Proceedings of the American Control Conference
San Francisco, California • June 1993



# Inductive Inference of Invariant Subspaces

Michael Lemmon *
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

**Abstract**

This paper shows that inductive inference protocols can learn invariant linear subspaces, used in the stabilization of variable structure systems, after a finite number of failed oracle queries. It is further shown that this convergence bound scales in a polynomial manner with the system's state space dimension.

## 1 Introduction

This paper is concerned with the behaviour of dynamical systems represented by the following differential equations,

$$\dot{\mathbf{x}} = \begin{cases} f^-(\mathbf{x}) & \text{if } \mathbf{s}'\mathbf{x} > 0 \\ 0 & \text{if } \mathbf{s}'\mathbf{x} = 0 \\ f^+(\mathbf{x}) & \text{if } \mathbf{s}'\mathbf{x} < 0 \end{cases}, \qquad (1)$$

where $\mathbf{x} \in \Re^n$ is the state vector and $\mathbf{s}$ is an $n$-dimensional vector characterizing a hyperplane in $\Re^n$ called the **switching surface**. The mappings, $f^+$ and $f^-$, are smooth mappings from $\Re^n$ onto $\Re^n$. The system in equation 1 is a variable structure system (VSS) [7] because the structure of the system's vector field changes discontinuously across a hyperplane defined by $\mathbf{s}'\mathbf{x} = 0$. The problem addressed by this paper concerns the determination of $(n-1)$-dimensional subspaces of $\Re^n$ which are attracting invariant sets with respect to the flow generated by equation 1.

The subspaces under consideration are linear spaces of the form $H_{\mathbf{s}} = \{\mathbf{x} : \mathbf{s}'\mathbf{x} = 0\}$ This set forms an $(n-1)$-dimensional hyperplane which is characterized by the $n$-dimensional vector $\mathbf{s}$. If $H_{\mathbf{s}}$ is an attracting invariant set of the flow, then all state trajectories must eventually be captured by this set. Switching surfaces which are also attracting invariant sets will be called sliding modes. A variable structure system of the form shown in equation 1 will be said to have been stabilized with respect to $H_{\mathbf{s}}$ if $H_{\mathbf{s}}$ is a sliding mode.

If the set of vector fields $\{f^+, f^-\}$ are already known, then the invariant sets can be computed directly. This paper, however, focuses on situations where the vector fields are unknown. In such situations, it is necessary that the invariant subspaces be determined directly from the system's observed behaviour. In other words, this paper is concerned with the identification of system invariants on the basis of observed behaviour. For this reason, the problem considered in this paper will be referred to as the invariant subspace identification (ISID) problem.

The technique used in this paper for solving the ISID problem is inductive inference [1]. This method was motivated by the simple observation that the problem of determining invariant subspaces bears some similarity to the problem of iteratively training linear classifiers. Since there exist well known finite-time algorithms [5] for training such classifiers, it was conjectured such techniques might yield direct adaptive controllers with provable finite time convergence.

The following sections substantiate that conjecture. Section 2 states the algorithm and section 3 derives specific components of the algorithm. Section 4 proves the paper's principal results concerning finite time convergence. An example of the algorithm's application is presented in section 5. Section 6 summarizes this paper's results and indicates future directions.

## 2 Invariant Subspace Identification Algorithm

Inductive inference is a machine learning protocol in which a system learns by example [1]. The inductive protocol developed in this paper consists of four fundamental components.

- Hypothesis: The hypothesis is characterized by a symmetric matrix $\mathbf{Q} \in \Re^{n\times n}$ which defines an ellipsoidal cone, $C(\mathbf{Q})$, defined as follows, $C(\mathbf{Q}) = \{\mathbf{s} \in \Re^n : \mathbf{s}'\mathbf{Q}\mathbf{s} < 0\}$ The matrix $\mathbf{Q}$ has $n-1$ positive eigevalues and 1 negative eigenvalue ordered. Assume that the eigenvalues, $\lambda_i$ are ordered so that $\lambda_i > \lambda_{i+1}$. Let $\mathbf{e}_i$ be the eigenvector associated with eigenvalue $\lambda_i$. The normalized eigenvalue matrix is denoted as $\mathbf{R} = \text{diag}(\lambda_1, \ldots, \lambda_{n-1})/|\lambda_n|$ and the positive eigenvector matrix $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_{n-1}) \in \Re^{n\times n-1}$.

