Learning Retrodictive Knowledge from Scientific Laws: The Case of Chemical Kinetics

Raúl E. Valdés-Pérez

August 1989

CMU-CS-89-179

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

We consider the problem of extracting retrodictive knowledge from scientific laws used ordinarily only to predict. In particular, a method is developed which synthesizes rules of experiment-interpretation from the basic law of chemical kinetics.

Previous work in AI on transforming predictive knowledge into convenient retrodictive knowledge has been within the subfield of diagnosis. The current work extends the idea to the domain of elucidation of causal mechanism.

Refutation rules are synthesized by discovering invariants within a parameterized system of equations. The choice of invariants to look for is guided by four criteria. A principle of stable refutation, based on the character of experimental data, is derived from the non-rescindible nature of refutation. Three other criteria contribute to the practicality, generality, and reliability of the rules.

The invariants chosen are tested by systematic sampling of a system parameter-space. Hence, the rules, which check that an invariant holds for experimental data, are established by induction from simulation data.

The synthesized rules serve in practice as reliable disconfirmatory evidence, rather than refutations, due to their inductive origin as well as to the uncertainty of experimental data. The rules will be applied within the context of ongoing work on elucidation of chemical-reaction networks.

The author is supported by a Graduate Fellowship from the Avionics Laboratory at Wright-Patterson Air Force Base, awarded by Universal Energy Systems. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. government or Universal Energy Systems.

1. Introduction

How can one extract convenient *retrodictive* knowledge from scientific laws used ordinarily only to predict? In particular, given laws that characterize observable manifestations of a physical mechanism, how can one turn those laws to use for the purpose of elucidating the structure of that mechanism?

This report addresses these questions in the context of the scientific laws of chemical kinetics. Parsimonious rules that interpret data from chemical-reaction experiments are synthesized automatically from the kinetical equations. These rules could be used as refutations of mechanistic hypotheses in the case of noiseless data, and as reliable contrary evidence in the case of noiseful data. For example, on the hypothetical set of concurrent reactions:

$$\begin{aligned} A + B &\rightarrow T + X \\ B + X &\rightarrow 2T \end{aligned}$$

the following rule was synthesized:

> If the concentration of B exceeds that of A at any time, refute (or disconfirm) the hypothesis.

Our method involves:

- Discovery of invariants within a parameterized system of ordinary, non-linear, differential equations (the kinetical equations).
- Development of several criteria to guide the choice of the invariants to look for. One criterion is derived from the non-rescindibility of refutation.
- Sampling the parameter space of the kinetical equations, followed by numerical integration to obtain the solution, as a means for discovery of invariants. Our chosen invariants are not generally inferrable analytically directly from the kinetical equations.
- Formulation of experiment-interpretation rules based on the invariants, which test experimental data for violation of particular invariants. A violation serves to refute (or disconfirm) any hypothesis for which the invariant holds.

The sequel is organized as follows. Our motivation and context are discussed in the next section. Then we sketch some chemistry knowledge helpful to understand the role of the equations in this report. We next present related ideas from structural diagnosis (of satellite subsystems, linear accelerators, and the human heart). Then, our method is developed in detail on the equations of chemical kinetics, and practical use of the synthesized experiment-interpretation rules is discussed. Finally, the last section recapitulates the present contribution, and points out the limitations. The appendices develop a technical point useful to reduce computation time, give further examples of program output, and relate the kinetical equations to theorems of differential equations.

2. Motivation and Context

Our larger motivation is to automate the elucidation of reaction networks of simple to moderate complexity. The work described by this report contributes one source of knowledge, or constraint,

to be used for elucidation.

The chemistry problem of reaction-network elucidation is to find the actual set of constituent reactions underlying a given, overall reaction. For example, suppose that a chemist reads or discovers that the reagents H_2 and Br_2 yield HBr, i.e., $H_2 + Br_2 \rightsquigarrow HBr$. He might then study the reaction experimentally in his laboratory, to determine the set of elementary reaction steps, or reaction network, which constitute the overall reaction:

$$\begin{aligned} Br_2 &\rightleftharpoons 2Br \\ Br + H_2 &\rightleftharpoons HBr + H \\ H + Br_2 &\rightarrow HBr + Br \end{aligned}$$

Determining the reaction network is a complex problem of inductive inference. The space of candidates is huge, and even verifying with certainty that a particular candidate is the "true" one is problematic, because more than one network may give rise to similar experimental data.

The best way to elucidate reaction structure would be to appeal to a scientific theory that states uniquely the course of reactions, given only the initial starting materials and reaction conditions. Unfortunately, no such theory is currently available, so other means are employed.

Our approach is to automate the elucidation of reaction structure via laboratory experimentation. The intent is to embrace the complete cycle of experiment design, execution, and interpretation, in order to advance our understanding of how best to design an assistant experimental-scientist.

The motivation of the work presented in this report, then, is to extract rules of experiment interpretation from chemical laws themselves, where by "experiment interpretation" is meant inference having theoretical bearing on the problem at hand: elucidation of a particular structure. The rules obtained are not available in chemistry textbooks and journals, nor readily from chemists, hence the question of comparison with human knowledge-acquisition is not an issue.

3. Relevant Chemistry Knowledge and Definitions

A modest knowledge of chemistry concepts helps to understand the role of the synthetic rules, as well as the chemical laws from which they are derived; this section presents these concepts. Also, some chemical terminology is not precise enough for our purposes, so we introduce definitions to be followed throughout.

To describe how substances are transformed into new substances, chemists use notation such as $A + B \rightarrow T + X$, to summarize a process, or reaction, by which two chemical substances A and B are transformed into new substances T and X.

