Data-driven Epidemic Inference Through Decomposition of Large-scale Nonlinear Optimization Problems

Laurens R. Luega, Michael L. Bynumb, Carl D. Lairda\*, Lorenz T. Bieglera

aChemical Engineering Department, Carnegie Mellon University, Pittsburgh, PA, USA

bSandia National Laboratories, Albuquerque, NM, USA

claird@andrew.cmu.edu

Abstract

The COVID-19 pandemic has demonstrated that a systematic understanding of the dynamics of infectious diseases is crucial to reduce their impact on a population. To this end, the use of epidemiological models has proved effective for both parameter estimation and epidemic control problems. In this work, we formulate a compartmental epidemic model on a US national scale with county-level granularity and fit relevant spatio-temporal parameters to data collected during the early stages of the COVID-19 pandemic. Conventional, serial solvers struggle to solve large-scale, nonlinear optimization problems such as this within reasonable time limits. Instead, we exploit the problem’s structure to apply a Schur complement decomposition. This allows for the computationally most expensive operation in an interior point method, the computation of the step direction, to be parallelized. We demonstrate promising scaling properties of this method when applied to large-scale epidemic inference problems on a moderately-sized parallel computer.

**Keywords**: large-scale nonlinear optimization, parallel computing, decomposition, epidemic modeling

* 1. Introduction

The outbreak of the COVID-19 pandemic had an unprecedented impact across the world. Significant excess deaths, stress on medical infrastructures and sustained economic fallout were amongst the most acute direct consequences. Toward the goal of guiding public health responses, epidemiological models are an important tool to understand the dynamics of infectious diseases and plan mitigation strategies accordingly. In this work, we present efficient computational approaches to solve large-scale parameter estimation problems arising from epidemic models.