*I would like to acknowledge the partial financial support of the National Science Foundation grant IRI-91-09298

The hypothesis is that the cone $C(\mathbf{Q})$ contains the normal vectors to all sliding modes and that the negative eigenvalue of $\mathbf{Q}$ is a sliding mode.

- Experiment: The algorithm's next major component is an "experiment" for measuring the system's current state.
- Oracle Query: The third component is an algorithm called the oracle. This component uses the experimentally gathered data and makes a declaration concerning the consistency of that data with the hypothesis. The declaration is made by a Boolean functional called the invariance oracle.
- Update: The oracle's repsonse is a MAYBE or FALSE declaration. If the answer is MAYBE then nothing is done. If the answer is FALSE, however, the current hypothesis is modified by the update algorithm. The update algorithm used in this paper is a modification of the central-cut ellipsoid method [6] which recomputes the symmetric matrix $\mathbf{Q}$. In modifying the hypothesis, the update procedure attempts to generate a new hypothesis which is consistent with prior experimental data. This basic cycle of experiment, query, and update continues until an attracting invariant subspace is found.

With the preceding outline and establishment of notational conventions, the ISID algorithm can now be formally stated. Specific components of this algorithm are derived in following section.

1. **Initialize:** Initialize an $n$ by $n$ symmetric matrix, $\mathbf{Q}$, which has $n-1$ positive eigenvalues and 1 negative eigenvalue such that if $H_{\mathbf{z}}$ is a sliding mode, then $\mathbf{z}'Q\mathbf{z} < 0$.
2. **Form Hypothesis:** Compute the eigendecomposition of $\mathbf{Q}$ to obtain matrices $\mathbf{R}$ and $\mathbf{E}$. Set the system's current switching surface, $\mathbf{s}$, equal to the negative eigenvector, $\mathbf{e}_n$, of $\mathbf{Q}$.
3. **Experiment:** Measure the system's state and state velocity, $\mathbf{x}$ and $\dot{\mathbf{x}}$.
4. **Query:** Compute the invariance oracle's response,
$$I_1(\mathbf{x}, \dot{\mathbf{x}}, \mathbf{s}) = \begin{cases} 0 & \text{if } (\mathbf{s}'\mathbf{x})(\mathbf{s}'\dot{\mathbf{x}}) < 0 \\ 1 & \text{otherwise} \end{cases} \qquad (2)$$
5. **Update Hypothesis:** If the oracle returns 1, then recompute $\mathbf{Q}$ using the following equations,
$$\mathbf{c} = \mathrm{sgn}(\mathbf{e}_n'\mathbf{x})\mathbf{E}'\dot{\mathbf{x}}, \qquad (3)$$
$$\mathbf{b} = \frac{\mathbf{R}^{-1}\mathbf{c}}{\sqrt{\mathbf{c}'\mathbf{R}^{-1}\mathbf{c}}}, \qquad (4)$$
$$\mathbf{a} = -\frac{1}{n}\mathbf{b}, \qquad (5)$$
$$\overline{\mathbf{R}}^{-1} = \frac{(n-1)^2}{(n-1)^2-1}\left(\mathbf{R}^{-1} - \frac{2}{n}\mathbf{b}\mathbf{b}'\right), \qquad (6)$$
$$\mathbf{x}_a = \mathbf{E}\mathbf{a} + \mathbf{e}_n, \qquad (7)$$
$$\overline{\mathbf{Q}} = (\mathbf{I} - \mathbf{e}_n\mathbf{x}_a')\mathbf{E}\overline{\mathbf{R}}\mathbf{E}'(\mathbf{I} - \mathbf{x}_a\mathbf{e}_n'), \qquad (8)$$
Set $\mathbf{Q}$ equal to $\overline{\mathbf{Q}}$.
6. If the oracle returns 0, then do nothing.
7. **Loop:** go to step 2.

