The importance of being computationally accurate: the case of covid-19

Christina C. Christara
University of Toronto
Toronto, Ontario M5S 3G4, Canada
ccc@cs.toronto.edu

Abstract

This winter term went through some rough times towards its end, due to the sudden outbreak of covid-19. This paper serves as end-of-year contribution to my CSC336 (winter term) students, as well as to all Computer Science students, who, in the absence of any other end-of-year celebration, plan to publish an electronic yearbook. Instead of some general (academic or practical) advice, I present some thoughts on the importance of accuracy of computational results, in hope you will find them useful and interesting. If not anything else, they are timely, as they are driven by the various discussions on mathematical models for covid-19.

1 Introduction

Numerical methods are an indispensable tool for solving many mathematical models of real world phenomena. With the wide availability of (variable quality) numerical software, it is relatively easy to get numerical solutions to various mathematical problems. What is harder and requires deeper knowledge is the appropriate use of numerical methods and software and the correct judgement of the results. Driven by the outbreak of covid-19, we take on an example of a simple epidemic model and demonstrate how different numerical methods may give rise to quite different numerical results. In Section 2, we briefly present a model describing the evolution of an epidemic by a system of Ordinary Differential Equations (ODEs). In Section 3, we summarize the general ideas and properties of numerical methods for ODEs, we well as some features of modern software implementing the methods, while in Section 4, we present results of some simulations using numerical ODE methods, and raise the readers' attention to important issues related to computational accuracy. We conclude in Section 5.

2 Mathematical models for epidemics

A widely used mathematical model for epidemics is based on differential equations, more specifically, a system of Ordinary Differential Equations (ODEs). Differential equations describe the change of unknown quantities in terms of other known and unknown quantities. Here is a typical system of ODEs modelling epidemics:

$$\frac{dS}{dt} = \mu N - \mu S - \frac{\beta}{N} I S \quad (1)$$

$$\frac{dE}{dt} = \frac{\beta}{N} I S - aE - \mu E \quad (2)$$

$$\frac{dI}{dt} = aE - \gamma I - \mu I \quad (3)$$

$$\frac{dR}{dt} = \gamma I - \mu R \quad (4)$$

The above model is referred to as SEIR, with $S(t)$, $E(t)$, $I(t)$ and $R(t)$ denoting the (unknown) numbers of Susceptible, Exposed, Infectious and Recovered, respectively, species at a time point t, while μ, β

and γ are the replenishment, transmission and recovery rates, respectively, a^{-1} is the average incubation period, and N is the total number of species. All of μ, β, γ, a and N are assumed to be given (known). The initial state $S(0)$, $E(0)$, $I(0)$ and $R(0)$ is also assumed to be given, and the model predicts future states. This type of problem is refereed to as an Initial Value Problem (IVP) for ODEs.

Just as any mathematical model for a real problem, the SEIR model makes several assumptions and simplifications. For example, it assumes that N remains constant, that μ, β, γ and a are the same for all individuals, that all recovered are immune for life, and that all newborns are susceptible.

The model can be extended to include, for example, different age compartments (this will result in more ODEs in the system), different μ, β, γ and a for different compartments, different spatial compartments (to model rural/isolated areas in a different way from urban/dense areas), or to introduce spatial diffusion, which will turn the model into a Partial Differential Equation (PDE), undetected and detected infectious individuals, quarantined and non-quarantined infectious individuals, vaccinated and non-vaccinated individuals, and so on. Once the outbreak of covid-19 started, researchers rushed to get predictions using variations of ODE models [4, 2]. Others use a data driven approach [3], resulting in a nonlinear least squares problem, while some use combinations of machine learning and ODEs [1].

The aim of the paper is not to give a comprehensive description of models for epidemics, so we will assume the above simple SEIR model.

3 Numerical/computational methods

Numerical methods for IVP-ODEs approximate the solution of IVP-ODEs by generating a sequence of approximate states at a set of increasing points in time. Explicit methods compute a new state using simple relations between old (already computed) states. Implicit methods use relations between the new and old states to compute a new state, resulting in solving linear and nonlinear systems at each timestep.

In general, the smaller the timestep size, the smaller the error of the approximation. But too small timestep sizes, result in large number of timesteps, for a fixed horizon of simulation, therefore, the simulation becomes inefficient.

Methods exhibit a certain order of convergence, that is, rate by which the error is decreased, as the timestep size is decreased.