* 1. Methodology

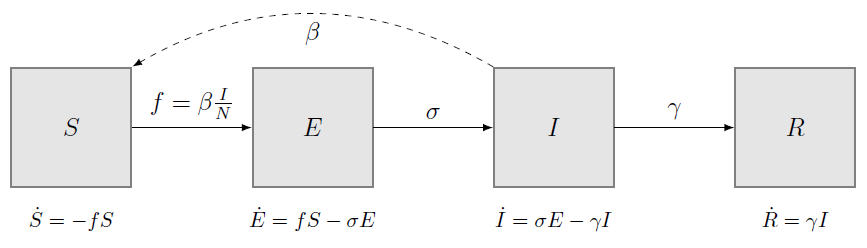
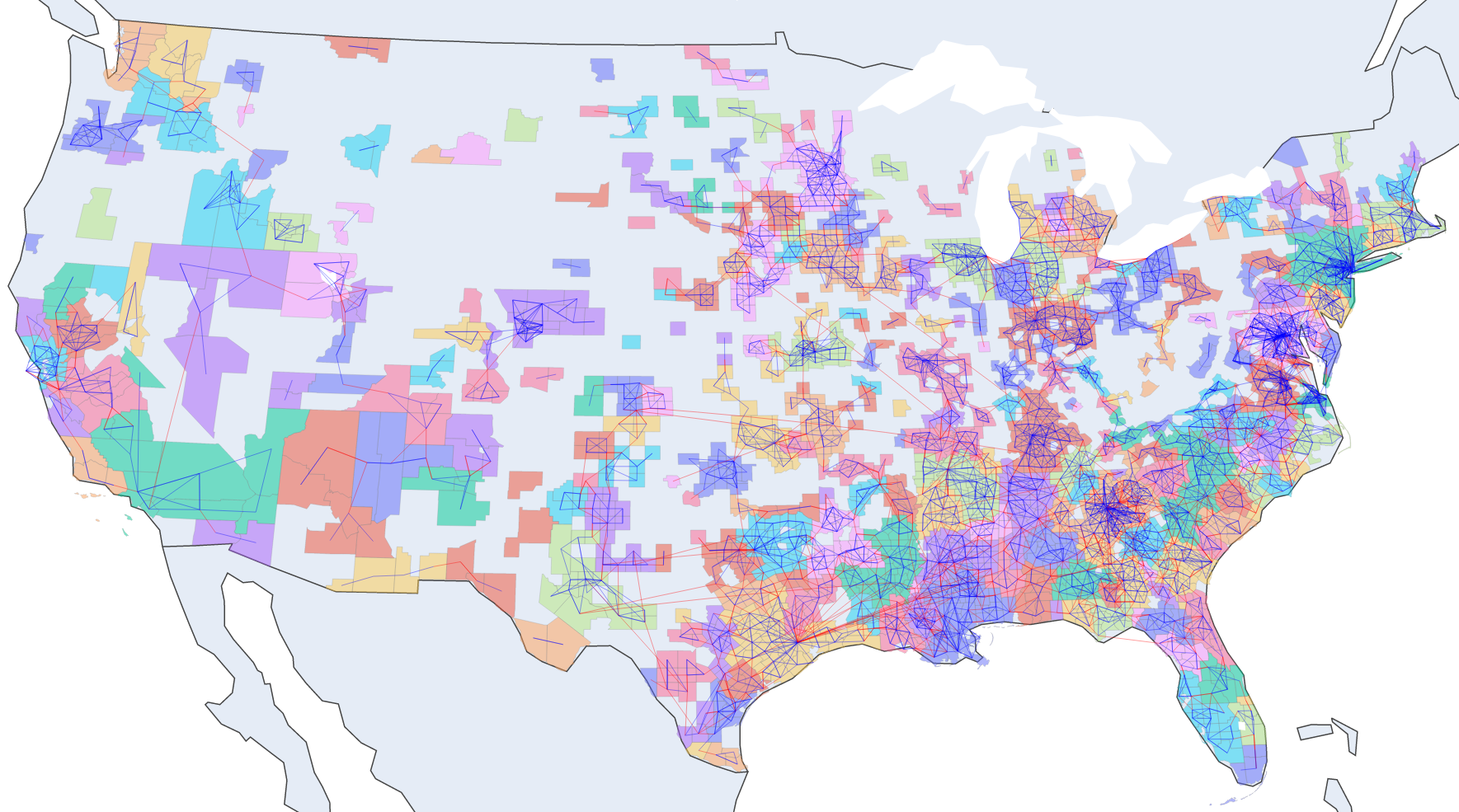


Figure 1: SEIR model for a single population. Rectangles depict the compartments, solid arrows, the movements of individuals between them. The dashed arrow visualizes contact between compartments, leading to transmission of the disease.

Compartmental models have proved effective in describing the spread of various infectious diseases (Rock et al., 2014). The population is divided into compartments, which represent different stages of a disease, i.e., susceptible (*S*), exposed (*E*), infectious (*I*) or recovered (*R*). The dynamics of individuals between compartments over time are determined by the contact, incubation, and recovery rates (β, σ and γ, respectively), which may vary across time and space. For a visualization of this dynamic model for a single population see Fig. 1.

In this work, we applied a large-scale, model-based approach to estimate epidemic model parameters based on policy-related descriptors. To reflect the spatial heterogeneity in the US, separate epidemiological compartments are defined on a county level. Interaction between counties is modeled using information on mobility patterns, in this work we use census data on commuting flows (US Census Bureau, 2020). The contact rate parameters for each county are defined as piecewise-constant functions, depending on the implementation of public health responses, such as mask mandates, over time in each county. Additional terms accounting for demographic or seasonal factors are possible. This defines a large-scale, nonlinear system of ordinary differential equations with ~104 differential state variables (epidemiological compartments for ~3000 counties in the US), discretized over a time horizon of several months. The parameters of this model are fitted to case data collected during the beginning of the COVID-19 pandemic.

Optimization problems such as this are usually solved using interior point algorithms. Serial implementations struggle to solve problems of this scale within reasonable time limits. Instead, we exploit the problem’s structure to apply a Schur complement decomposition (Kang et al., 2014). This allows for the computationally most expensive operation in an interior point method, the computation of the step direction, to be parallelized. The epidemic model at hand is amenable to this decomposition, as there are only a few shared variables between counties, i.e., the factors associated with the effect of different non-pharmaceutical interventions (NPIs) on the local contact rates. Below, the promising scaling properties of this method are demonstrated.