## 3 Algorithm Components

The invariance oracle will be a Boolean functional which declares whether or not a given subspace, $H_{\mathbf{s}}$, is attracting and $\Phi$-invariant. This test is based on testing a Lyapunov inequality. The following theorem states the desired inequality.

**Theorem 1** *Let $\mathbf{s} \in \Re^n$ and let $\dot{\mathbf{x}}$ be given by equation 1. If for all $\mathbf{x} \notin H_{\mathbf{s}}$,*
$$(\mathbf{s}'\mathbf{x})(\mathbf{s}'\dot{\mathbf{x}}) < 0, \qquad (9)$$
*then the subspace, $H_{\mathbf{s}}$, is an attracting $\Phi$-invariant set.*

**Proof:** See theorem 8 in [7].

Equation 9 can be recast as a logical function making a declaration about the consistency of the measured state and state velocity with the hypothesis that $H_{\mathbf{s}}$ is a sliding mode. This motivates the following definition for an "invariance" oracle.

**Definition 1** *The Boolean functional, $I_1 : \Re^{3n} \to \{0, 1\}$, defined by equation 2 will be called an* **invariance oracle.**

Let the set $A_1$ denote the sets of attracting invariant subspaces which are declarable by the oracle, $I_1$. Let the set of all attracting invariant subspaces be denoted as $A$. Clearly, $A_1 \subset A$ so that the declarations (0/1) made by the oracle can be given the semantic interpretations of MAYBE/FALSE.

Note that the set of $\mathbf{s}$ determined by equation 9 forms a pair of halfspaces forming a convex cone in $\Re^n$. Since $A_1$ is the intersection of all these cones, the following lemma can be immediately deduced.

**Lemma 1** *$A_1$ is a convex cone centered at the origin.*

The significance of the preceding lemma is that it suggests $A_1$ may be well approximated by ellipsoidal cones, $C(\mathbf{Q})$. Since the preceding lemma implies that $A_1$ is a convex set, then if it is also bounded, there exists a unique ellipsoid of minimal volume which contains $A_1$ [3].

If the oracle declares the current experimental data to be inconsistent with the hypothesis, then the following lemma provides a characterization of the subspaces inconsistent with the hypothesis characterized by $\mathbf{Q}$.

**Lemma 2** *Let $C(\mathbf{Q})$ be an ellipsoidal cone with negative eigenvector, $\mathbf{e}_n$. Let $\mathcal{X}$ be a data collection for which the invariance oracle, $I_1$, declares a failure, $I_1(\mathcal{X}, \mathbf{e}_n) = 1$. If $A_1 \subset C(\mathbf{Q})$, then $A \subset C(\mathbf{Q}) \cap H(\mathcal{X}, \mathbf{e}_n)$ where*

$$H(\mathcal{X}, \mathbf{e}_n) = \{\mathbf{s} \in \Re^n : \mathbf{s}'\dot{\mathbf{x}} < \mathrm{sgn}(\mathbf{e}_n'\mathbf{x})\mathbf{e}_n'\dot{\mathbf{x}}\}. \quad (10)$$

*The set $H(\mathcal{X}, \mathbf{e}_n)$ will be called the* <u>*inconsistent set*</u> *generated by $\mathcal{X}$.*

**Proof:** If a perfect invariance oracle $I_1$ returns 1 for $\mathcal{X}$ given the subspace represented by $\mathbf{e}_n$, then the following inequality holds. $(\mathbf{e}_n'\dot{\mathbf{x}})(\mathbf{e}_n'\mathbf{x}) > 0$. Note that for all $\mathbf{z}$ such that $\mathbf{z}'\dot{\mathbf{x}} \geq \mathbf{e}_n'\dot{\mathbf{x}}$, it can be inferred by the comparison principle that $\mathbf{z}'\mathbf{x} \geq \mathbf{e}_n'\mathbf{x}$. Similar arguments apply if the inequalities are reversed. Therefore any subspace, $H_{\mathbf{z}}$, such that

$$\mathbf{z}'\dot{\mathbf{x}} \geq \mathrm{sgn}(\mathbf{e}_n'\mathbf{x})\mathbf{e}_n'\dot{\mathbf{x}}, \quad (11)$$

cannot possibly be an attracting invariant set. Using the halfspace defined by the preceding inequality yields equation 10. **QED**.

