Distilling Free-Form Natural Laws from Experimental Data

By Michael Schmidt and Hod Lipson

Presenter David Alsip

“Even though it looks like it’s changing erratically, there is always something deeper there that is always constant”
Lipson

Introduction

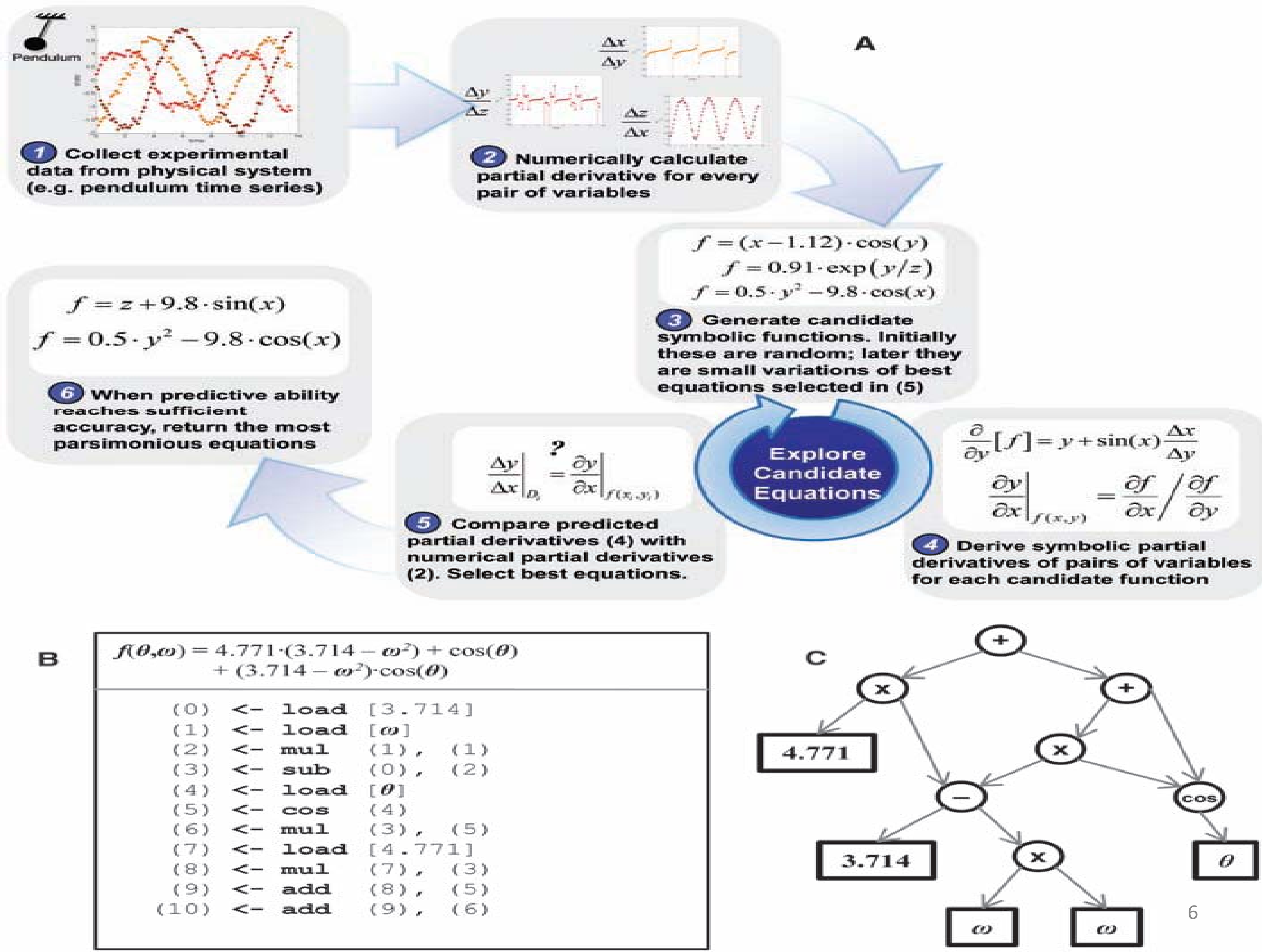
- Scientists for centuries have attempted to identify natural laws that underpin physical phenomena in nature
- Resistance to automation, despite occurrence of increased computer power
- Key to this is finding algorithms correlate correctly to observed data
- The paper proposes a principle for identification of the nontrivial

