Performance evaluation of a novel implementation of NumPy library for numerical computing and machine learning
(Utvrđivanje brzine nove implementacije NumPy biblioteke za numeričko računarstvo i mašinsko učenje)

Ibrahim Numanagić
University of Victoria
Victoria, BC, Canada

Python is arguably the most popular programming language nowadays, and is thus widely used for scientific computing and machine learning. The central component of Python's scientific ecosystem is NumPy,[1] a library that provides an implementation of tensor (multi-dimensional array) data structure and a set of associated algebraic tensor operations. However, the base Python runtime is notorious for its lack of performance, and even NumPy, whose core is implemented in Fortran and C, is not as performant as it can be. We have recently developed Codon,[2] a static ahead-of-time compiler for Python, and a new highly optimized implementation of NumPy that is powered by Codon. In this project, we are looking to compare this new implementation with the reference NumPy implementation, as well as the other performant NumPy implementations (e.g., Cython,[3] Numba,[4] Pythran,[5] etc.) through the standardized benchmark sets such as NPBench.[6]

Tasks:

- Read and understand the literature
- Set up the various NumPy environments
- Implement and adapt NPBench and other benchmark suites in Codon
- Compare the results between Codon NumPy and other implementations
- Write a report detailing the improvements and the differences

References:

1. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
2. Shajii, A. *et al.* Codon: A Compiler for High-Performance Pythonic Applications and DSLs. in *Proceedings of the 32nd ACM SIGPLAN International Conference on Compiler Construction* (ACM, New York, NY, USA, 2023). doi:10.1145/3578360.3580275.
3. Behnel, S. *et al.* Cython: The best of both worlds. *Comput. Sci. Eng.* **13**, 31–39 (2011).
4. Lam, S. K., Pitrou, A. & Seibert, S. Numba. in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC* (ACM, New York, NY, USA, 2015). doi:10.1145/2833157.2833162.
5. Guelton, S. Pythran: Crossing the python frontier. *Comput. Sci. Eng.* **20**, 83–89 (2018).
6. Ziogas, A. N., Ben-Nun, T., Schneider, T. & Hoefler, T. NPBench. in *Proceedings of the ACM International Conference on Supercomputing* (ACM, New York, NY, USA, 2021). doi:10.1145/3447818.3460360.