**Theorem 2** *Let $C(\mathbf{Q})$ be an ellipsoidal cone with negative eigenvector $\mathbf{e}_n$ such that $A_1 \subset C(\mathbf{Q})$. Let $\mathcal{X}$ be a data collection for which $I_1(\mathcal{X}, \mathbf{e}_n) = 1$. There exist ellipsoidal cones, $C(\underline{\mathbf{Q}})$ and $C(\overline{\mathbf{Q}})$, such that $C(\underline{\mathbf{Q}}) \subset H(\mathcal{X}, \mathbf{e}_n) \cap C(\mathbf{Q}) \subset C(\overline{\mathbf{Q}})$. Furthermore $\overline{\mathbf{Q}}$ is given by equations 3 through 8.*

**Proof:** Let $S$ be an $n-1$-dimensional subspace of $\Re^n$ and let $\mathbf{x} \in \Re^n$. The <u>linear variety</u> of $S$ generated by $\mathbf{x}$ is the set $V(S, \mathbf{x}) = \{\mathbf{s} + \mathbf{x} : \mathbf{s} \in S\}$ From the above definition it can be easily shown that the intersection of $C(\mathbf{Q})$ with $V(\mathrm{sp}(\mathbf{E}), \mathbf{e}_n)$ is an ellipsoid of the following form $E(\mathbf{R}^{-1}, 0) = \{\mathbf{w} \in \Re^{n-1} : \mathbf{w}'\mathbf{R}\mathbf{w} < 1\}$ . where spE is the span of the eigenvectors of $\mathbf{Q}$ with positive eigenvalues ($\mathbf{E}$ is called the positive eigenvalue matrix), $\mathbf{R} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n-1})/|\lambda_n|$, and $\lambda_i$ are the eigenvalues of $\mathbf{Q}$ in decreasing order.

The intersection of the inconsistent set, $H(\mathcal{X}, \mathbf{e}_n)$, and the linear variety, $V(\mathrm{sp}(\mathbf{E}), \mathbf{e}_n)$, can be shown to form an $n-1$-dimensional halfspace, $H$, given by $H = \{\mathbf{w} \in \Re^{n-1} : \mathbf{w}'\mathbf{c} < 0\}$ , where $\mathbf{c} = \mathrm{sgn}(\mathbf{e}_n'\mathbf{x})\mathbf{E}'\dot{\mathbf{x}}$. This is shown by noting that any vector $\mathbf{z}$ in the linear variety can be written as $\mathbf{z} = \mathbf{E}\mathbf{w} + \mathbf{e}_n$. Inserting this into inequality 11 yields the above halfspace.

The above remarks therefore suggest that the intersection of $C(\mathbf{Q})$, $V(\mathrm{sp}(\mathbf{E}), \mathbf{e}_n)$, and $H(\mathcal{X}, \mathbf{e}_n)$ will be an $n-1$-dimensional convex body, $K$. It is well known [3] that any bounded convex body can be contained within a unique ellipsoid of minimal volume called the Lowner-John ellipsoid. Computing the Lowner-John ellipsoid [3] [6] for $K$ will yield the ellipsoid $E(\overline{\mathbf{R}}^{-1}, \mathbf{a})$ where $\overline{\mathbf{R}}$ and $\mathbf{a}$ are as given in the theorem (see equations 3, 4, 5, and 6).

The $n-1$-dimensional Lowner-John ellipsoid computed in equations 3 through 6 can be extended into an $n$-dimensional ellipsoidal cone in the following way. Let $\mathbf{s}$ be any point in the cone generated by the ellipsoid $E(\overline{\mathbf{R}}^{-1}, \mathbf{a})$. There exists an $\alpha \in \Re$ such that $\alpha\mathbf{s}$ is in the linear variety, $V(\mathrm{sp}(\mathbf{E}), \mathbf{e}_n)$. The $\alpha$ for which this is true must satisfy the orthogonality condition,

$$\begin{aligned} 0 &= \mathbf{e}_n'(\alpha\mathbf{s} - \mathbf{e}_n) \quad &(12)\\ &= \alpha\mathbf{e}_n'\mathbf{s} - 1, \quad &(13)\end{aligned}$$

which implies that $\alpha = 1/\mathbf{e}_n'\mathbf{s}$.

