Are you a scientist with interesting yet unexplained data that you don't have the time to analyse? You might want to get in touch with two physicists in the US who have created an algorithm that can deduce physical laws from raw experimental data with little help from humans.

Without any knowledge of physics or geometry, the algorithm discovered exact energy and momentum relations governing the dynamics of mass-spring systems as well as single and double pendulums. The researchers envisage such algorithms speeding up the scientific process by reducing the time needed to identify potentially interesting models of particular systems.

Since the 1960s scientists have been using artificial intelligence to design and run experiments, developing ever more powerful programs to generate, collect and store data. However, they have had less success in automatically distilling these data into new scientific laws.

Modelling algorithms can allow scientists to focus on developing new theories rather than spending their time comparing models with data

Hod Lipson, Cornell University

Now, two new papers in the journal *Science* confront this problem. One describes the development of a robot that can generate and then test hypotheses about biological systems, while the other, by
computational biologist Michael Schmidt and engineer and computer scientist Hod Lipson of Cornell University in the US, explains how conservation laws can be generated automatically (Science 324 81).

**Meaningful or trivial correlations?**

Schmidt and Lipson say that the biggest difficulty in searching for new conservation laws using a computer is identifying meaningful, as opposed to trivial, correlations within data. They point out that the experimental data from a physical system can yield an infinite number of invariant equations, but that only a few of these will have anything interesting to say about that system. Their solution to this problem is to say that an equation is useful only if it can predict how the system’s sub components affect each other over time.

To put this into practice, Schmidt and Lipson set up an algorithm that takes measurements of certain variables over time within a particular physical system, such as the x, y and z coordinates of a pendulum.

The algorithm numerically calculates the partial derivatives for every pair of variables; then generates functions that might describe the behaviour of the system by randomly sticking together algebraic operators (+, -, ÷, ×), analytical functions (such as sine and cosine), constants and variables; and then works out the partial derivatives of each of these functions.

The best candidate functions are those whose partial derivatives most closely match the numerical partial derivatives. These functions can then be further refined until they reach a certain level of accuracy.

**Chaos in simple systems**

To test their algorithm, the researchers investigated four different physical systems – a single mass held between two springs; two masses held between three springs; a single pendulum; and a double pendulum (one pendulum swinging off the bottom of another).

Given time-varying position and velocity data, the algorithm was able to identify the energy laws of each system — the Hamiltonian (total energy) and Lagrangian (kinetic energy minus potential energy). When it was also supplied with acceleration data, it generated the equations of motion corresponding to Newton’s second law for each system.

The algorithm does not produce a unique equation in each case but a shortlist of around ten candidate equations. These represent the most accurate equations for a range of complexities (i.e. number of terms in each equation). It is then down to the scientist to choose his or her favourite.

**Plugging theoretical gaps**

According to Lipson, this automated approach to law discovery could help in those areas of science where there is “a theoretical gap despite abundance of data”. Cosmology, he says, would be one such area, and biology in general another. “In biology there are many systems where we do not know their dynamics or the rules that they obey,” he adds. “Detecting an invariant could help scientists focus more quickly on an interesting aspect of the system, even if it is not fully understood.”
Robert Crease, a philosopher at Stony Brook University in the US, believes that algorithms can help to advance science “within an already-understood horizon” but says that this is not the interesting part of science. That, he says, “involves discovering something that transforms our current horizon, using a taste for the interesting, the willingness to entertain paradox, and the sense that something is a puzzle and not just an error or absurd contradiction.”

**Humans are still needed**

Lipson does not claim that automation can replace scientists, because, he says, humans are still needed to choose which data to collect, what the building blocks for equations should be, and also to “give meaning to the results”. However, he believes that algorithms like theirs can speed up the investigation and modelling of new phenomena. “Just as automated design algorithms allow engineers to delegate mundane tasks to computers,” he says, “modelling algorithms can allow scientists to focus on developing new theories rather than spending their time comparing models with data.”

David Waltz, a computer scientist at Columbia University in the US and coauthor of a commentary on the two *Science* papers, does not believe that the Schmidt and Lipson algorithm is likely to produce a truly profound result in the near future. But he believes that the general approach could become much more sophisticated, envisaging that intelligent systems could continuously look for correlations in the data from an ever larger range of experiments in such areas as astronomy, geophysics and particle physics. “I expect that computational systems will exhibit increasing amounts of what we would today say requires human insight,” he adds.

**About the author**

Edwin Cartlidge is a freelance journalist based in Rome, Italy