Furthermore, methods incur a certain computational cost per timestep, with higher order methods often being more expensive than lower order, and implicit methods being more expensive than explicit.

All the properties of the problem need to be studied and all the properties of the method need to be considered in order to find a appropriate problem and method pair for accurate and efficient simulations.

Modern computer implementations of numerical methods for IVP-ODEs involve more features, such as *adaptive timestep size selection* (so that small step sizes are used when the computed quantities vary a lot, and large for slowly varying quantities), *error control* (so that the user sets a tolerance and the software adjusts the parameters of the method so that the error is guaranteed to stay below the tolerance and the simulation be most efficient), as well as *event detection* (techniques that allow the user to declare "events" when some parameters of the problem and/or method are adjusted depending on the needs of the problem).

Computational methods for IVP-ODEs are included in CSC436 (Numerical Algorithms), while methods for solving linear and nonlinear systems are studied in CSC336 (Numerical Methods). Numerical methods for linear least squares problems (with a touch on nonlinear least squares) are part of CSC436.

4 Numerical results

We take on the example of system of ODEs (1)-(4) with a given initial state and given values of μ, β, γ, a and N, and carry the simulations with the MATLAB implementations of `ode15s` and `ode45`, two numerical methods for ODEs. The former uses a variable order method (orders 1 through 5), adaptively chosen by the software depending on the properties of the problem, and is designed for "stiff" problems, i.e. problems that involve rapid change of computed quantities in short periods of time. The latter uses a pair of methods of orders 4 and 5, is designed for the more standard problems, and is one of the most popular methods used. Both implementations include adaptively chosen step size, and error control, for balancing efficiency and accuracy.

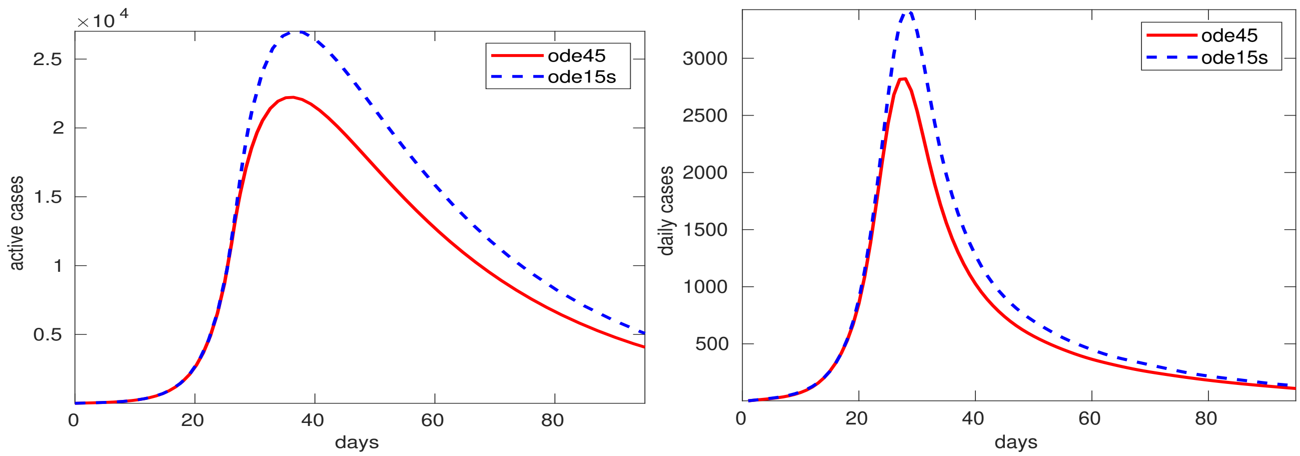

Figure 1: Total active and daily infectious cases computed by `ode45` and `ode15s`. Simulation for 95 days.

In Figure 1, we see results of simulations by the two methods applied to (1)-(4) with the same initial conditions, and with the same parameters μ, β, γ, a and N. In both simulations, we assume that, after the 27th day, the transmission rate decreases by a certain factor, due to social distancing measures. The "trend" of the results for both (total) active and (new) daily cases is the same. The cases increase sharply, following the so-called exponential growth, then decrease, due to social distancing measures. The daily cases start decreasing immediately after the measures take into effect, while the active ones start decreasing a little later, due to the incubation period. If one looked at only one of these simulations (only one method), one would be satisfied of the results, as the look "reasonable".

