Name of Submitter

Matt Haberland

Your Email

mhaberla@calpoly.edu

Is your project Affiliated or Sponsored?

Sponsored

Select Your Project

SciPy

Proposal Title

Introducing Users to Powerful New Features of SciPy

Two Sentence Summary of Proposal

Several upcoming and recently released SciPy capabilities are featured as highlights in the release notes, yet currently we have only application programming interface (API) documentation to introduce users to their functionality. We propose to remedy this by 1) adding articles to the user guide for the new `scipy.sparse` arrays and matrix decompositions, the new mixed-integer linear programming features of `scipy.optimize`, and the new fitting and resampling methods of `scipy.stats`; and 2) spreading awareness about these powerful new features at the SciPy conference.

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

Out of hundreds of bug fixes and new features listed in each biannual release of SciPy, only a handful are selected as "Highlights" in the release notes. Often, these are features that have been requested many times and developed over the course of several years. These features have the potential to dramatically expand the types of problems SciPy's userbase can solve, but that potential may not be realized if the SciPy API Reference [1] is the only source of information about their use. This motivates the need for tutorials in the SciPy User Guide [2]: to introduce users to features they didn't know they needed and to teach them how to use these features.

Of the three features highlighted in the release notes of 1.8.0 and the two features already slated to be highlighted in the release notes of 1.9.0, only one has a corresponding tutorial. We propose to fill this gap for these important new SciPy capabilities.

- The `scipy.sparse` array API and PROPACK method of `scipy.sparse.svd`: We will show how to select the appropriate array type (e.g. compressed sparse row/column, coordinate format, diagonal format) for various tasks and how to perform essential operations and decompositions (e.g. LU, SVD with the new PROPACK).
- Mixed integer linear programming using `scipy.optimize.milp`: We will demonstrate setting up and solving an assignment problem, examples of which are as ubiquitous in everyday life (e.g.

scheduling classes at a school, picking individuals for a team) as they are in science and engineering.

- Resampling methods of `scipy.stats`: the new `scipy.stats.permutation_test` and `scipy.stats.monte_carlo_test` are extremely powerful, with the ability to replicate the results of almost all other hypothesis tests in SciPy, but without the inaccuracies commonly caused by small sample sizes or ties in the data. `scipy.stats.bootstrap` returns standard errors and confidence intervals for virtually any statistic. We will show users how to harness these tools to dramatically expand the capabilities of `scipy.stats`.
- New tools for fitting distributions to data: the new `scipy.stats.fit` fulfills a longstanding need to fit SciPy's discrete distributions to data. We will demonstrate this new capability, and also show how `scipy.stats.fit` gives users more control over fitting continuous distributions than the older `stats.rv_continuous.fit` method.

In addition, to raise awareness about these new features, we will:

- deliver an annual update about the SciPy package as a "Tools Talk",
- present our submission "Optimal Review Assignments for the SciPy Conference Using Binary Integer Linear Programming in SciPy 1.8", and
- host a SciPy development sprint session

at the SciPy 2022 conference.

Consideration of Feedback from 2021 Cycle 3 Proposal

"Clear benefit to the SciPy project and to the ecosystem. The description and justification of overhead is appreciated."

Thank you for the feedback; we will continue to include a careful justification of the budget.

References:

[1] The SciPy Community. "SciPy API Reference". https://docs.scipy.org/doc/scipy/reference/index.html

[2] The SciPy Community. "SciPy User Guide". https://docs.scipy.org/doc/scipy/tutorial/index.html

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

Please explain the benefit of this proposal including:

- *Impact to the project*
- *Impact to the scientific ecosystem*
- *Impact to the community*

This work will impact the SciPy Library project by generating essential documentation for several of its most important new features. Also, through representation at the SciPy 2022 Conference, the SciPy Library project will attract new users and developers.

This work will benefit the scientific Python ecosystem by introducing users to several powerful new features of the SciPy library, enabling them to solve problems that were previously much more difficult or impossible to solve using widely available, permissively-licensed, open-source tools.

This work will strengthen the community by welcoming new users with tutorials (which are much more approachable than API reference documentation), by enabling developers to meet with users and one another at the SciPy conference (despite hundreds of hours working together virtually, Nicholas and Matt have never met in person!), and by encouraging new developers to join the project at the conference sprints.

Amount Requested

$8,005

Brief Budget Justification - How will the money be spent?

$4353 ($2290 for Nicholas McKibben, $2063 for Matt Haberland) is for airfare, lodging, registration fees, and per diem for the authors to attend the SciPy conference.

To ensure high quality, each tutorial will require approximately 8 hours of "author" time and 4 hours of "reviewer" time; each proposer will author two tutorials and review the other two.
$1,200 is for 24 hours of work by Nicholas McKibben at a rate of $50/hr. Nicholas will author the `scipy.sparse` and `scipy.optimize.milp` tutorials and review the two `scipy.stats` tutorials. Note: these funds are not to be paid through Cal Poly; rather, they would be paid directly by NumFOCUS to Nicholas McKibben.

$1,197 ($1,096 salary; $101 fringe benefits/payroll taxes) will compensate overload work by Matt Haberland. Matt will author the two `scipy.stats` tutorials and review the `scipy.sparse` and `scipy.optimize.milp` tutorials.

$1,255 is for Cal Poly recovery of indirect costs for Matt Haberland's portion of the award. Processing Matt's portion of the award through Cal Poly reinforces the connection between this work and Matt's research, which is what allows Matt to work on SciPy for hundreds of hours each year.

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 2021-2022 Fiscal year (July 1 – June 30). Benefits for CPSU Faculty summer and overload work include FICA, SUI, and Workers Compensation are calculated at the DHHS pooled rate of 9.2%. Cal Poly's federally negotiated indirect rate is 38.5% of Modified Total Direct Costs, effective July 1, 2020. 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. The rates in effect at the time the work is performed will be charged to the sponsor.

Timeline of Deliverables

4/15/2022 – decision notification

5/15/2022 – draft `scipy.stats.fit` and `scipy.optimize.milp` tutorials

6/15/2022 – draft `scipy.sparse` and `scipy.stats` resampling methods tutorials

7/11/2022-7/17/2022 – SciPy Conference (presentations, outreach, sprints)

8/15/2022 – tutorial review, edits, and merge

9/15/2022 – submit final report

Has someone been identified to carry out the work in the proposal?

Yes: Matt Haberland and Nicholas McKibben. Nicholas and Matt are ideally suited to the work because they implemented most of the features in question, they are both project maintainers with commit rights, and they have a strong record of contributions to the project (e.g. combined, over 150 PRs merged, many of which implement major new features).

Please list the name and email address of a project leader(s) who has approved this proposal. *

Ralf Gommers <ralf.gommers@gmail.com>

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

I agree.