Next are some definitions. The target reaction is the reaction, such as $A + B \rightsquigarrow T$, whose structure is to be elucidated. We may also call this the aggregate reaction, to emphasize the likelihood that the reaction has underlying structure. The reagents are the chemicals that are placed in a reaction vessel to start an experiment; A,B are reagents of the target reaction just shown. A chemical species is a distinct molecular substance having a particular chemical behavior.[1]

[1] Two species may have the same molecular formula, yet exhibit different chemical behavior, due to different atomic

An important distinction is that between an aggregate reaction and an elementary reaction (the latter term is standard; we have introduced the first term). An elementary reaction is regarded as a single act: one or more molecules re-arrange themselves into new products, but no constituent reactions underlie the elementary reaction. If the reaction involves two or more molecules, they must meet in physical space for the reaction to occur. An aggregate reaction, on the other hand, may be composed of any number of steps, either on a path to the target product, or off it. We allow the class of aggregate reactions to include that of elementary reactions, i.e., an elementary reaction *is* an aggregate reaction, by our definition. We remark that the arrow notation for reactions implies nothing about whether a reaction is elementary. For example, the aggregate reaction $A + B \rightarrow T$ may be composed of two elementary reactions: $A + B \rightarrow X$ and $X \rightarrow T$.

A reaction network is (ordinarily) a set of elementary reactions; if the network reactions are aggregate, and not elementary, that will be stated explicitly. Each reaction of a reaction network is called a step.

The scope of our work is restricted to chemical reactions that occur entirely in one phase, e.g., if a reaction occurs in liquid solution, no mass is transformed into solid or gas phase. Reactions occurring in a single phase are termed homogeneous. Furthermore, we make the standard assumption that the volume of solution hosting the reaction remains constant.

The field of chemical kinetics studies the rates of chemical reactions. The basic law of kinetics asserts that, under certain conditions, the rates of change of concentrations can be characterized by a set of ordinary, non-linear, differential equations derived in a straightforward manner from the elementary reaction steps. For example, if the reaction $A + B \rightarrow C$ is elementary and occurs at constant volume, and if the solution is well-stirred, then the law asserts the following:

$$\begin{aligned} d[A]/dt &= -k_1[A][B] \\ d[B]/dt &= -k_1[A][B] \\ d[C]/dt &= k_1[A][B] \end{aligned}$$

where brackets denote the instantaneous concentration of the enclosed species. This reaction, like all elementary reactions, has a characteristic rate k_1 which is constant for fixed temperature, pressure, catalyst, and solvent. It is noteworthy that this law does not depend on the identity of the species, although the rates certainly do.

For the case of concurrent chemical reactions, the derivation of equations is well illustrated by example. The species concentrations undergoing the two reactions $A + B \rightarrow C$ and $2C \rightarrow D$ evolve as follows:

$$\begin{aligned} d[A]/dt &= -k_1[A][B] \\ d[B]/dt &= -k_1[A][B] \\ d[C]/dt &= k_1[A][B] - 2k_2[C]^2 \\ d[D]/dt &= k_2[C]^2 \end{aligned}$$

For each species, the effect of a reaction involving the species becomes a term on the right-hand side of its equation - positive for productive reactions, and negative for consumptive reactions.

A reversible reaction is one that can proceed in either direction, e.g., $A + B \rightleftharpoons C$. There is a rate constant associated with each direction. If the reactants and products of a reversible,

---

connectivity, or to different spatial configuration. Such molecules are called isomers.

elementary reaction are involved in no other reaction, then a dynamic equilibrium is reached at concentrations dependent on the two rate constants. Reactions still occur in both directions, but without observable effect on concentrations, which are in equilibrium at non-zero values.

A set of concurrent, elementary reactions at constant ambient conditions determines a system of differential equations having fixed rate constants. A network, as defined above, corresponds to a system of *parameterized* differential equations, i.e., the rates appear as symbolic constants in the system of equations.

A network, then, is an abstraction over an infinitude of fully-specified reactions (i.e., *reactions-with-rates*). These networks aggregate over instances whose concentration-versus-time trajectories are very dissimilar, due to the effect of variations in the rates. Therefore, these networks make much weaker predictions than the reactions-with-rates; for example, they may predict certain equilibrium points, such as that the final, asymptotic concentration of a reactant species is zero. In contrast, the reactions-with-rates predict unique trajectories over all time.

Finally, the single-phase chemical reactions that we study conserve mass, as well as number and type of atoms of each element. This implies that the system is stable, in the sense that any solution to the kinetical system is bounded.

4. Rules from Predictive Knowledge

The data-driven synthesis of retrodictive rules from predictive knowledge involves first carrying out predictions over a wide range of conditions. The data are then processed to form rules that associate predictions (i.e., observations) with a set of possible conditions or causes.

We next present the idea as developed within the subfield of diagnosis, and then show how structure elucidation can benefit from the same theme.

4.1. Rules for diagnosis

Diagnostic rules are synthesized via the repeated simulation of a predictive model under a wide range of model inputs and faults, to obtain outputs of the model.[2] These simulations yield triples of input, faults, output, which are processed to learn rules that associate input/output (mis)behaviors with possible faults.

This idea has been exploited in the AI diagnosis literature for models of:

- A satellite electrical-power subsystem [1]
- A particle-accelerator beam line [2]
- The human heart [3,4]

[2]By 'model' is meant any formal (e.g. computational, mathematical) means to infer a system's outputs from inputs, without actually exciting the system.