Goals

- To find natural relations where they exist
- Identify linked factors within a dataset
- Use unconstrained analytical expression that explain symbolically precise relations
- With the ultimate goal of uncovering an “alphabet” to describe the system

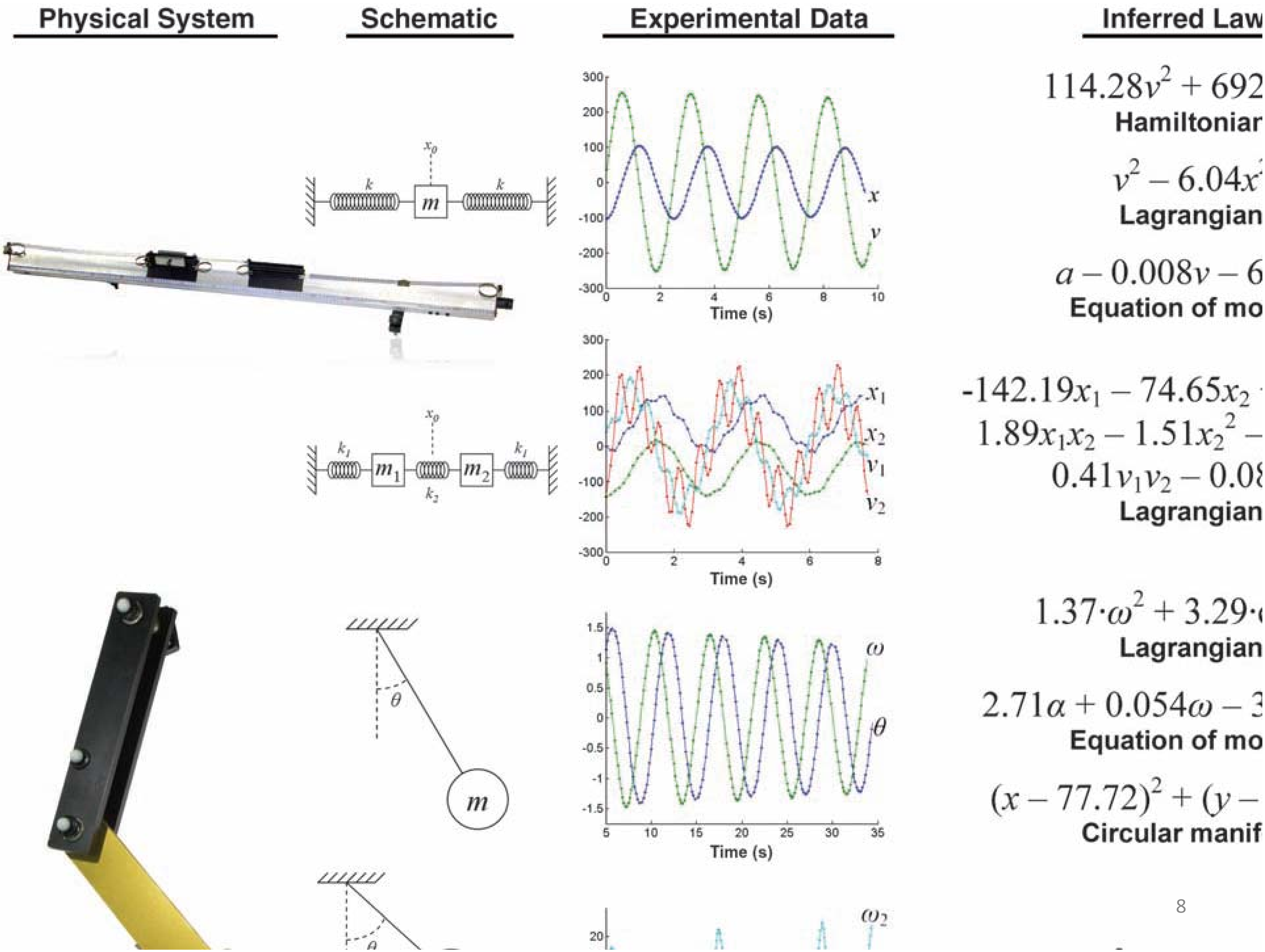
Experiment Setup

- Their process begins by taking the derivatives of every variable observed with respect to every other
- This is mathematical way of measuring how one quantity changes as another changes
- The program creates equations at random using various constants and variables from the data


