Journal of Statistical Software

*October 2015, Volume 67, Book Review 1.* *doi: 10.18637/jss.v067.b01*

Reviewer: James P. Howard, II
University of Maryland University College

---

**Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data**

J. Nathan Kutz
Oxford University Press, Oxford, 2013.
ISBN 978-0199660346. 656 pp. USD 49.95 (P).
https://global.oup.com/academic/product/9780199660346

---

J. Nathan Kutz's *Data Driven Modeling & Scientific Computation* is a new text presenting scientific computing methods in MATLAB. Unlike other scientific computing books, Kutz also takes on the broader topic of data science methods. The author, in his introduction, states two goals for his text. First, the text should be suitable for advanced undergraduate or introductory graduate instruction. Second, the book should be usable as a reference text for practitioners. Kutz divides the text into four parts, really representing two fields.

The first part, "Basic Computation and Visualization," comprises six chapters and is keyed to a traditional course in numerical analysis, excepting differential equations. Kutz provides a single chapter on MATLAB, linear systems, curve fitting, differentiation and integration, optimization, and visualization. The rapid introduction to MATLAB's key features in the first chapter provides a good grounding for what is to come. Additional chapters provide a course in numerical analysis, much of it around data science applications. The second part, "Differential and Partial Differential Equations," is five additional chapters on differential equations. There are independent chapters on boundary value problems, finite differences, stepping schemes, spectral methods, and finite element methods. Together, these chapters provide the introduction to differential equations missing from part one, but at substantially greater detail.

In the second half of the book, Kutz moves into data science methods. Part three, "Computational Methods for Data Analysis," provides twelve additional chapters focusing on data analysis techniques. Opening with a chapter on statistical methods, Kutz rapidly moves into image processing, sensing, and dynamical systems. In addition, Kutz weaves in additional numerical analysis topics through applications, such as singular value decomposition for principal component analysis. The final part, "Scientific Applications," brings three closing chapters that provide high level discussions of applied problems ranging from celestial mechanics to image denoising to musical analysis.

Kutz's strength lies in the elegance of his mathematical exposition. For instance, the chapter on statistical methods includes a brief, though understandable, review of hypothesis testing.

In another example, linear programming is explained before the method solutions are presented for consumption. Through this, there are refreshers for material that may be somewhat distant for some readers. In addition, these discussions provide just enough background to support the scientific computing discussion without forcing the reader to review other texts.

In addition, Kutz covers other topics that are normally overlooked in both scientific computing and data science courses. One of these focuses on data assimilation and how to handle the dirty problems of real world data acquisition. Reminding the reader, and especially the student, of the real challenges in many data science projects is important since understanding the sources of data is often skipped in data science treatments.

This book's largest problem is that it attempts to accomplish too much. Kutz initially focuses on scientific computing, which is grounded principally in numerical analysis. He then shifts to an introduction to data science, a discipline grounded in statistics. The overlap is not built in to the material. In some ways, the author has written one book when two would have served better. Both areas are handled relatively well, but the split focus leaves out important details.

A book on scientific computing is necessarily a cookbook that should be rife with working examples and functioning code. Of note, in the chapter on integration, an implementation of basic methods, such as the trapezoid rule, is not provided. As an introductory text, the material comes up short and instead relies upon the reader to the turn formulae and description into functioning code. Neither upper level undergraduate nor first year graduate student is likely to possess that command of both MATLAB and numerical analysis, in an introductory course. Because of this, students will need additional introductory material to read side-by-side with this book. More code examples are provided in the data science chapters, though.

In addition, and most glaringly, there is no systematic treatment of numerical error. Modern computers and double precision floating point numbers place numerical error just outside the reach of most applications. However, someone with a background in scientific computing should fairly understand the limits of the computational engine just as she is familiar with the limits of a graduated cylinder. Concepts of error are included, however, with the algorithmic discussions, though formal analysis is infrequent.

For these reasons, the book would not work in a scientific computing classroom without substantial supplementary material. Implementation details, and the nuance of the language, whether MATLAB, Java, C, or something else, is the hangup for students. Glossing these over at the start leaves the implementation opaque. The advantage, though, is that book is not so intimately tied to MATLAB as to be unusable to learn scientific computing in other languages, depending on the supplementary material provided. In an applied data science class, one focusing on specific examples, the text provides a stronger base.

However, as a standard reference, for those who already possess that command of the material, the book is outstanding. It provides cleanly and succinctly the basic material from which someone familiar with numerical analysis and statistical techniques can create functioning applications. One of the clearest examples of this is the review of image recognition, using cats and dogs as the example.

Of final note, and only because it is so extreme, it is worth mentioning that the price of the paperback edition, under USD 50 (and EUR), is very low. This alone makes the book a standout and stronger option for students. As faculty see increased pressure to reduce textbook costs, Kutz and the Oxford University Press have presented a sufficient text at an excellent price.

Kutz provides a usable text that can be used as a standard reference for MATLAB users and users of other applications, too. On the shelf, it is an excellent complement to other reference texts in the fields. However, those texts should include functioning code examples, such as Press, Teukolsky, Vetterling, and Flannery (2007). The book can be used in the classroom, but the instructor will need to provide additional supplementary material to support students who need more concrete examples to assimilate the material.

## References

Press W, Teukolsky S, Vetterling W, Flannery B (2007). *Numerical Recipes*. 3rd edition. Cambridge University Press, Cambridge.

## Reviewer:

James P. Howard, II
Mathematics and Statistics
University of Maryland University College
E-mail: jh@jameshoward.us or jphoward@faculty.umuc.edu
Twitter: @howardjp
URL: http://jameshoward.us/

*Journal of Statistical Software* http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
October 2015, Volume 67, Book Review 1 *Published:* 2015-10-06
doi:10.18637/jss.v067.b01