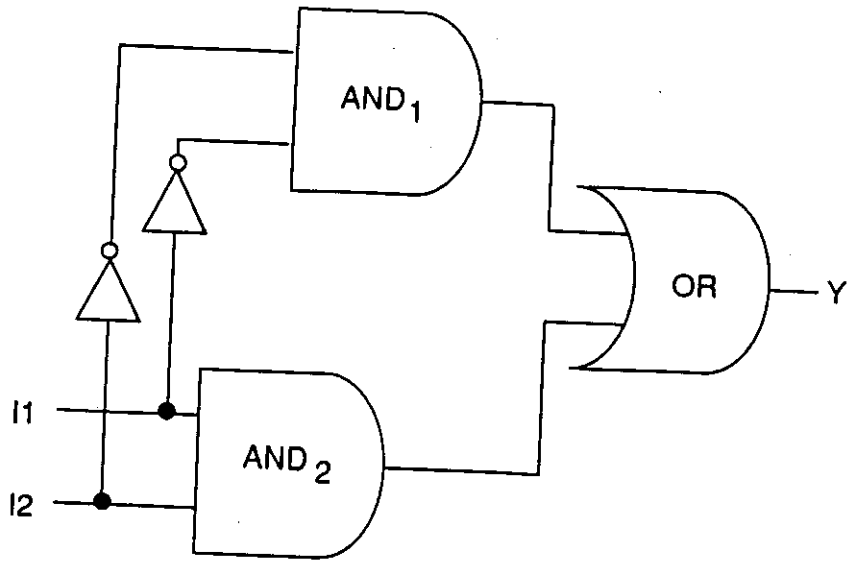

Figure 1: Simple Logic Circuit

An earlier advocacy of the idea is in [5].

To illustrate the technique in more detail, imagine that we have a model of the simple circuit of Figure 1. This circuit tests equality between its two inputs I_1,I_2: output Y is 1 exactly when equality holds. There are five logic gates: two two-input AND-gates, one two-input OR-gate, and two inverters. Let's assume that experience has taught that the only possible fault within such circuits is a permanent input-stuck-at: the *input* at a gate is *stuck at* logic 1 or 0, regardless of the logic value along its feeding connection. A complete set of diagnostic rules, expressed in terms of controllable values (the two inputs) and observable values (the output) obtains by simulating the circuit model on the cross-product of the set of possible inputs (2x2=4) with the set of gate-input states ($3^{3\cdot2+2\cdot1=8}$), because each gate-input is functioning, stuck at 1, or stuck at 0), and calculating the circuit output. In this way, *four* rules are synthesized, which associate a set of faulty gate-inputs with the four possible instances of incorrect output: a Y-output bit-flip for each of four possible inputs. Here is what a rule might look like:

If $I_1 = 1, I_2 = 0$, *and the output Y is discrepant, then*

$$\left(\begin{array}{c} input_1(OR) \text{ is stuck at } 1 \\ \vee \\ input_2(OR) \text{ is stuck at } 1 \\ \vee \\ input(INV_1) \text{ is stuck at } 0 \wedge input(INV_2) \text{ is stuck at } 0 \\ \vee \\ \vdots \end{array} \right)$$

Within a rule, each disjunct of the right-hand-side disjunction is a distinct, consistent, and complete explanation of the misbehavior covered by that rule.

Often the assumption is made that only a single fault (or say, at most two) occurs at one time, thus greatly reducing the size of the disjunctive right-hand-side of the synthesized rules. Distinct explanations arising from use of the rules on several, actual device tests can be intersected, to narrow the space of candidates still more.

This method presupposes a model of the mechanism of interest, and a catalogue of possible faults.

In the general case, there arises the difficulty of dealing with analog inputs, or with faults expressed over a continuous domain. Since not all points can be simulated, a sampling policy must be selected, and then the comprehensiveness of the derived rules becomes uncertain. Even discrete variables of infinite extent pose these problems. A plausible recourse is an empirical test of the synthesized rules, possibly by simulation, in which values not corresponding to any sample from the training set are tried. If the rules reliably include the true fault(s) in their diagnosis, then the rules are validated.

4.2. Analogy of structure elucidation with diagnosis

The goal of diagnosis is a physical explanation of misbehavior in terms of a deviation from a working model. In practice, those explanations postulating fewer deviations are entertained earlier.

The goal of structure elucidation is to find the true mechanistic structure, or reaction network in our case. This goal is similar to the diagnostic goal, except that a space of mechanisms is searched, rather than a space of faults. Simpler mechanisms are preferred, although what constitutes simplicity depends on the domain. In reaction chemistry, networks having fewer steps or fewer species are simpler.

4.3. Rules for structure elucidation

Elucidation rules are synthesized via the repeated simulation of many structural models under a range of model inputs, to obtain outputs of the model. These simulations yield triplets of input, structure, output, which are processed to acquire rules that associate input/outputs with structures.[3]

Our method learns rules that refute a single structural hypothesis. For a finite hypothesis space, a set of such rules having identical tests could be combined into a single rule that tests the same condition, but infers the list of unrefuted hypotheses, just like the diagnostic example of Figure 1. Our presentation will not take that second combination step; we merely remark on the equivalence here.

---

[3]Ours is not the first instance of rule induction for structure elucidation. The Meta-DENDRAL work induced rules for molecular-structure elucidation (distinct from reaction-network elucidation) from empirical, molecule/mass-spectral-data pairs [6,7]. We propose to induce them from uniquely predictive scientific laws, which were not available for the Meta-DENDRAL task.