A
Pendulum
1 Collect experimental data from physical system (e.g. pendulum time series)
2 Numerically calculate partial derivative for every pair of variables
3 Generate candidate symbolic functions. Initially these are random; later they are small variations of best equations selected in (5)
$f = (x - 1.12) \cdot \cos(y)$
$f = 0.91 \cdot \exp(y/z)$
$f = 0.5 \cdot y^2 - 9.8 \cdot \cos(x)$
Explore Candidate Equations
4 Derive symbolic partial derivatives of pairs of variables for each candidate function
$\frac{\partial}{\partial y}[f] = y + \sin(x)\frac{\Delta x}{\Delta y}$
$\left.\frac{\partial y}{\partial x}\right|_{f(x,y)} = \frac{\partial f}{\partial x} \Big/ \frac{\partial f}{\partial y}$
5 Compare predicted partial derivatives (4) with numerical partial derivatives (2). Select best equations.
6 When predictive ability reaches sufficient accuracy, return the most parsimonious equations
$f = z + 9.8 \cdot \sin(x)$
$f = 0.5 \cdot y^2 - 9.8 \cdot \cos(x)$
B
$f(\theta,\omega) = 4.771 \cdot (3.714 - \omega^2) + \cos(\theta) + (3.714 - \omega^2) \cdot \cos(\theta)$
(0) <- load [3.714]
(1) <- load [ω]
(2) <- mul (1), (1)
(3) <- sub (0), (2)
(4) <- load [θ]
(5) <- cos (4)
(6) <- mul (3), (5)
(7) <- load [4.771]
(8) <- mul (7), (3)
(9) <- add (8), (5)
(10) <- add (9), (6)
C
4.771
3.714
cos
θ
ω
ω

