto a singleton, a linear subspace, or an affine subspace $\mathbb{P} \subset \mathbb{R}^{n\times n}$ where matrices in $\mathbb{R}$ carry a certain desired structure, e.g., the canonical form.

- Consider the functional $F: G \rightarrow \mathbb{R}$

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

  - Want to minimize $F$ over $G$.

- Flow approach:
  - Compute $\nabla F(g)$.
  - Project $\nabla F(g)$ onto $T_g G$.
  - Follow the projected gradient until convergence.
Some Old Examples

- Brockett's double bracket flow (Brockett'88).
- Least squares approximation with spectral constraints (Chu&Driessel’90).
  \[
  \frac{dX}{dt} = [X, [X, P(X)]].
  \]

- Simultaneous reduction problem (Chu’91),
  \[
  \frac{dX_i}{dt} = \left[ X_i, \sum_{j=1}^{p} \frac{[X_j, P_j^T(X_j)] - [X_j, P_j^T(X_j)]^T}{2} \right]
  X_i(0) = A_i
  \]

- Nearest normal matrix problem (Chu’91),
  \[
  \frac{dW}{dt} = \left[ W, \frac{1}{2} \{ [W, \text{diag}(W^*)] - [W, \text{diag}(W^*)]^* \} \right]
  W(0) = A.
  \]
- Matrix with prescribed diagonal entries and spectrum (Schur-Horn Theorem) (Chu’95),

\[ \hat{X} = [X, [\text{diag}(X) - \text{diag}(a), X]] \]

- Inverse generalized eigenvalue problem for symmetric-definite pencil (Chu&Guo’98).

\[
\begin{align*}
\hat{X} &= -((XW)^T + XW), \\
\hat{Y} &= -((YW)^T + YW), \\
W &:= X(X - P_1(X)) + Y(Y - P_2(Y)).
\end{align*}
\]

- Various structured inverse eigenvalue problems (Chu&Golub’02).

- Remember the list of applications that Nicoletta gave on Monday!!!???
The idea of group actions, least squares, and the corresponding gradient flows can be generalized to other structures such as

- Stiefel manifold $\mathcal{O}(p, q) := \{Q \in \mathbb{R}^{p \times q} | Q^T Q = I_q\}$.
- The manifold of oblique matrices $\mathcal{OB}(n) := \{Q \in \mathbb{R}^{n \times n} | \text{diag}(Q^T Q) = I_n\}$.
- Cone of nonnegative matrices.
- Semigroups.
- Low rank approximation.

- Using the product topology to describe separate groups and actions might broaden the applications.
- Any advantages of using the isometry group over the orthogonal group?
Stochastic Inverse Eigenvalue Problem

- Construct a stochastic matrix with prescribed spectrum
  - A hard problem (Karpelevic’51, Minc’88).

\[ \begin{array}{cccc}
1 & 0.5 & 0 & -0.5 \\
0.5 & 1 & -0.5 & 0 \\
0 & -0.5 & 1 & 0.5 \\
-0.5 & 0 & 0.5 & 1
\end{array} \]

Figure 1: $\Theta_4$ by the Karpelević theorem.

- Would be done if the nonnegative inverse eigenvalue problem is solved – a long standing open question.
• Least squares formulation:

\[
\begin{align*}
\text{Minimize} \quad & F(g, R) := \frac{1}{2} \| g J g^{-1} - R \circ R \|_2^2 \\
\text{Subject to} \quad & g \in \text{Gl}(n), \ R \in \text{gl}(n).
\end{align*}
\]

\( J \) = Real matrix carrying spectral information.

\( \circ \) = Hadamard product.