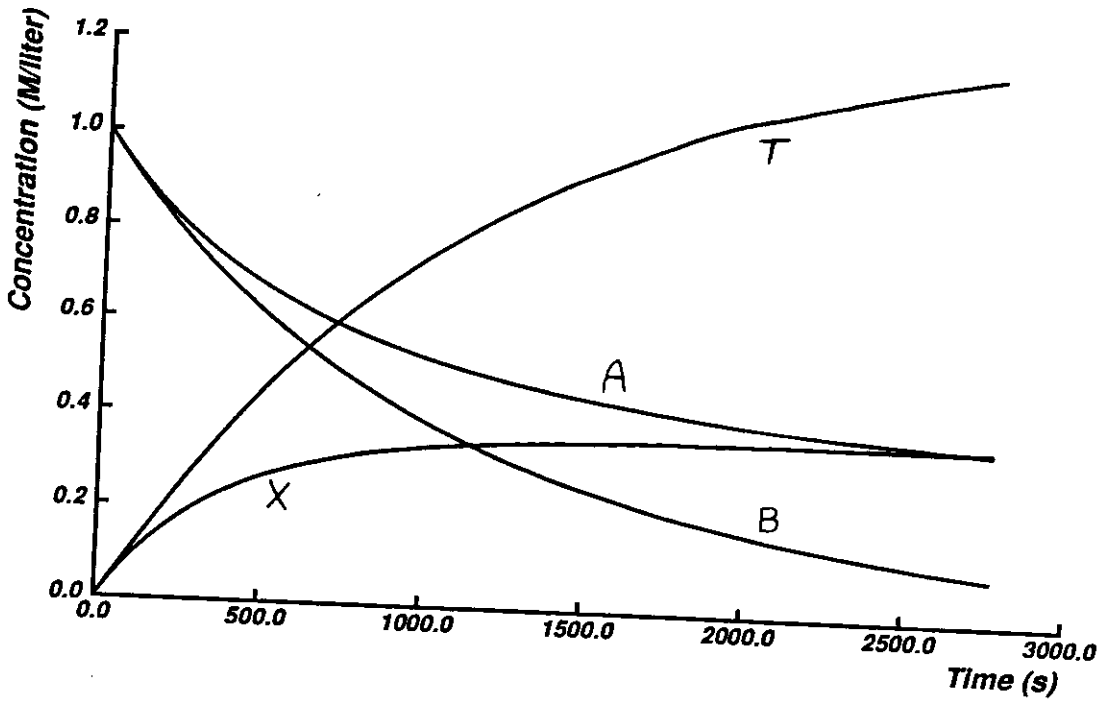

Figure 2: Course of the Reaction Concentrations

5. Development of the Method

5.1. Example network

We shall illustrate our method always with reference to the following reaction network (we shall refer to it as Ψ):

$$\begin{aligned} A + B &\rightarrow T + X \\ B + X &\rightarrow 2T \end{aligned}$$

for which the basic law of kinetics implies this system of differential equations (we shall refer to the system by Σ):

$$\begin{aligned} d[A]/dt &= -k_1[A][B] \\ d[B]/dt &= -k_1[A][B] - k_2[B][X] \\ d[T]/dt &= k_1[A][B] + 2k_2[B][X] \\ d[X]/dt &= k_1[A][B] - k_2[B][X] \end{aligned}$$

which characterizes the concentrations over time (trajectories) for each of the four species. The reference experiment is the reaction of the reagents A,B at equal, initial concentrations of 1 unit.

Figure 2 shows the course of the reaction on the reference experiment under the following reaction rates:

$$\begin{aligned} A + B &\overset{.001}{\rightarrow} T + X \\ B + X &\overset{.001}{\rightarrow} 2T \end{aligned}$$

5.2. A granularity mismatch

Experimental data from a reaction consist of numerous concentration measures over time. However, our hypotheses are reaction networks, which correspond to a system of parameterized differential

equations, such as Σ. Therefore, a hypothesis does not in general make unique, detailed predictions of concentration. This gives rise to a granularity mismatch between experimental data and what hypotheses predict.[4]

Our solution here is to extract predictions for the reference experiment that are unique over the entire span of the system parameters, i.e, network invariants. We have seen that such invariants cannot be the detailed concentration trajectories, but other, more abstract invariants are possible.

If a network invariant is violated by experimental data on the reference experiment, then the network predicting the invariant cannot be the true network.

From the network equations, derived from the basic law of chemical kinetics, we will synthesize convenient *retrodictive* knowledge: rules that test experimental concentrations data in order to refute the hypothesis that a certain network gave rise to the data.

5.3. Deciding on appropriate invariants

Our parameterized systems may make several invariant predictions on the reference experiment. For example, system Σ predicts that the equilibrium value for [B] is zero, and that d[B]/dt + d[T]/dt vanishes at time 0. Are these invariants interesting for our purpose?

To determine what are good invariants, we first establish some criteria, partly by examining more closely the character of experimental data.

5.3.1. Laboratory sampling and cut-off

Discrete, simulated data on species concentrations look like Figure 3; experimental data are not ordinarily as smooth. Noticeably, the curves are not dense; there are significant time gaps between measurements, due to the manipulations required to extract a sample of the reaction solution and analyze it. Also, the reaction is cut off after a time, either when the chemist believes equilibrium is reached, or when the resources (e.g. space, apparatus) are needed.