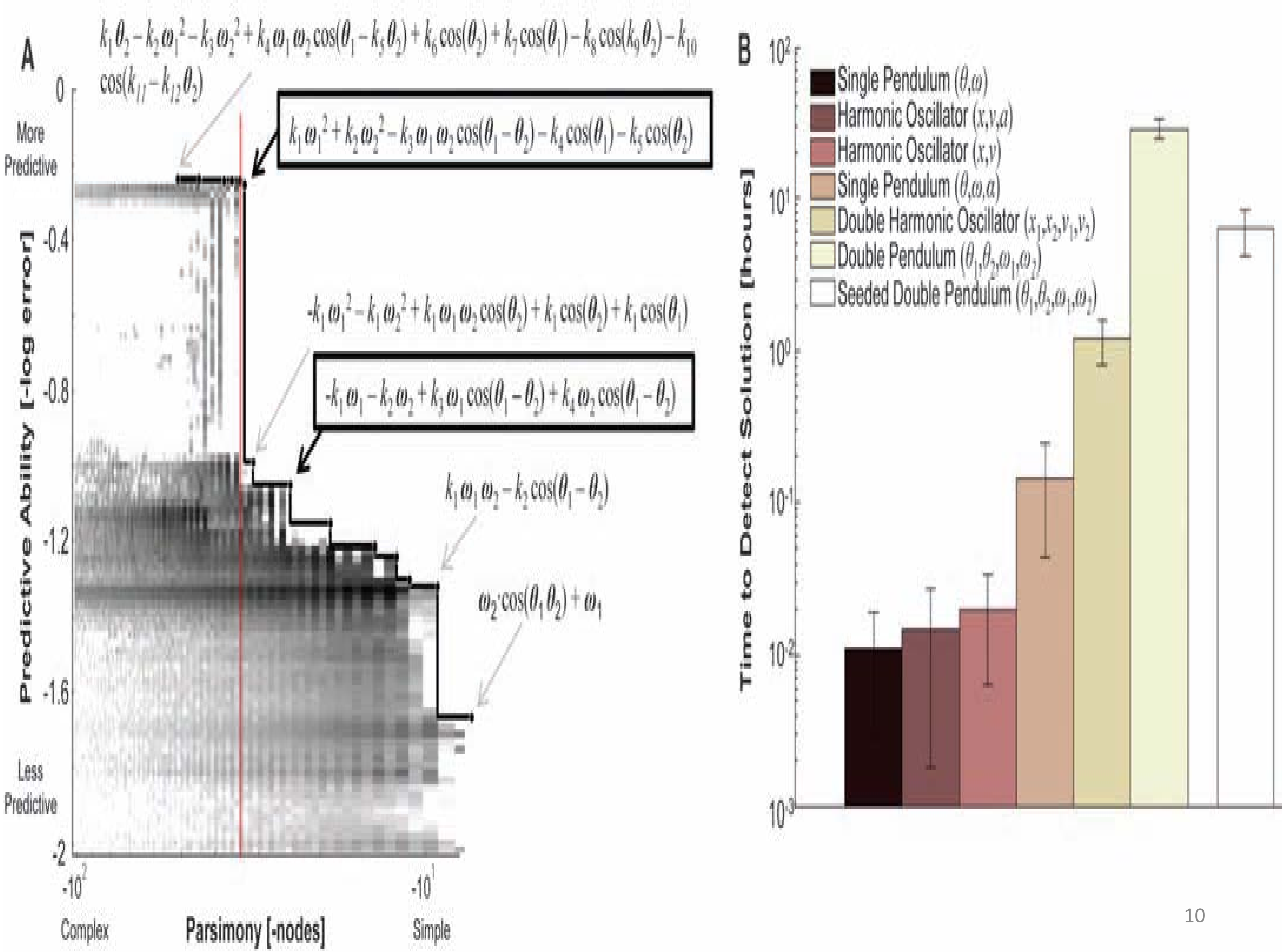
Outcome

- Initially the output failed to explain the data
- But some failures were slightly less wrong than others
- A genetic algorithm the program modified the most promising failures and tested the again chose the best and repeated
- This was repeated until a set of equations evolved to describe the system


Physical System
Schematic
Experimental Data
Inferred Law
114.28v² + 692
Hamiltoniar
v² – 6.04x
Lagrangian
a – 0.008v – 6
Equation of mo
-142.19x₁ – 74.65x₂
1.89x₁x₂ – 1.51x₂² –
0.41v₁v₂ – 0.0
Lagrangian
1.37·ω² + 3.29·
Lagrangian
2.71α + 0.054ω – 3
Equation of mo
(x – 77.72)² + (y –
Circular manif
Time (s)
x
v
x₁
x₂
v₁
v₂
ω
θ
ω₂
k
k₁
k₂
x₀
m
m₁
m₂

Conclusion

- The program evolves these laws without any prior knowledge of physics, kinematics or geometry
- But time consuming
- On a parallel computer with 32 processors, simple linear motion could be analyzed in few minutes, but the complex double pendulum required 30 to 40 hours
- Could be applied to more complex systems such as biology to cosmology



A
$k_1\theta_2 - k_2\omega_1^2 - k_3\omega_2^2 + k_4\omega_1\omega_2\cos(\theta_1 - k_5\theta_2) + k_6\cos(\theta_2) + k_7\cos(\theta_1) - k_8\cos(k_9\theta_2) - k_{10}\cos(k_{11} - k_{12}\theta_2)$
$k_1\omega_1^2 + k_2\omega_2^2 - k_3\omega_1\omega_2\cos(\theta_1 - \theta_2) - k_4\cos(\theta_1) - k_5\cos(\theta_2)$
$-k_1\omega_1^2 - k_1\omega_2^2 + k_1\omega_1\omega_2\cos(\theta_2) + k_1\cos(\theta_2) + k_1\cos(\theta_1)$
$-k_1\omega_1 - k_2\omega_2 + k_3\omega_1\cos(\theta_1 - \theta_2) + k_4\omega_2\cos(\theta_1 - \theta_2)$
$k_1\omega_1\omega_2 - k_2\cos(\theta_1 - \theta_2)$
$\omega_2 \cdot \cos(\theta_1 \theta_2) + \omega_1$
More Predictive
Less Predictive
Predictive Ability [-log error]
0
-0.4
-0.8
-1.2
-1.6
-2
-10^2
-10^1
Complex
Parsimony [-nodes]
Simple
B
Time to Detect Solution [hours]
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}
Single Pendulum (θ,ω)
Harmonic Oscillator (x,v,a)
Harmonic Oscillator (x,v)
Single Pendulum (θ,ω,a)
Double Harmonic Oscillator (x_1,x_2,v_1,v_2)
Double Pendulum $(\theta_1,\theta_2,\omega_1,\omega_2)$
Seeded Double Pendulum $(\theta_1,\theta_2,\omega_1,\omega_2)$

- By "bootstrapping" or seeding the complex pendulum problem with terms from equation for the simple pendulum cut the processing time to 7 or 8 hours.