

# SymPy: Symbolic Computing in Python

Aaron Meurer<sup>1</sup>, Christopher P. Smith<sup>2</sup>, Mateusz Paprocki<sup>3</sup>, Ondřej Čertík<sup>4</sup>, Sergey B. Kirpichev<sup>5</sup>, Matthew Rocklin<sup>6</sup>, AMiT Kumar<sup>7</sup>, Sergiu Ivanov<sup>8</sup>, Jason K. Moore<sup>9</sup>, Sartaj Singh<sup>10</sup>, Thilina Rathnayake<sup>11</sup>, Sean Vig<sup>12</sup>, Brian E. Granger<sup>13</sup>, Richard P. Muller<sup>14</sup>, Francesco Bonazzi<sup>15</sup>, Harsh Gupta<sup>16</sup>, Shivam Vats<sup>17</sup>, Fredrik Johansson<sup>18</sup>, Fabian Pedregosa<sup>19</sup>, Matthew J. Curry<sup>20</sup>, Andy R. Terrel<sup>21</sup>, Štěpán Roučka<sup>22</sup>, Ashutosh Saboo<sup>23</sup>, Isuru Fernando<sup>24</sup>, Sumith Kulal<sup>25</sup>, Robert Cimrman<sup>26</sup>, and Anthony Scopatz<sup>27</sup>

<sup>1</sup>University of South Carolina, Columbia, SC 29201 ([asmeurer@gmail.com](mailto:asmeurer@gmail.com)).

<sup>2</sup>Polar Semiconductor, Inc., Bloomington, MN 55425 ([smichr@gmail.com](mailto:smichr@gmail.com)).

<sup>3</sup>Continuum Analytics, Inc., Austin, TX 78701 ([mattpap@gmail.com](mailto:mattpap@gmail.com)).

<sup>4</sup>Los Alamos National Laboratory, Los Alamos, NM 87545 ([certik@lanl.gov](mailto:certik@lanl.gov)). The Los Alamos National Laboratory is operated by Los Alamos National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy under Contract No. DE-AC52-06NA25396.

<sup>5</sup>Moscow State University, Faculty of Physics, Leninskie Gory, Moscow, 119991, Russia ([skirpichev@gmail.com](mailto:skirpichev@gmail.com)).

<sup>6</sup>Continuum Analytics, Inc., Austin, TX 78701 ([mrocklin@gmail.com](mailto:mrocklin@gmail.com)).

<sup>7</sup>Delhi Technological University, Shahbad Daultpur, Bawana Road, New Delhi 110042, India ([dtu.amit@gmail.com](mailto:dtu.amit@gmail.com)).

<sup>8</sup>Université Paris Est Créteil, 61 av. Général de Gaulle, 94010 Créteil, France ([sergiu.ivanov@u-pec.fr](mailto:sergiu.ivanov@u-pec.fr)).

<sup>9</sup>University of California, Davis, Davis, CA 95616 ([jkm@ucdavis.edu](mailto:jkm@ucdavis.edu)).

<sup>10</sup>Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh 221005, India ([singhsartaj94@gmail.com](mailto:singhsartaj94@gmail.com)).

<sup>11</sup>University of Moratuwa, Bandaranayake Mawatha, Katubedda, Moratuwa 10400, Sri Lanka ([thilinarmtb.10@cse.mrt.ac.lk](mailto:thilinarmtb.10@cse.mrt.ac.lk)).

<sup>12</sup>University of Illinois at Urbana-Champaign, Urbana, IL 61801 ([sean.v.775@gmail.com](mailto:sean.v.775@gmail.com)).

<sup>13</sup>California Polytechnic State University, San Luis Obispo, CA 93407 ([ellisonbg@gmail.com](mailto:ellisonbg@gmail.com)).

<sup>14</sup>Center for Computing Research, Sandia National Laboratories, Albuquerque, NM 87185 ([rmuller@sandia.gov](mailto:rmuller@sandia.gov)). Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

<sup>15</sup>Max Planck Institute of Colloids and Interfaces, Department of Theory and Bio-Systems, Science Park Golm, 14424 Potsdam, Germany ([francesco.bonazzi@mpikg.mpg.de](mailto:francesco.bonazzi@mpikg.mpg.de)).

<sup>16</sup>Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India ([hargup@protonmail.com](mailto:hargup@protonmail.com)).

<sup>17</sup>Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India ([shivamvats.iitkgp@gmail.com](mailto:shivamvats.iitkgp@gmail.com)).

<sup>18</sup>INRIA Bordeaux-Sud-Ouest – LFANT project-team, 200 Avenue de la Vieille Tour, 33405 Talence, France ([fredrik.johansson@gmail.com](mailto:fredrik.johansson@gmail.com)).

<sup>19</sup>INRIA – SIERRA project-team, 2 Rue Simone IFF, 75012 Paris, France ([f@bianp.net](mailto:f@bianp.net)).

<sup>20</sup>Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131 ([mattjcurry@gmail.com](mailto:mattjcurry@gmail.com)).

<sup>21</sup>Fashion Metric, Inc, Austin, TX 78681 ([andy.terrel@gmail.com](mailto:andy.terrel@gmail.com)).