These realities affect refutation in the following way. Let's consider the noiseless case. A refutation, by definition, is decisive: it cannot be rescinded. If a refutation is based on data such as in Figure 3, then it cannot be the case that filling in the curve with more data, as well as extending it until equilibrium, invalidates the refutation. Therefore, our invariants must permit stable, or monotonic refutations, according to the following Principle of Stable Refutation:

Principle 1 *The absence of an invariant is stable with respect to more data from increased reaction sampling or extent in time.*

An invariant that violates this principle due to cut-off is the equilibrium concentration for a species. One never knows whether by allowing a reaction to proceed longer, the species concentration might

[4] A network hypothesis *does* predict that there exist values for the unknown parameters that entail the experimental data, with allowance for noise and experimental uncertainty.

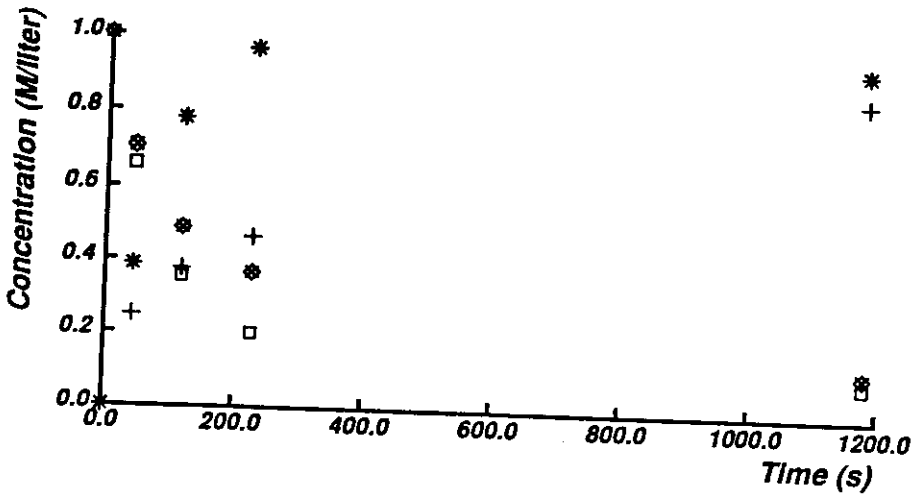

Figure 3: Typical Concentrations Data

reach zero. Note that it is only the failure of the invariant to hold, not its fulfillment, that must be stable in the sense of Principle 1.

An invariant that violates Principle 1 due to sampling would be that two trajectories $[X_1]$ and $[X_2]$ always cross at some time. To illustrate why with an extreme case, consider two sampled trajectories that appear as parallel lines. Between two consecutive sample times, the trajectories could cross and resume their seemingly straight path in time for the next sample. Hence, basing a refutation on the seeming absence of crossings would be unjustified, so the crossing invariant violates the principle.

5.3.2. Operationality

An acceptable invariant must be operational in practice, given the intended source of experimental data. Potentially informative features, such as the shape of a trajectory's rate-of-change, are too sensitive to noiseful data. Note that deciding whether this criterion holds is not clear-cut like the others to be stated here; rather it requires some judgment and knowledge of the accuracy of experimental data.

5.3.3. Generality

It would very inconvenient if there were needed a separate batch of rules to interpret these two experiments:

- Combine reagents at equal concentrations of .01 moles/liter.
- Combine reagents at equal concentrations of .005 moles/liter.

Particular reagents may not easily be monitorable at low concentrations, others at higher concentrations, but the same rules should apply to the experiment where both reagents have equal initial concentrations.

To assure that experiment interpretation does not depend on the magnitude of initial concentrations, we require that an acceptable invariant not depend on any absolute level of concentration,

only on relative levels.

5.3.4. Reliability

There can be invariants of a system of differential equations which are not easily proven analytically, because the mathematical techniques called for are not available. A recourse is to test for a proposed invariance by sampling the system's parameter space, performing numerical integration of the equations on the reference experiment, followed by testing whether the property holds for the concentrations trajectories. If the property holds at all tested points in the parameter space, then it is a plausible invariant of the system. Of course, the sampling procedure may miss a region in which the property fails to hold. If we lack analytical tools, our confidence that the system sample has discovered a persistent invariant relies on the extent of sampling, together with any background knowledge about system behavior.

Our systems of differential equations possess the strong constraint of conservation of mass: exactly that reagent mass present at the start of an experiment remains in the reaction, only being distributed among the new products.[5] Hence, the stability (i.e. boundedness) of these dynamical systems is guaranteed. Also, as discussed in Appendix 3, the solution to a chemical-kinetic system is an analytic function of the initial conditions and parameters, so that a certain smoothness is assured. This makes sampling a more credible tactic.

However, to increase further our confidence that the samples adequately capture the behaviors spanned by that space, we shall require that the behavior found at a sample point S not change under certain mappings of S to other points in the parameter space, as discussed next.

One convenient mapping is that of uniform multiplication: If the coordinates of a sample point (i.e. values for the system parameters) are uniformly multiplied by a constant, then the fulfillment of system invariants *shall not change*. In this way, a single sample of the parameter space gives at no cost an infinitude of other system samples, including parameter values of much greater or lesser magnitude than the "root" sample. How to guarantee fulfillment under such a mapping is the next issue.

We note that the equations of the differential system are *linear* in the parameters, for example:

$$d[T]/dt = k_1[A][B] + 2k_2[B][X]$$

This means that uniform multiplication of the parameters k_i by a constant is equivalent in effect to changing the units of time used, say from seconds to milliseconds, or to megaseconds.

Therefore, to assure that uniform multiplication by a constant not change fulfillment of an invariant, we shall require that an acceptable invariant not depend on the time units. We obtain in this way the mentioned infinitude of "free samples."