• Steepest descent flow:

\[
\begin{align*}
\frac{dg}{dt} &= [(g J g^{-1})^T, \alpha(g, R)] g^{-T} \\
\frac{dR}{dt} &= 2\alpha(g, R) \circ R.
\end{align*}
\]

\( \alpha(g, R) := g J g^{-1} - R \circ R. \)
• ASVD flow for $g$ (Bunse-Gerstner et al’91, Wright’92):

$$
g(t) = X(t)S(t)Y(t)^T$$

$$
\dot{g} = \dot{X}SY^T + X\dot{S}Y^T + XS\dot{Y}^T
$$

$$
X^TgY = \begin{bmatrix} X^T & X & X \\ Y & Y & Y \end{bmatrix} \begin{bmatrix} \dot{X} \dot{S} \dot{Y} \end{bmatrix}
$$

Define $Q := X^TgY$. Then

$$
\frac{dS}{dt} = \text{diag}(Q).
$$

$$
\frac{dX}{dt} = XZ.
$$

$$
\frac{dY}{dt} = YW.
$$

$\diamond$ $Z,W$ are skew-symmetric matrices obtainable from $Q$ and $S$. 
Nonnegative Matrix Factorization

- For various applications, given a nonnegative matrix $A \in \mathbb{R}^{m \times n}$, want to
  $$\min_{0 \leq V \in \mathbb{R}^{m \times k}, 0 \leq H \in \mathbb{R}^{k \times n}} \frac{1}{2} \| A - VH \|_F^2.$$

  ◦ Relatively new techniques for dimension reduction applications.
    - Image processing — no negative pixel values.
    - Data mining — no negative frequencies.
  ◦ No firm theoretical foundation available yet (Tropp'03).

- Relatively easy by flow approach!
  $$\min_{E \in \mathbb{R}^{m \times k}, F \in \mathbb{R}^{k \times n}} \frac{1}{2} \| A - (E \circ E)(F \circ F) \|_F^2.$$

- Gradient flow:
  $$\frac{dV}{dt} = V \circ (A - VH)H^\top,$$
  $$\frac{dH}{dt} = H \circ (V^\top(A - VH)).$$

  ◦ Once any entry of either $V$ or $H$ hits 0, it stays zero. This is a natural barrier!
  ◦ The first order optimality condition is clear.
Image Articulation Library

- Assume images are composite objects in many articulations and poses.
- Factorization would enable the identification and classification of intrinsic “parts” that make up the object being imaged by multiple observations.
- Each column $a_j$ of a nonnegative matrix $A$ now represents $m$ pixel values of one image.
- The columns $v_k$ of $V$ are $k$ basis elements in $\mathbb{R}^m$.
- The columns of $H$, belonging to $\mathbb{R}^k$, can be thought of as coefficient sequences representing the $n$ images in the basis elements.
\[ A \in \mathbb{R}^{19200 \times 10} \] Representing 10 Gray-scale 120 \times 160 Irises
Basis Irises with $k = 2$
(Wrong?) Basis Irises with $k = 4$
Conclusion

- Many operations used to transform matrices can be considered as matrix group actions.

- The view unifies different transformations under the same framework of tracing orbits associated with corresponding group actions.
  - More sophisticated actions can be composed that might offer the design of new numerical algorithms.
  - As a special case of Lie groups, (tangent space) structure of a matrix group is the same at every of its element. Computation is easy and cheap.

- It is yet to be determined how a dynamical system should be defined over a group so as to locate the simplest form.
  - The notion of “simplicity” varies according to the applications.
  - Various objective functions should be used to control the dynamical systems.
  - Usually offers a global method for solving the underlying problem.

- Continuous realization methods often enable to tackle existence problems that are seemingly impossible to be solved by conventional discrete methods.

- Group actions together with properly formulated objective functions can offer a channel to tackle various classical or new and challenging problems.
Some basic ideas and examples have been outlined in this talk.

- More sophisticated actions can be composed that might offer the design of new numerical algorithms.
- The list of application continues to grow.

New computational techniques for structured dynamical systems on matrix group will further extend and benefit the scope of this interesting topic.

- Need ODE techniques specially tailored for gradient flows.
- Need ODE techniques suitable for very large-scale dynamical systems.
- Help! Help! Help!