Since, $\mathbf{s} = \mathbf{E}\mathbf{w} + \mathbf{e}_n$, the ellipsoid equation for $E(\overline{\mathbf{R}}^{-1}, \mathbf{a})$ is

$$\begin{aligned} 1 &> (\mathbf{w} - \mathbf{a})'\overline{\mathbf{R}}(\mathbf{w} - \mathbf{a}) \quad &(14)\\ &> (\mathbf{s} - \mathbf{x}_a)'\mathbf{E}'\overline{\mathbf{R}}\mathbf{E}(\mathbf{s} - \mathbf{x}_a), \quad &(15)\end{aligned}$$

where $\mathbf{x}_a = \mathbf{E}\mathbf{a} + \mathbf{e}_n$. The vector $\mathbf{s}$ in this equation must, of course, lie in the linear variety generated by $\mathbf{e}_n$, $V(\mathrm{sp}(\mathbf{E}), \mathbf{e}_n)$. From our preceding discussion, any vector in the cone can be pulled back to the variety by appropriate renormalization with $\alpha$. This then implies that if $\mathbf{s}$ is any vector in the cone, then

$$\left(\frac{\mathbf{s}}{\mathbf{s}'\mathbf{e_n}} - \mathbf{x}_a\right)'\mathbf{E}'\overline{\mathbf{R}}\mathbf{E}\left(\frac{\mathbf{s}}{\mathbf{e_n}'\mathbf{s}} - \mathbf{x}_a\right) < 1. \quad (16)$$

Multiplying through by $|\mathbf{s}'\mathbf{e}_n|^2$, we obtain

$$\mathbf{s}'\left[(\mathbf{I} - \mathbf{e}_n\mathbf{x}_a')\mathbf{E}'\overline{\mathbf{R}}\mathbf{E}(\mathbf{I} - \mathbf{x}_a\mathbf{e}_n') - \mathbf{e}_n\mathbf{e}_n'\right]\mathbf{s} < 0. \quad (17)$$

This inequality determines an ellipsoidal cone and the term within the square brackets is $\overline{\mathbf{Q}}$ as computed in equation 7 and 8.

$\underline{\mathbf{Q}}$ is obtained by noting that if $E(\overline{\mathbf{R}}^{-1}, \mathbf{a})$ is a Lowner-John ellipsoid for $K$, then $E(\overline{\mathbf{R}}^{-1}/(n-1)^2, \mathbf{a})$ is an ellipsoid contained within $K$ [3]. The preceding construction is then repeated with this smaller ellipsoid. **QED**

## 4 Convergence and Complexity

Establishing the convergence results of this section requires some way of measuring a ellipsoidal cone's volume.

**Definition 2** *Let $C(\mathbf{Q})$ be an ellipsoidal cone and let the eigenvalues of $\mathbf{Q}$ be ordered as $\lambda_i > \lambda_{i+1}$. The* **volume** *of cone $C(\mathbf{Q})$ is defined to be* $\mathrm{vol}C(\mathbf{Q}) = \sqrt{\prod_{i=1}^{n-1}(|\lambda_n|/\lambda_i)}$ .

The preceding definition is using the volume of the $n-1$-dimensional ellipsoid contained in the linear variety $V(\mathrm{sp}(\mathbf{E}), \mathbf{e}_n)$ as the "volume" of the cone.

The following theorem shows that the ISID algorithm must locate an attracting invariant subspace after a finite number of failed queries to a perfect invariance oracle. A perfect oracle is an oracle which never makes a mistake in its declaration.