[5] As stated in section 3, the scope of our work is limited to homogeneous reactions, e.g., loss of mass from undetectable precipitation is excluded. Also, phenomena such as evaporation are ignored.

5.3.5. Proposed invariants

We recapitulate here the four criteria for acceptable invariants developed in the previous subsections. These are:

- Invariants must permit stable refutation in the face of data from more frequent reaction sampling, or from extending the experiment longer.
- Invariants must be operational within the context of particular experimental apparatus.
- Invariants should depend only on relative concentrations, not on their absolute levels.
- Invariants should not depend on the units used for time.

The invariants used currently are the following:

1. A species trajectory never increases.
2. A species trajectory never decreases.
3. The trajectory for one species dominates (i.e. is always $\geq$) the trajectory for a second species.

These invariants satisfy the four criteria listed. In general, the presence of these invariants in a kinetical system of equations having fixed initial conditions seems not inferrable analytically. Hence, the fourth criterion, of independence of time units, is critical to augment the sampling coverage.

5.4. Invariants of example network

Our example network,

$$\begin{aligned} A + B &\xrightarrow{k_1} T + X \\ B + X &\xrightarrow{k_2} 2T \end{aligned}$$

exhibits the following invariants:

- A,B monotonically decrease
- T monotonically increases
- A dominates B
- T dominates X

These invariants were discovered by sampling the parameter space k_1, k_2 at $7 \times 7 = 49$ points, using a 7-level factorial design of the parameter space at values of successive powers of ten.

The invariants found for this case are seen correct by reasoning at the level of reaction steps. For example, T dominates X because they are produced equally by the first reaction step, whereas an X is consumed to produce 2T at the second step. However, reasoning at the reaction-step level is not powerful enough to account for the invariants found in many cases, as will be seen on the examples in Appendix 2.

If concentrations data from an experiment violate any of the invariants of the example network, then the network could not have given rise to the data (assuming no noise). To be explicit, the rule derived from the second invariant above is:

If the trajectory for species T ever decreases, then the network is refuted.

5.5. Parameter-space sampling

The purpose of sampling is to discover system invariants. Each parameter represents the speed of a reaction step; for instantaneous reactant concentrations, it determines the step's contribution to instantaneous rate-of-change of the species in the step. Reaction speeds are positive quantities, and vary over huge ranges, with little constraint a priori on possible values.

One reasonable sampling policy varies the reaction speeds in a systematic manner, so that each step has a chance to dominate with regard to speed. Our current approach is a full 7-level factorial design in the parameter space, in which each parameter assumes a value from the set $10^{k,k=1,...,7}$. According to the factorial-design regimen, the number of points tested is 7^N, where N is the number of factors (parameters in our case).

Appendix 1 points out redundance in the factorial design that can be exploited to reduce the amount of computation.

6. Practical Use of Refutation Rules

The preceding sections developed a method for synthesis of idealized refutation rules with no attention to their practical use. This section discusses the practical aspects, and illustrates the potential discriminatory power on a real chemical reaction.

6.1. The rules are experiment-specific

The invariants currently used result in rules that interpret the reference experiment, in which the two reagents are combined at equal concentrations. The rules are applicable regardless of the initial magnitudes of concentration. However, the rules are sensitive to the ratios of the reagent concentrations; an experiment at a 2:1 ratio cannot be interpreted with the rules synthesized on the reference experiment. Although it would be very convenient to apply the same rules regardless of the reagent ratios, much fewer invariants would be detected over such a wide range of conditions.

In this report we only consider the reference experiment, in which only the reagents are present at the reaction's start. Synthesizing rules for other reaction experiments is conceptually identical, and only involves setting different initial conditions for the numerical integrations.

6.2. Disconfirmation versus refutation

Because noise is present in any experimental data,[6] definitive rejection of a hypothesis when data do not satisfy its predictions is risky. Therefore, to decrease the risk of rejecting the correct hypothesis, any single evidence based on experiment should only disconfirm, i.e., augment inconclusively disbelief in the hypothesis.

Our rules interpret experimental data, so it follows that they should not be used to refute, only to disconfirm. The reason for stressing refutation in their development is to localize to the extent possible the sources of uncertainty attending their use.

Moreover, even on idealized data our rules are not certain, since they are derived by induction on simulated data; the rules themselves nevertheless have a high degree of confirmation, because the invariants that they test proved true on all the parameter points examined.

Rules have already been generated for all networks having two reaction steps, to be used as disconfirmatory evidence within our reaction-structure elucidation programs.

6.3. Discriminatory potential

We illustrate the potential of the rules to discriminate among hypotheses by the following real example. The synthesis of $C_{31}H_{37}N_2I$ within the following scheme:

$$C_{13}H_{18}NI + C_{17}H_{16}N_2 \rightsquigarrow C_{31}H_{37}N_2I + C_{24}H_{27}N_2I + C_6H_7N$$

from the two reagents $C_{13}H_{18}NI$ and $C_{17}H_{16}N_2$ involves also the formation of the two products shown at the right, above.

Three networks for this reaction were hypothesized by a separate program, which generates initial hypotheses based only on the molecular formulas of reagents and known products, and on simplicity.

[6] For our purpose, noise is any deviation from the concentrations that are predicted by the equations of the "true" reaction network. It is not necessarily any fault of measurement or data collection, e.g., the solution might not have been stirred adequately, so that the basic law of kinetics is obeyed only approximately.