However, having them together in single plots, demonstrates how badly `ode45` misses the rapid exponential growth after some time, as well as the value (and even time) of peaks. Consider that several critical decisions, such as decisions on the number of ICU beds, quantities of medication and medical material stock, number of first responders, and more, that may affect people's lives, depend on results of such simulations. I emphasize that both methods and implementations are correct. Both methods and implementations are reputable, and useful (each one on specific classes of problems). It is the knowledge on the specifics of the problem that can lead the educated user to the correct choice. Or, if such a knowledge is not already acquired, the consistent questioning, research and effort for deep understanding can eventually lead to the desired outcome.

During my last electronic class meeting this term, one your fellow students asked me whether I have any "cool" simulations I could show. What is cool and what not is certainly subjective, but, in any case, I will show simulations for disease outbreak under the assumption that the government takes social distancing measures when the number of exposed persons goes above a given limit (taking into account the capabilities of the health system), but eases the measures when the number of exposed persons goes below a given limit (taking into consideration the negative effects of measures to people, the economy, etc). Such a simulation can be done using the event detection abilities of modern numerical ODE software.

Figure 2 presents the results, with the upper limit set to 25,000, and the lower to 10,000. Notice that the number of infectious rise above the exposed set limit, due to the incubation period. Also, note it takes longer to reduce the exposed below the limit, when the number of infectious is higher, as, the more the infectious, the more the spreading. Such a situation can go for very long, unless an effective medication is found, which will increase the recovery rate, thus avoiding the large exponentional growth in the absence of measures, or unless a vaccine is found, which will significantly change the model and give rise to more prosperous results.

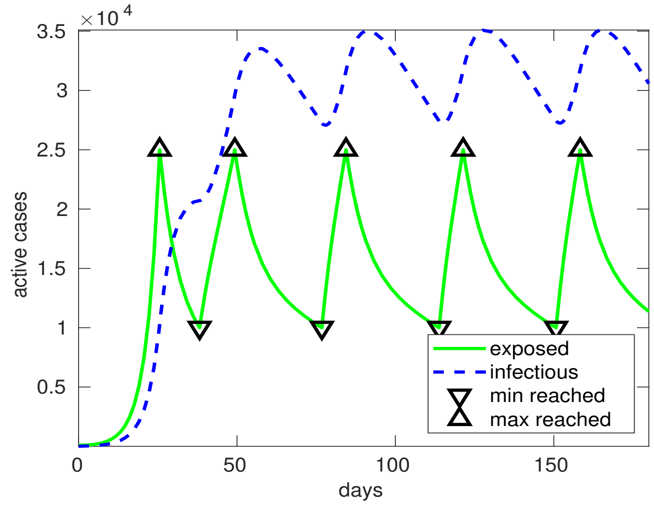

Figure 2: Total active exposed and infectious cases computed by `ode15s` with event detection. Simulation for 180 days.

5 Conclusions

We have shown examples of simulation of an epidemic outbreak using a simple ODE model and certain numerical methods and associated software features. Attention is raised on issues of computational accuracy. Knowledge on problem and method properties, and continuous research provide the best advice for which problem and method pair gives the most accurate results.

Disclaimer: The aim of this paper is not to provide predictions for covid-19.

References

[1] R. Dandekar and G. Barbastathis, *Quantifying the effect of quarantine control in Covid-19 infectious spread using machine learning*, (2020), pp. 1–13. Preprint, http://www.doi.org/10.1101/2020.04.03.20052084.

[2] B. Ivorra, M. Ferrandez, M. Vela-Perez, and A. Ramos, *Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections: the case of China*, (2020), pp. 1–28. Preprint, http://www.doi.org/10.13140/RG.2.2.21543.29604.

[3] C. J. Murray, *CurveFit tool for Covid-19 model estimation*, (2020), pp. 1–4. Preprint, IHME COVID-19 health service utilization forecasting team https://www.medrxiv.org/content/medrxiv/suppl/2020/03/30/2020.03.27.20043752.DC1/2020.03.27.20043752-1.pdf.

[4] A. R. Tuite, D. N. Fisman, and A. L. Greer, *Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada*, (2020), pp. 1–9. Early release, https://www.cmaj.ca/content/cmaj/early/2020/04/09/cmaj.200476.full.pdf.