**Theorem 3** *Initialize the ISID algorithm with an ellipsoidal cone whose volume is unity and which is known to contain $A_1$. Let $\epsilon$ denote the volume of the smallest ellipsoidal cone containing $A_1$. If $n$ is the state space dimension, then the ISID algorithm will determine an attracting invariant subspace after no more than $2(n-1)\ln\epsilon^{-1}$ failed queries to a perfect invariance oracle.*

**Proof:** Consider the ellipsoidal cone $C(\mathbf{Q}_i)$ after the $i$th failed invariance test. Let $\mathbf{E}$ and $\mathbf{L}$ be the positive eigenvector and eigenvalue matrices of $\mathbf{Q}_i$, respectively. The volume of this ellipsoid will be given by $\mathrm{vol}C(\mathbf{Q}_i) = \sqrt{\frac{1}{\prod_{j=1}^{n-1}\lambda_j(\mathbf{R})}}$ , where $\lambda_j(\mathbf{R})$ is the $j$th positive eigenvalue of $\mathbf{R}$ and $\mathbf{R} = \mathbf{L}/|\lambda_n|$. Consider the ellipsoidal cone obtained using equations 3 through 8. The symmetric matrix characterizing this cone is $\overline{\mathbf{Q}} = \mathbf{X}'\mathbf{Y}\mathbf{X}$ where

$$\mathbf{X} = \begin{pmatrix} \mathbf{E}'(\mathbf{I}-\beta\mathbf{e}_a\mathbf{e}_n') \\ \mathbf{e}_n' \end{pmatrix}, \qquad (18)$$

$$\mathbf{Y} = \begin{pmatrix} \mathbf{R} & 0 \\ 0 & -1 \end{pmatrix}. \qquad (19)$$

where $\beta = ||\mathbf{x_a}||$. Applying the orthogonal transformation,

$$\mathbf{P} = \begin{pmatrix} \mathbf{E} & \mathbf{e}_n \end{pmatrix}, \qquad (20)$$

to $\mathbf{X}$, yields

$$\mathbf{P}'\mathbf{X}' = \begin{pmatrix} \mathbf{I} & 0 \\ -\beta\mathbf{e}_a'\mathbf{E} & 1 \end{pmatrix} \qquad (21)$$

where $\beta = ||\mathbf{x}_a||$ and $\beta\mathbf{e}_a = \mathbf{x}_a$. Recall that $\mathbf{x}_a$ is the center of the updated ellipsoid in the linear variety $V(\mathrm{sp}(\mathbf{E}), \mathbf{e}_n)$. For convenience, let $\mathbf{v}' = -\beta\mathbf{e}_a'E$.

Since the eigenvalues of $\overline{\mathbf{Q}}$ are unchanged by an orthogonal transformation, the eigenvalues of $\mathbf{P}'\mathbf{X}'\mathbf{Y}\mathbf{X}\mathbf{P}$ can be used to compute the volume of $\overline{\mathbf{Q}}$. This transformed matrix has the form

$$\mathbf{P}'\overline{\mathbf{Q}}\mathbf{P} = \mathbf{P}'\mathbf{X}'\mathbf{Y}\mathbf{X}\mathbf{P} \qquad (22)$$

$$= \begin{pmatrix} \overline{\mathbf{R}} & \overline{\mathbf{R}}\mathbf{v} \\ \mathbf{v}'\overline{\mathbf{R}} & \mathbf{v}'\overline{\mathbf{R}}\mathbf{v}-1 \end{pmatrix}. \qquad (23)$$

Note that $\overline{\mathbf{R}}$ is an $n-1$ by $n-1$ leading principal submatrix of $\mathbf{P}'\overline{\mathbf{Q}}\mathbf{P}$, the eigenvalues of the two matrices satisfy the interlacing property, $\lambda_{i+1}(\overline{\mathbf{Q}}) \leq \lambda_i(\overline{\mathbf{R}}) \leq \lambda_i(\overline{\mathbf{Q}})$ for $i = 1,\ldots,n-1$ [2].

