Email address

mhaberla@calpoly.edu

Name of Submitter

Matt Haberland

Project Submitting this Proposal

SciPy

Is your project Affiliated or Sponsored?

Sponsored

Proposal Title

Add PROPACK Sparse SVD to SciPy

Two Sentence Summary of Proposal

Scientists and engineers worldwide rely on SciPy for sparse linear algebra computations, and SciPy in turn relies on the Fortran library ARPACK (ARnoldi PACKage) for sparse singular value decomposition (SVD); however, ARPACK is not always reliable, and it is not the most performant tool available. Our proposal is to remedy this by vendoring the BSD-licensed library PROPACK and exposing it to users as a solver option of `scipy.sparse.linalg.svds`.

Description of Proposal (<750 words, < 4500 chars)

The singular value decomposition (SVD) [1] is a matrix factorization technique with widespread applications in science, engineering, and machine learning, which often rely on its ability to identify the most "important" factors in data. The SciPy function `scipy.sparse.linalg.svds` [2] currently provides interfaces to two solvers that perform SVD on sparse matrices: ARPACK (ARnoldi PACKage) [3], a Fortran library that uses the Implicitly Restarted Arnoldi Method [4], and LOBPCG, a Python implementation of the Locally Optimal Block Preconditioned Conjugate Gradient method [5]. With nearly 10,000 code results on GitHub that refer to `scipy.sparse.linalg.svds` [6, 7], `svds` is one of SciPy's most important features, but as today's mathematical problems get larger and more challenging, there is a need for faster and more robust methods for performing sparse SVD. Since 2006 [8], SciPy's proposed solution has been to add a wrapper for PROPACK [9], a BSD-licensed, Fortran implementation of the Lanczos Bidiagonalization Algorithm with Partial Reorthogonalization [10]. However, even after nearly six years on the SciPy roadmap [11], this feature is not complete. With the aid of a Small Development Grant, we will fulfill this longstanding need for improved sparse singular value decomposition by adding PROPACK to SciPy. Specifically, we will:

- vendor PROPACK version 2.1,
- wrap PROPACK's essential Fortran routines using f2py,
- use the wrappers to add "`propack`" as a `solver` of `scipy.sparse.linalg.svds`,

- update the `scipy.sparse.linalg.svds` documentation to reflect PROPACK's unique solver options,
- add unit tests to verify the new functionality, and
- add benchmarks to compare the performance of PROPACK against the existing solvers.

References:

[1] Klema, Virginia, and Alan Laub. "The singular value decomposition: Its computation and some applications." *IEEE Transactions on automatic control* 25.2 (1980): 164-176.

[2] The SciPy Community. "scipy.sparse.linalg.svds. " *SciPy v1.6.1 Reference Guide,* https://docs.scipy.org/doc/scipy/reference/generated/scipy.sparse.linalg.svds.html. Accessed 19 February 2021.

[3] Lehoucq, Richard B., Danny C. Sorensen, and Chao Yang. *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. Society for Industrial and Applied Mathematics, 1998.

[4] Lehoucq, Richard B. "Implicitly restarted Arnoldi methods and subspace iteration." *SIAM Journal on Matrix Analysis and Applications* 23.2 (2001): 551-562.

[5] Knyazev, Andrew V. "Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method." *SIAM journal on scientific computing* 23.2 (2001): 517-541.

[6] Query results for "from scipy.sparse.linalg import svds". *GitHub.* https://github.com/search?q=%22from+scipy.sparse.linalg+import+svds%22&type=Code. Accessed 19 February 2021.

[7] Query results for "scipy.sparse.linalg.svds". *GitHub.* https://github.com/search?q=%22scipy.sparse.linalg.svds%22&type=Code. Accessed 19 February 2021.

[8] The SciPy Community. "A Wrapper for PROPACK (Trac #330) ". *scipy/scipy - GitHub*. https://github.com/scipy/scipy/issues/857. Accessed 19 February 2021.

[9] Larsen, Rasmus. M. *PROPACK – Software for Large and Sparse SVD Calculations*. http://sun.stanford.edu/~rmunk/PROPACK/index.html. Accessed 19 February 2021.

[10] Larsen, Rasmus M. "Lanczos bidiagonalization with partial reorthogonalization." *DAIMI Report Series* 537 (1998).

[11] The SciPy Community. "DOC: add new feature suggestions to the SciPy 1.0 roadmap." *scipy/scipy - GitHub*. https://github.com/scipy/scipy/commit/6c9c99113b81b4545ff70e3a527f41cc828947d9. Accessed 19 February 2021.

Benefit to Project/Community (<400 words, < 2500 chars)

This project would resolve a few of SciPy's issues and an unfinished PR on GitHub, and it addresses the item "Wrappers for PROPACK for faster sparse SVD computation" on SciPy's roadmap. Having maintainers familiar with each of SciPy's subpackages is important to SciPy's health; this project would add to the experience of Matt Haberland and Nicholas McKibben, a new core developer, in

`scipy.sparse`. Most importantly, it would improve the performance and reliability of one of SciPy's essential sparse linear algebra functions, benefitting both SciPy users and users of downstream projects (e.g. scikit-learn).

Amount Requested

$2,050

Brief Budget Justification - How will the money be spent?

$1,500 is for an estimated 30 hours of work by Nicholas McKibben at a rate of $50/hr. Nicholas' primary responsibility will be to write the wrappers for PROPACK, integrate it into `scipy.sparse.linalg.svds` function, implement unit tests, and update the documentation. ***NOTE: THESE FUNDS ARE NOT TO BE PAID THROUGH CAL POLY AND WILL BE PAID DIRECTLY BY NUMFOCUS TO NICHOLAS MCKIBBEN.***

$1,480 ($1,369 salary; $111 fringe benefits/payroll taxes) will compensate the proposer for 20 hours of overload work. The proposer's primary responsibilities will be to review the PROPACK interface written by Nicholas McKibben and to add benchmarks.

$570 is for Cal Poly recovery of indirect costs.

The Cal Poly salary and wage rates are based on the California Polytechnic State University (CPSU) and Cal Poly Corporation (CPC), jointly Cal Poly, established salary and wage rates paid during the 2020-2021 Fiscal year (July 1 – June 30). Benefits for CPSU Faculty summer and overload work include FICA, SUI and Workers Compensation and are calculated at the DHHS pooled rate of 8.1%. CPC undergraduate student benefits include SUI and Worker's Compensation. The DHHS pooled rate of 2.7% is used for budgetary purposes. Only rates in effect at the time the work is performed will be charged to the project. Cal Poly's federally negotiated indirect rate is 38.5% of modified total direct costs, effective July 1, 2019. Modified total direct costs exclude equipment, capital expenditures, charges for patient care, tuition remission, rental costs of off-site facilities, scholarships, and fellowships, participant support costs, and the portion of each subaward in excess of $25,000.

Timeline of Deliverables

4/15/2021 – Acceptance notification

6/15/2021 – Complete PROPACK wrappers and add PROPACK option to scipy.sparse.linalg.svds

7/15/2021 – Finish unit tests and documentation; review and merge implementation PR

8/15/2021 – Implement benchmarks; review and merge benchmark PR

9/15/2021 – Submit final report

Project Team

Matt Haberland, Nicholas McKibben

I agree to submit a grant report-back if my proposal is selected for funding.

I agree.