Benchmarks

In the past five years I have been involved as a reviewer and author in a wide range of conferences, including the American Control Conference (ACC), the IEEE Conference on Decision and Control (CDC), AIAA Guidance Navigation and Control (GNC), the International Conference on Robotics and Automation (ICRA), and the International Conference on Intelligent Robots and Systems (IROS), as well as Neural Information Processing Systems (NIPS), the International Conference on Machine Learning (ICML), and the Conference on Uncertainty in Artificial Intelligence (UAI). This has led me to reflect on the different expectations for papers written in various fields.

Benchmarks can be in the form of dynamic system models or standardized image databases.

In particular, my experience is that most conferences in the fields of inference and planning systems require proposed algorithms to be compared with other techniques, which are hopefully the state of the art, on both common benchmark problems and/or published data sets. Failure to do so will result in almost-certain rejection of the paper. Typical benchmark problems are readily accessible and easily coded, and there are generally agreed-upon settings of parameters that make the problems interesting.

Benchmarks can be in the form of dynamic system models or standardized image databases [1]. For example, the reinforcement learning community has numerous benchmark problems, such as the mountain car [2], blocks world [3], and grid world [4]. Though often derided as "toy problems," some of the benchmarks, such as the helicopter benchmark [5], are quite complex and more realistic.

A similar approach is taken in other related fields, such as the development of optimization algorithms. For example, the Computational Infrastructure for Operations Research (COIN-OR) [6] repository usually provides head-to-head comparisons of each submitted algorithm on a range of challenging benchmark problems.

With a steady progression of papers using similar benchmarks, it is relatively easy for authors to demonstrate the contributions of their work and for reviewers to assess those contributions. Furthermore, it is typically easier for other researchers to determine the limitations of the techniques that exist and whether further research might be required. Thus, not only is benchmarking perceived to be a good idea, the culture within the community, enforced through the review process, is that authors are expected to do this comparison. While "over designing" the algorithm to just solve the problem posed and not generate a general solution technique is a potential problem, that is why several different benchmark problems that share core features are required to ensure that the algorithms are not fragile.

However, in reviewing ACC and CDC papers, the control community seems quite different. There is much less focus on comparisons with the state of the art. Furthermore, there is no clear consensus for the benchmark problems to use for comparisons. These observations appear to echo the International Federation of Automatic Control (IFAC) report from 1990 [7], which lists 13 benchmark problems in total, with the observation that "it was believed by the IFAC theory committee that it would be useful to have a collection of standard problems for comparing the benefits of 'new or existing control system design tools,' i.e. at present, every new design is applied to some ad hoc example, and it is difficult to determine a meaningful comparison between existing techniques." Given the effort that was exerted to collect those benchmark problems, it is somewhat surprising that, according to Google Scholar, the report has only been cited a total of 47 times in the past 25 years.

More recent examples of attempts to develop benchmark problems include the two mass-spring system that attracted a lot of interest from researchers looking at the complex (and real parametric) robust control problem associated with lightweight space structures. This benchmark led to several sessions at the 1992 ACC, with a special journal issue [8]. Based on the citation count, there has been good sustained interest in this simple problem. A related nonlinear control problem is the translational oscillations with a rotational actuator [9], a nonlinear, fourth-order dynamical

Digital Object Identifier 10.1109/MCS.2014.2364690
Date of publication: 19 January 2015

system. The missile autopilot has been quite broadly applied by the gain-scheduling and linear-parameter-varying control communities [10], [11].

Despite the existence of these benchmarks and more that I haven't mentioned, several observations and questions still arise. Can the control field develop and publish a set of benchmark problems that capture and clearly identify the current key challenges in the field, such as control of networked systems, modeling and control of financial systems, and control of nonlinear flexible aircraft? Good benchmark problems should be complex enough to highlight issues in controller design or implementation but simple enough to provide easily understood comparisons. Furthermore, can these problems be extended to include experimental control challenges as well, perhaps capitalizing on the recent proliferation of motion capture systems and quadrotor flight test beds?

Should the authors of papers in the control field be expected to compare their algorithms on a benchmark problem against the perceived state of the art, possibly leading to a culture in which papers that do not carefully and thoroughly compare the proposed work with the state of the art are deemed unacceptable by the community and, in particular, the paper reviewers?

Addressing these points should improve the quality of the papers presented at the conferences and in the journals, ultimately strengthening our community.

REFERENCES

[1] (2014, Nov. 20). [Online]. Available: http://www.vision.caltech.edu/archive.html

[2] A. Moore, "Efficient memory-based learning for robot control," Ph.D. dissertation, Univ. Cambridge, Cambridge, U.K., Nov. 1990.

[3] N. Gupta and D. S. Nau, "On the complexity of blocks-world planning," *Artif. Intell.*, vol. 56, no. 2, pp. 223–254, 1992.

[4] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[5] (2014, Nov. 20). [Online]. Available: http://library.rl-community.org/wiki/Helicopter_(Java)

[6] (2014, Nov. 20). [Online]. Available: http://www.coin-or.org/

[7] E. J. Davison, "Benchmark problems for control system design: Report of the IFAC Theory Committee," Int. Federation Autom. Contr., 1990.

[8] B. Wie and D. S. Bernstein, "Benchmark problems for robust control design," *J. Guid. Control Dyn.*, vol. 15, no. 5, pp. 1057–1059, 1992.

[9] D. S. Bernstein, Ed., "Special issue: A nonlinear benchmark problem," *Int. J. Rob. Nonlinear Contr.*, vol. 8, nos. 4–5, pp. 307–310, 1998.

[10] R. T. Reichert, "Dynamic scheduling of modern-robust-control autopilot designs for missiles," *IEEE Control Syst.*, vol. 12, no. 5, pp. 35–42, 1992.

[11] J. S. Shamma and J. R. Cloutier, "Gain-scheduled missile autopilot design using linear parameter varying transformations," *J. Guid. Control Dyn.*, vol. 16, no. 2, pp. 256–263, 1993.

Jonathan P. How

The Wright Contributions

Wright is a great name in aeronautics because sixty two years ago Wilbur and Orville demonstrated man carrying flight at Kittyhawk. This event is universally recognized; details of the mechanization are definitely known through design data and other records of the original flyer, which itself still exists in the Smithsonian Institute, whereas the nature of the inventions involved have long since been established for the legal record during the progress of many court actions. In my opinion, all these things are important, but, even in the aggregate, they do not truly represent the very significant ideological contributions of the Wright brothers to the progress of modern engineering and technology. The nature of these contributions appears not so much from what was achieved, but rather from the mental attitudes and methods that were applied to the attack, and first-time successful resolution of a long-term challenge to the imagination and ingenuity of mankind. … The genius of the Wrights lay in their ability to recognize the real world situations involved, to resolve these circumstances into clearly defined regions, and to pioneer effective means of attack on each separate problem while always keeping in mind the fact that all parts had to function as working components of the same over-all system.

— Charles S. Draper, "The Role of Informetics in Modern Flight Systems, 29th Wright Brothers Lecture," *AIAA Journal of Spacecraft and Rockets*, vol. 3, no. 6, pp. 769–779, June 1996