Since $\lambda_n(\overline{\mathbf{Q}})$ is negative, it can be shown that $\lambda_n(\mathbf{P}'\mathbf{X}'\mathbf{Y}\mathbf{X}\mathbf{P}) \leq \sigma_n^2(\mathbf{P}'\mathbf{X}')\lambda_n(\mathbf{Y})$ , where $\sigma_n(\mathbf{P}'\mathbf{X}')$ is the smallest singular value of $\mathbf{P}'\mathbf{X}'$ and $\lambda_n(\mathbf{Y})$ is the negative eigenvalue of $\mathbf{Y}$ [2]. Note that this eigenvalue must be negative one (by construction of $\mathbf{Y}$). Also note that the singular value must satisfy the following inequality for any $\mathbf{x} \in \Re^n$,

$$\sigma_n^2(\mathbf{P}'\mathbf{X}') \leq \frac{\mathbf{x}'\mathbf{P}'\mathbf{X}'\mathbf{X}\mathbf{P}\mathbf{x}}{\mathbf{x}'\mathbf{x}}. \qquad (24)$$

In particular, if we let $\mathbf{x} = (0\cdots01)'$, then the smallest singular value must be less than unity. It can therefore be concluded that $|\lambda_n(\overline{\mathbf{Q}})| < 1$.

With the preceding observations about the interlaced eigenvalues, it can be concluded that

$$\mathrm{vol}C(\overline{\mathbf{Q}}) = \sqrt{\prod_{j=1}^{n-1}\frac{|\lambda_n(\overline{\mathbf{Q}})|}{\lambda_j(\overline{\mathbf{Q}})}} \qquad (25)$$

$$\leq \sqrt{\prod_{j=1}^{n-1}\frac{1}{\lambda_j(\overline{\mathbf{R}})}} \qquad (26)$$

$$\leq e^{-\frac{1}{2(n-1)}}\mathrm{vol}C(\mathbf{Q}). \qquad (27)$$

Inequality 26 is a consequence of the bound on the absolute value of the negative eigenvalue as well as the interlacing property. Inequality 27 is a consequence of a well-known relationship on the quotient of ellipsoid volumes obtained using the central-cut ellipsoid method [3].

Since the initial ellipsoidal cone's volume is unity, the ellipsoidal cone's volume after the $L$th failed query must be bounded as follows, $\mathrm{vol}C(\mathbf{Q}_L) \leq \exp(-L/(2n-2))$ . However, $C(\mathbf{Q}_L)$ cannot be smaller than $\epsilon$ by assumption, therefore the number of failed queries, $L$, must satisfy $\epsilon \leq \exp(-L/(2n-2))$ . Rearranging this inequality to extract $L$ shows that the number of failed invariance queries can be no larger than the bound stated by the theorem. **QED**

The following corollary for the preceding theorem establishes the polynomial oracle-time complexity of the ISID algorithm.

**Corollary 1** *Assume that $A_1$ is a set which is contained within an ellipsoidal cone characterized by a matrix, $\mathbf{Q}$, whose normalized positive eigenvalues satisfy the inequality $\frac{|\lambda_n|}{\lambda_i} < \gamma$ for $1 > \gamma > 0$ and $i = 1,\ldots,n-1$. Under the assumptions of theorem 3, the ISID algorithm will determine an attracting invariant subspace after no more than $2(n-1)^2\ln(n-1) + (n-1)^2\ln\gamma^{-1}$ MAYBE declarations by the invariance oracle.*

**Proof:** Because of the constraints on $\mathbf{Q}$, the volume of the smallest bounding ellipsoid will be no greater than $\gamma^{(n-1)/2}(n-1)^{-n+1}$. The assumed comonstraints on $\lambda_i$ yield the first term. The second term comes from the fact that the volume of an $n$-d sphere with unit radius can be no smaller than $n^{-n}$. Inserting this into the bound of theorem 3 yields the asserted result. **QED**

## 5 Example: AUV Stabilization

The following simulation results illustrate how the ISID algorithm can quickly stabilize an AUV's dive plane dynamics. The simplified equations of motion for vehicle (pitch) angle of attack, $\theta$, in the dive plane as a function of velocity, $v$, may be written as

$$\ddot{\theta} = K_1\dot{\theta} + K_2\theta|\theta| + K_3\theta|v| + u_\theta, \qquad (28)$$
$$\dot{v} = -v + K_4|\theta|v + u_v, \qquad (29)$$

where $K_1$, $K_2$, $K_3$, and $K_4$ are hydrodynamic force coefficients. $u_v$ and $u_\theta$ represent control forces applied in the velocity and angle of attack channels, respectively. These equations clearly show how nonlinearities enter the dynamics through the hydrodynamic cross coupling between $\theta$ and $v$. Uncertainty arises from the simple fact that the hydrodynamic coefficients may be poorly known.

