Review for *The End of Error: Unum computing* by John L. Gustafson

Frédéric Goualard

The flaws and idiosyncrasies of floating-point arithmetic are well known and their proper handling constitutes a sizable portion of any curriculum on Numerical Analysis. Even though many schemes have been proposed as (partial) replacements—Interval Arithmetic [MR2482682] being one of the better known contenders—, none has displaced floating-point arithmetic so far, it being so ingrained in the fabric of modern processors and of standard numerical algorithms.

With his book "The end of error: Unum Computing", Dr. Gustafson strives to introduce unums as a complete replacement for floating-point numbers, which he claims to avoid all the pitfalls of floating-point arithmetic while being mostly compatible with it. A large part of the book is also devoted to presenting how to best use unums to solve nonlinear as well as linear problems.

As stated by the author himself in the first pages, the book is written for the (educated) "layman" rather than specifically targeting numerical analysis experts. That intent shows in several ways:

- Many examples and pictures in color illustrate the text, which greatly enhances the readability of the book as whole;
- Jargon is mostly avoided throughout, which makes for a more pleasant read; for better or for worse, Dr. Gustafson is true to his goal as he also successfully refrains from using any academic tone, even sprinkling a good dose of jokes—albeit of poor taste for some of them, I might say (see Section 14.2's title for an example). Dr. Gustafson is also no stranger to provocative statements and shortcuts, the very title of the book being the first testimony to that.

The book is organized into three parts:

- Part 1 presents unums, the new format for arithmetic on reals, and their environment;
- Part 2 explains in great details how to modelize problems from various fields (engineering, physics, . . .) in order to best use unums strengths;

- Part 3 is a set of appendices, which present the prototype of an unum-based arithmetic library written in Mathematica.

Part 1 makes for an interesting read, as it presents clearly the sources of the problems arising from floating-point arithmetic use, and how unums avoid them. For all intents and purposes, unums may be considered as better IEEE 754 floating-point numbers (with only two NaNs) whose exponent and fractional part may vary in size, together with an inexact bit stating whether the object represents an exact result (a rational) or an inexact result (a canonical interval that does not contain its bounds). Using pairs of unums ("ubounds"), it is possible to construct a closed arithmetic with many similarities to interval arithmetic, with the added benefit of being able to represent domains with less bits than in interval arithmetic by adjusting the size of the unums used to the precision required. On that subject, Dr. Gustafson makes several harsh comments regarding the inability of "naive" interval arithmetic to represent domains with open bounds, which disregards superbly the entire body of work on interval arithmetic with closed/open bounds (See, for example: J. G. Cleary. Logical Arithmetic, Future Computing Systems, 2(2), p. 125–149, 1987) as well as complete systems like Prolog IV, which supported intervals with closed/open bounds represented by either floating-point numbers or arbitrary precision rational numbers. Unfavorable comparisons with unums/ubounds are also made regarding the inability of interval arithmetic to solve difficult problems tailored to exhibit floating-point pathological behavior; however, Dr. Gustafson always considers naive interval algorithms only, disregarding once more all the better methods that have been developed in the interval arithmetic community. Despite the unfairness of these comparisons, Part I is still commendable for the breadth of its presentation, as it describes unums from the user point to view down to the definition of the operators—with links to Mathematica code in the appendices—and their efficient implementation in hardware.

Part 2 introduces a naive—in a non-derogatory way—, versatile, and highly parallelizable algorithm based on "uboxes" (Cartesian product of unums) to solve nonlinear and linear problems. The algorithm reconstructs systematically the solution space from a seed, taking advantage of the variable precision afforded by unums. Physics problems (pendulum, n-body problems) are treated at length, from their modeling to their solving. At its root, the algorithm presented in this part is not new as it is akin to a tessellation with labelled boxes (true/false/unknown) of the search space that could be performed with interval arithmetic. The use of unums make it much more efficient through variable precision of the bounds. As in Part I, Dr. Gustafson fails to consider related algorithms from the interval arithmetic community (e.g. SIVIA and friends [MR1989308]) when assessing the relative merits of interval arithmetic and unums.

Part 3 describes the implementation in Mathematica of unums and their operators. It should be of interest only to readers committed to understand the finest points and to implementors who would like to design unum-based libraries in other languages.

Even though this book presents radical ideas that often depart from existing works, the bibliography still appears tellingly terse, for a 400 pages book, at only fourteen entries. I believe that major works that have been overlooked in Part 1 and Part 2 would have deserved their own entries (see at least the aforementioned references). On the other hand, the index is very rich and greatly facilitates searches in the paper version of the book.

Overall, this book is both intriguing and interesting as it addresses profound problems in a somewhat novel way. Even though it has some few distracting typos that might hinder the understanding of casual readers (there is an errata sheet on the publisher's webpage for the book, along with the archive for the Mathematica code of the prototype), it is, overall, written with care and rigor as for the mathematical aspects. More annoying are the repeated charges on interval arithmetic, which never compare state-of-the-art algorithms against unum-based methods. Also troubling is the fact that this book does exist in a vacuum: to my knowledge there has been no independent work at all on unums so far. To the credit of Dr. Gustafson, he has taken the trouble of offering a prototype alongside the book for readers to play with. It is however difficult to assess the true merits of the approach with it, as it is written in Mathematica—a tool not available to all interested, and sufficiently slow to forbid trying solving large problems with it with performances in mind. Only time will tell if this book is a premature foray into a dead-end or the seminal source of many better numerical libraries.