Schematically, these networks are as follows:

$A+B$	$\rightarrow$	$X+Y$		
$2X$	$\rightarrow$	$T+Z$		
$(\rightarrow\rightarrow)$	$\Downarrow(A,B)$	$\Uparrow(T,Y,Z)$	$=(AB,TZ)$	$\geq(Y/TXZ)$
$(\rightarrow\rightleftharpoons)$	$\Downarrow(A,B)$	$\Uparrow(T,Y,Z)$	$=(AB,TZ)$	$\geq(Y/TXZ)$
$(\rightleftharpoons\rightarrow)$	$\Downarrow(A,B)$	$\Uparrow(T,Y,Z)$	$=(AB,TZ)$	$\geq(Y/TXZ)$
$(\rightleftharpoons\rightleftharpoons)$	$\Downarrow(A,B)$	$\Uparrow(T,Y,Z)$	$=(AB,TZ)$	$\geq(Y/TXZ)$

$A+B$	$\rightarrow$	$X+Y$		
$B+X$	$\rightarrow$	$T+Z$		
$(\rightarrow\rightarrow)$	$\Downarrow(A,B)$	$\Uparrow(T,Y,Z)$	$=(TZ)$	$\geq(A/BTZ,Y/TXZ)$
$(\rightarrow\rightleftharpoons)$	$\Downarrow(A,B)$	$\Uparrow(Y)$	$=(TZ)$	$\geq(A/BTZ,Y/TXZ)$
$(\rightleftharpoons\rightarrow)$	$\Downarrow(B)$	$\Uparrow(T,Z)$	$=(TZ)$	$\geq(A/BTZ,Y/TXZ)$
$(\rightleftharpoons\rightleftharpoons)$			$=(TZ)$	$\geq(A/BTZ,Y/TXZ)$

$A+B$	$\rightarrow$	$X+Y$		
$A+X$	$\rightarrow$	$T+Z$		
$(\rightarrow\rightarrow)$	$\Downarrow(B,A)$	$\Uparrow(T,Y,Z)$	$=(TZ)$	$\geq(B/ATZ,Y/TXZ)$
$(\rightarrow\rightleftharpoons)$	$\Downarrow(B,A)$	$\Uparrow(Y)$	$=(TZ)$	$\geq(B/ATZ,Y/TXZ)$
$(\rightleftharpoons\rightarrow)$	$\Downarrow(A)$	$\Uparrow(T,Z)$	$=(TZ)$	$\geq(B/ATZ,Y/TXZ)$
$(\rightleftharpoons\rightleftharpoons)$			$=(TZ)$	$\geq(B/ATZ,Y/TXZ)$

Below each network is a list of possible reaction-step directions, together with the invariants found on the reference experiment for each direction pair. An entry involving '=' indicates that between two species each dominates the other, hence their trajectories are coincident.

The three hypotheses, even after aggregating over possible directions, predict different relations between reagents A and B. The first network predicts coincident trajectories; the second and third networks predict respectively that A dominates B, and vice versa. If the true network is one of these three, then the reference experiment can partially rule out two others, in case that the concentration of one reagent is at any time significantly greater than the other.

Note that if the two reagent trajectories cross, so that neither dominates the other, then all three hypotheses are disconfirmed.

7. Conclusion

This work has extended the idea of synthesizing convenient, retrodictive knowledge from predictive knowledge to the domain of structure elucidation, within the context of chemical kinetics. The emphasis was on synthesis of refutation-style rules for the interpretation of experiments, which in practical use would serve as reliable, disconfirmatory evidence.

The possibility of refutation arises from discovering invariants in a parameterized system of ordinary, non-linear, differential equations, via induction over many instances of simulated data obtained by numerical integration using different parameter values.

Four criteria were developed to guide the selection of invariants. The Principle of Stable Refuta-

tion uses the logical non-rescindibility of refutation to constrain the invariants that are suitable for sampled, cut-off experimental data. Other criteria referring to operationality, generality, and reliability were presented. The invariants chosen - monotonicity and pair-wise dominance - are not generally inferrable analytically, so simulation experiments are done to discover them by induction from data.

The reliability of the invariants, and of the rules derived from them, is supported by choosing invariants that do not depend on the units of time. By so doing, behaviors at an infinitude of points in parameter space are captured by examining a single point in the space. Hence, the synthetic rules achieve a considerable degree of confirmation, although they are not proven true.

An advantage of the rules is that they are crisp and intelligible to the human chemist, who might be asked to judge whether a gross violation of an invariant has occurred.

7.1. Limitations

The limitations to the method are several. First, the invariants are established by a factorial design on the parameter space, so computation time increases exponentially with the number of parameters. The run time for our example network, which involves 13 numerical integrations, is about 1 minute on an IBM RT-PC with floating point accelerator. On the other hand, calculations need be done only once for all time, because the reaction networks are expressed in terms of variables A,B,T, etc., which abstract the identities of molecules.

A second issue to be addressed is the stability of the invariants with respect to errors in the experimental initial conditions. It is not possible experimentally to combine two reagents in exactly equal concentrations, so that rules developed for the reference experiment are subject also to that uncertainty. One remedy to this problem could be to verify during simulations that the invariants persist despite slight perturbations in initial conditions. Instead, we prefer to apply the rules conservatively, i.e., to decide that an invariant has been violated experimentally only when the violation is gross.

Appendix 3 shows that the solution to a chemical-kinetic system is stable, in the sense that small perturbations of the initial conditions lead to bounded perturbations of the solution, for any finite time interval. This fact lends some credibility to the stability of our invariants with respect to perturbations in the initial conditions.

8. Acknowledgments

Thanks are due to Prof. Herbert Simon for technical suggestions on this work. Any shortcomings remain the responsibility of the author.