In the following simulations, a hierarchical variable structure controller with boundary layer was designed. The control hierarchy was designed so that the system nulls angle of attack prior to nulling commanded velocity errors.

Figure 1 shows the AUV's performance with the hierarchical sliding mode controller after a system failure causes the initially chosen switching surfaces to no longer be invariant sets. As can be seen, the sliding controller is actually unstable with the system exhibiting large oscillations in $\theta$.

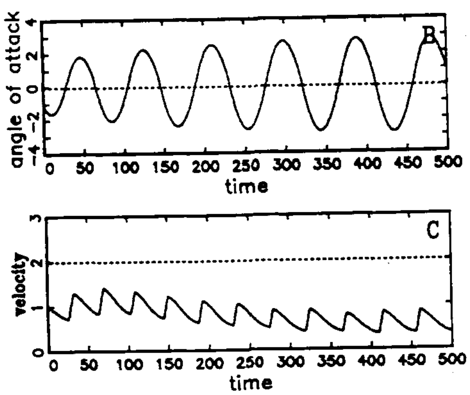


Figure 1: Simulated AUV dive with hierarchical sliding control in which sliding mode constraints are violated. Angle of attack, $\theta$, time history and velocity, $v$, time history.

Figures 2 shows the system's behaviour after two "learning" sessions with the ISID algorithm. A learning session involves starting the vehicle at the initial condition and then commanding it over to the desired state. In this case, it is clear that learning is complete. There are no readjustments of the sliding surface and the system wastes little effort in bringing the system to its commanded state.

Perhaps the most remarkable thing about this example is the apparent speed with which the sliding surface is learned. In these simulations, only 4 failed invariance tests were required before finding a sliding mode. This low number of failed tests was observed in other simulation runs where the system's initial conditions were randomly varied.

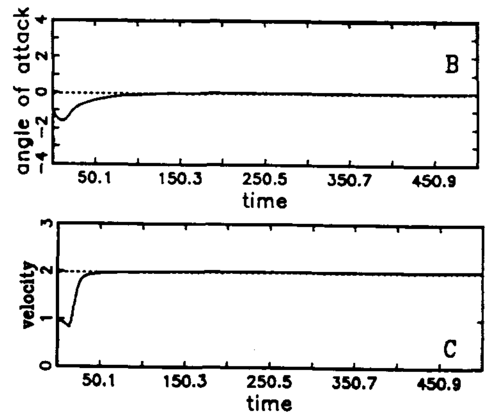


Figure 2: Simulated AUV dive where ISID algorithm is used to relearn hierarchical sliding mode controller (After 2 learning sessions). Angle of attack, $\theta$, time history and velocity, $v$, time history.

## 6 Summary

This paper has derived an inductive inference protocol for the on-line identification of sliding modes for a variable structure system. In that the sliding modes stabilize the system, the proposed inductive inference protocol is therefore a direct adaptive control algorithm. The principal result of this paper proves that the proposed protocol will find an attracting invariant set after a finite number of failed queries to a perfect invariance oracle. Under realistic assumptions, this finite time convergence bound is shown to scale as $O((n-1)^2 \ln(n-1)) \approx O((n-1)^{2.5})$ where $n$ is the state space dimension.

## References


[1] D. Angluin, C.H. Smith, *Computing Surveys*, 15(3):237-269, September 1983.

[2] G. Golub, C. Van Loan (1983), *Matrix Computation*, Johns Hopkins University Press, Baltimore, Maryland, 1983

[3] Groetshel, Lovasz, and Schrijver (1988), *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag, 1988.

[4] L.G. Khachiyan (1979), *Soviet Mathematics Doklady*, 20:191-194, 1979.

[5] F. Rosenblatt (1962), *Principles of Neurodynamics*, Spartan books, Washington D.C.

[6] N.Z. Shor (1977), *Cybernetics*, 13:94-96, 1977.

[7] V.I. Utkin (1977), *IEEE Transactions on Automatic Control*, Vol. AC-22:212-222.