* 1. A graph of a graph with a green line and a red x

     Description automatically generated Results

(a) (b)

Figure 2: (a) Map of counties considered for national scale model. Shading is based on the spatial decomposition of counties for parallelization. (b) Strong scaling of parallel Schur complement method. Speedup computed relative to two core execution. Solution time using *Ipopt* solver is shown for serial case.

The parameter estimation problem was solved simultaneously for 1900 counties in the US (see Fig. 2a) over the first 170 days of the pandemic. The full 3142 counties of the US were not considered because those with fewer total cases than 0.1% of the county population during this period were discarded. All parameters other than β were fixed to common values from literature (Yang et al., 2021). Three NPIs were included in the model: mask mandates, school closings and stay-at-home orders, with state-level implementation dates taken from the sources listed in Yang et al. (2021). All tests were performed using an x86 machine with 32 cores, each with a clock rate of 3.3 GHz. The parallel Schur complement method was applied to this problem using the *parapint* package (Rodriguez et al., 2023). The scaling of the solution time with an increasing number of parallel processors was observed, with results shown in Fig. 2b. In our model, each of the NPIs considered in this work reduced the contact rate by 11-15% when implemented in the respective counties. These values are reasonably close to statistical estimates listed in other works (Yang et al., 2021), but come with yet unquantified confidence intervals stemming from both model and measurement uncertainty.

* 1. Conclusion

We use a dynamic compartmental model for COVID-19 with county-level granularity, which considers inter-county infections through commuters and quantifies the effect of non-pharmaceutical interventions (NPIs) on the contact rate between individuals. The parallel Schur complement method was applied to solve this large-scale nonlinear programming problem on a US national scale, with significant speedup when using 32 cores, compared to the serial state-of-the-art solver *Ipopt* (Wächter and Biegler, 2006). The qualitative results suggest that the implementation of NPIs had an adverse effect on the spread of COVID-19. Uncertainty quantification for these results remains as future work.

Acknowledgements

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC (NTESS), a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy’s National Nuclear Security Administration (DOE/NNSA) under contract DE-NA0003525. This written work is authored by an employee of NTESS. The employee, not NTESS, owns the right, title and interest in and to the written work and is responsible for its contents. Any subjective views or opinions that might be expressed in the written work do not necessarily represent the views of the U.S. Government. The publisher acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this written work or allow others to do so, for U.S. Government purposes. The DOE will provide public access to results of federally sponsored research in accordance with the DOE Public Access Plan.

References

J. Kang, Y. Cao, D. P.Word, and C. D. Laird, 2014. An interior-point method for efficient solution of block-structured nlp problems using an implicit schur-complement decomposition. Computers & Chemical Engineering, 71:563–573.

K. Rock, S. Brand, J. Moir, and M. J. Keeling, 2014. Dynamics of infectious diseases. Reports on Progress in Physics, 77(2):026602.

J. S. Rodriguez, R. B. Parker, C. D. Laird, B. L. Nicholson, J. D. Siirola, and M. L. Bynum, 2023, Scalable parallel nonlinear optimization with pynumero and parapint. INFORMS Journal on Computing, 35(2):509–517.

United States Census Bureau, 2020 Census Results, retrieved in July 2023. URL: https://www.census.gov/data/tables/2020/demo/metro-micro/commuting-flows-2020.html

A. Wächter and L. T. Biegler, 2006, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, Mathematical programming, 106:25–57.

B. Yang, A. T. Huang, B. Garcia-Carreras, W. E. Hart, A. Staid, M. D. Hitchings, E. C. Lee, C. J. Howe, K. H. Grantz, A. Wesolowksi, et al., 2021, Effect of specific non-pharmaceutical intervention policies on sars-cov-2 transmission in the counties of the united states. Nature communications, 12(1):3560.