Profs. Jonathan Lindsey (Chemistry), Tom Mitchell, and Herbert Simon have regularly guided the larger network-elucidation effort. Prof. Bruce Buchanan (Univ. of Pittsburgh) has also made several valuable suggestions. François Lecouat was an early collaborator on the elucidation problem.

Dr. Scott Clearwater (Univ. of Pittsburgh) made helpful comments on a draft of this report.

Numerical-integration Fortran programs were contributed by Dr. Patrick McCroskey, now of the Dow Chemical Central Engineering Research Laboratory.

References

[1] D. Pearce, "The induction of fault diagnosis systems from qualitative models," in *Proceedings of AAAI*, pp. 353–357, 1988.

[2] B. Buchanan, S. Clearwater, J. Sullivan, and T. Cheng, "Simulation-assisted inductive learning," in *Proceedings of AAAI*, pp. 552–557, 1988.

[3] I. Mozetic, "Compression of the ecg knowledge-base using the aq inductive learning algorithm," Report UIUCDCS-F-85-943, Department of Computer Science, University of Illinois at Urbana-Champaign, 1985.

[4] I. Bratko, I. Mozetic, and N. Lavrac, *KARDIO: A Study in Deep and Qualitative Knowledge for Expert Systems.* MIT Press, 1989. In press.

[5] Y. Ishida and L. Eshelman, "Integrating model-based and syndrome-based diagnosis," Technical Report CMU-CS-87-111, Carnegie Mellon University, 1987.

[6] R. Lindsay, B. Buchanan, E. Feigenbaum, and J. Lederberg, *Applications of Artificial Intelligence for Organic Chemistry: The Dendral Project.* McGraw Hill, 1980.

[7] B. Buchanan and T. Mitchell, "Model-directed learning of production rules," in *Pattern Directed Inference Systems*, (D. Waterman and F. Hayes-Roth, eds.), pp. 297–312, Academic Press, 1978.

[8] A. Andronov, A. Vitt, and S. Khaikin, *Theory of Oscillators.* Pergamon Press, 1966.

[9] R. V. Churchill, J. W. Brown, and R. F. Verhey, *Complex Variables and Applications.* McGraw Hill, 1976. See theorem on Page 132.

A Exploiting redundance within the factorial design

Section 5.3.4 showed that the behavior of a system of kinetical equations with respect to our invariants is not changed by uniform multiplication of each system parameter by a constant. Hence, there is some redundance in the factorial design which can be eliminated.

For example, using the levels $10^{k,k=1,...,7}$ on a network having two parameters k_1,k_2, the case (k_1=10^2, k_2 = 10^3) is identical to the case (k_1=10, k_2 = 10^2). Exploitation of this redundance reduces the number of trials within the factorial design from q^n to $q^n - (q-1)^n$, where 'q' is the number of levels, and 'n' is the number of parameters. For us, q=7, so the fraction of computation needed $1 - (6/7)^n$.

One implements this savings within an algorithm as follows. If during an iteration all the parameters have values greater than the lowest level, then that case need not be done.

B Examples of invariants of other networks

Here is a two-step, undirected network of five species:

$$\begin{aligned} A + B &\approx X + Y \\ Y &\approx 2T \end{aligned}$$

where the symbol '$\approx$' abstracts with respect to the direction of the reaction (forward, backward, or reversible). The invariants for each assignment of directions on this network turn out identical, so that the invariants are also invariants of the undirected network. The invariants are:

- A,B never increase.
- T,X never decrease.
- A dominates B and vice versa (i.e., A=B).
- X dominates Y.

Next we consider the example network used in this report, in which the invariants are not identical over the different assignments of direction. The undirected network is:

$$\begin{aligned} A + B &\approx T + X \\ B + X &\approx 2T \end{aligned}$$

The invariants, grouped by reaction directions and using a more concise notation, are:

$$\begin{array}{ll} (\rightleftharpoons, \rightarrow) & B\downarrow, T\uparrow, A \geq B, T \geq X \\ (\rightleftharpoons, \leftarrow) & A\downarrow, X\uparrow, B \geq A, X \geq T \\ (\rightarrow, \rightarrow) & A\downarrow, B\downarrow, T\uparrow, A \geq B, T \geq X \\ (\rightarrow, \leftarrow) & A\downarrow, X\uparrow, B \geq A, X \geq T \\ (\rightarrow, \rightleftharpoons) & A\downarrow \\ (\rightleftharpoons, \rightleftharpoons) & \emptyset \end{array}$$

For example, '$B\downarrow$' means that the trajectory of species B never increases. Notably, the directed network $A + B \rightleftharpoons T + X, B + X \rightleftharpoons 2T$ has none of our invariants.

C Mathematical observations on chemical-kinetic systems

We refer here to the basic theorems as formulated in the appendix to a treatise on the theory of differential equations[8].

First, we note that the right-hand sides of the kinetical equations are everywhere analytic functions of their arguments (state variables and parameters), because their derivatives exist everywhere.

Below, let an initial-value problem refer to a chemical-kinetic system of equations together with initial conditions.

From the conditions of Theorem II on page 796, which are weaker than analyticity, one deduces that the solution to an initial-value problem is stable, in the sense that a small perturbation in the initial conditions issue in a small perturbation in the solution, over any finite time interval.

From the analyticity condition of Theorem VII on page 800, one deduces that the solution to an initial-value problem is an analytic function of the initial conditions and of the parameters.

A theorem from complex analysis[9] states that if a function is analytic at all points, then its derivatives of all orders are also analytic functions at all points.