Abstract

Motivatioin: Dynamic Systems have proved to be a major tool in modelling and understanding biological processes and chemical reactions. The reconstruction of system states under statistical errors using e.g. Kalman Filters and Particle Filters is a long discussed topic. However serious issues a likely to occur as soon as the the underlying systems equations contain systematic error, e.g. unknown interactions. The R package seeds provides two algorithms that aim a numerical estimation of systematic model and simultaneously estimate the system states.

Availability: The R package seeds is provided via CRAN.

Contact: seeds@hs-koblenz.de

1 Introduction

Cell biology, chemical reactions and mechanical systems are just some examples of processes that can be modelled and investigated using dynamic systems. In the mathematical formulation these take the form

$$\dot{x}(t) = f(x(t), u(t)) \qquad (1)$$

where $x(t)$ is the state of the system at time t, $u(t)$ is a known, time-dependent input and f is a mathematical model of the system. In addition it is in most cases not possible to measure the system state x directly, but only observables y via a observation function

$$y(t) = h(x(t)) \quad . \qquad (2)$$

It is more often the rule than the exception, that measurements y^{obs} and predictions of (1) and (2) do not satisfactorily coincide. If the discrepancies are due to stochastic processes or measurement noise, the known methods using varieties of Kalman Filters and Particle Filters, to name a few, will be able to reconstruct the states $x(t)$ given the data y^{obs}. As soon as the model f contains systematic errors, such as unknown interactions, these methods tend to fail or become very unstable. The DYNAMIC ELASTIC NET (DEN)[1] is a dynamic optimization procedure that handles systematic model errors as unknown inputs to the systems equations. The methods provided by seeds are able to optimize the unknown inputs such to archive a satisfactory fit to given data.

2 Methods

The seeds package provides two distinct ways to compute systematic errors. Both make use of the Hamilton formalism, a convenient tool in dynamic optimization. As a setup for the DEN you must provide

1. a mathematical model f of the dynamic system,

2. an observation function h,

3. measured data y^{obs}.

The greedyseeds method is a deterministic algorithm based on a dynamic version of the method of steepest descent, combined with a "greedy" selection process to archive a sparse solution. In addition to the systems equations it is possible to constrain the system to the solution space of algebraic equations, e.g. mass conservation. In particular

the case

$$\frac{\partial f_i}{\partial x_i} = 0 \quad , \tag{3}$$

where the index is understood as the i-th component, may lead to certain numerical difficulties and highly benefits of such an algebraic constraint.

The bayesianseeds method is a stochastic algorithm based on ideas of bayesian inference and uses LASSO regularization to generate sparsity of the solution [2]. In addition to a specific estimate of the unknown inputs it results in a probability density in the space of unknown input functions.

3 Example

As an exemplary scenario we take a look at the 13-dimensional UVB system and the greedyseeds. You can find the R notation of systems equations, observation function and a data set in the package. Figure 1 shows
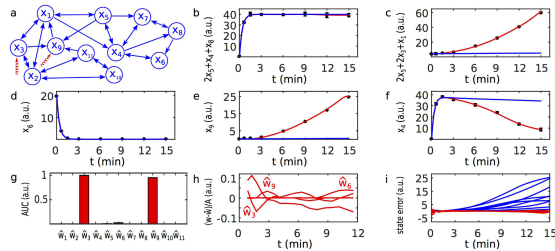

Figure 1: Results of the greedyseeds algorithm. The black markers represent the data y^{obs}, the blue lines represent the erroneous model predictions and the red lines show the predictions using the estimated unknown inputs $\hat{w}_i$.

the data and the five observables with erroneous system predictions as black markers

and blue lines, respectively. It also shows the estimated unknown inputs $\hat{w}_i$, the Area-Under-Curves and the predictions of the corrected model in red. You see that the algorithm results in a sparse set of unknown inputs and how the unknown inputs are able to correct the model predictions such that these fit the data.

4 Conclusion

As the example shows, the R package seeds is able to correct systematic errors in dynamic model on the level of the systems equations based on optimization principles of the DEN. It is able to detect the errors, and to simultaneously estimate the errors and the system state. We want to point out that, depending on the properties of the system, cases may occur where error are not observable. However the provided algorithms have proven to yield good results in various realistic problems and may be a fruitful tool whenever the lack of a precise model inhibits a satisfying state estimation.

References

[1] Benjamin Engelhardt *et al.* "Learning (from) the errors of a systems biology model". In: *Scientific Reports* (2016).

[2] Benjamin Engelhardt, Maik Kschischo, and Holger Fröhlich. "A Bayesian approach to estimating hidden variables as well as missing and wrong molecular interactions in ordinary differential equation-based mathematical models". In: *Journal of The Royal Society Interface* (2017).