<sup>22</sup>Faculty of Mathematics and Physics, Charles University in Prague, V Holešovičkách 2, 180 00 Praha, Czech Republic ([stepan.roucka@mff.cuni.cz](mailto:stepan.roucka@mff.cuni.cz)).

<sup>23</sup>Birla Institute of Technology and Science, Pilani, K.K. Birla Goa Campus, NH 17B Bypass Road, Zuarinagar, Sancoale, Goa 403726, India ([ashutosh.saboo96@gmail.com](mailto:ashutosh.saboo96@gmail.com)).

<sup>24</sup>University of Moratuwa, Bandaranayake Mawatha, Katubedda, Moratuwa 10400, Sri Lanka ([isuru.11@cse.mrt.ac.lk](mailto:isuru.11@cse.mrt.ac.lk)).  
<sup>25</sup>Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra 400076, India ([sumith@cse.iitb.ac.in](mailto:sumith@cse.iitb.ac.in)).  
<sup>26</sup>New Technologies – Research Centre, University of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic ([cimrman3@ntc.zcu.cz](mailto:cimrman3@ntc.zcu.cz)).  
<sup>27</sup>University of South Carolina, Columbia, SC 29201 ([scopatz@cec.sc.edu](mailto:scopatz@cec.sc.edu)).

## ABSTRACT

SymPy is an open source computer algebra system written in pure Python. It is built with a focus on extensibility and ease of use, through both interactive and programmatic applications. These characteristics have led SymPy to become a popular symbolic library for the scientific Python ecosystem. This paper presents the architecture of SymPy, a description of its features, and a discussion of select submodules. The supplementary material provide additional examples and further outline details of the architecture and features of SymPy.

Keywords: symbolic, Python, computer algebra system

## 1 INTRODUCTION

SymPy is a full featured computer algebra system (CAS) written in the Python [32] programming language. It is free and open source software, licensed under the 3-clause BSD license [48]. The SymPy project was started by Ondřej Čertík in 2005, and it has since grown to over 500 contributors. Currently, SymPy is developed on GitHub using a bazaar community model [43]. The accessibility of the codebase and the open community model allow SymPy to rapidly respond to the needs of users and developers.

Python is a dynamically typed programming language that has a focus on ease of use and readability.<sup>1</sup> Due in part to this focus, it has become a popular language for scientific computing and data science, with a broad ecosystem of libraries [37]. SymPy is itself used by many libraries and tools to support research within a variety of domains, such as SageMath [57] (pure and applied mathematics), yt [63] (astronomy and astrophysics), PyDy [19] (multibody dynamics), and SfePy [10] (finite elements).

Unlike many CAS's, SymPy does not invent its own programming language. Python itself is used both for the internal implementation and end user interaction. By using the operator overloading functionality of Python, SymPy follows the embedded domain specific language paradigm proposed by Hudak [24]. The exclusive usage of a single programming language makes it easier for people already familiar with that language to use or develop SymPy. Simultaneously, it enables developers to focus on mathematics, rather than language design. SymPy officially supports Python 2.6, 2.7 and 3.2–3.5.

SymPy is designed with a strong focus on usability as a library. Extensibility is important in its application program interface (API) design. Thus, SymPy makes no attempt to extend the Python language itself. The goal is for users of SymPy to be able to include SymPy alongside other Python libraries in their workflow, whether that be in an interactive environment or as a programmatic part in a larger system.

Being a library, SymPy does not have a built-in graphical user interface (GUI). However, SymPy exposes a rich interactive display system, and supports registering display formatters with Jupyter [29] frontends, including the Notebook and Qt Console, which will render SymPy expressions using MathJax [9] or L<sup>A</sup>T<sub>E</sub>X.

The remainder of this paper discusses key components of the SymPy library. Section 2 enumerates the features of SymPy and takes a closer look at some of the important ones. The section 3 looks at the numerical features of SymPy and its dependency library, mpmath.

---

<sup>1</sup>This paper assumes a moderate familiarity with the Python programming language.

Section 4 looks at the domain specific physics submodules for performing symbolic and numerical calculations in classical mechanics and quantum mechanics. Section 5 discusses the architecture of SymPy. Section 6 looks at a selection of packages that depend on SymPy. Conclusions and future directions for SymPy are given in section 7. All examples in this paper use SymPy version 1.0 and mpmath version 0.19.

The following statement imports all SymPy functions into the global Python namespace.<sup>2</sup> From here on, all examples in this paper assume that this statement has been executed.<sup>3</sup>

```
>>> from sympy import *
```

All the examples in this paper can be tested on [SymPy Live](http://sympy.live), an online Python shell that uses the Google App Engine [11] to execute SymPy code. SymPy Live is also integrated into the SymPy documentation at <http://docs.sympy.org>.

## 2 OVERVIEW OF CAPABILITIES

This section gives a basic introduction of SymPy, and lists its features. A few features—assumptions, simplification, calculus, polynomials, printers, solvers, and matrices—are core components of SymPy and are discussed in depth. Many other features are discussed in depth in the supplementary material.

### 2.1 Basic Usage

Symbolic variables, called symbols, must be defined and assigned to Python variables before they can be used. This is typically done through the `symbols` function, which may create multiple symbols in a single function call. For instance,

```
>>> x, y, z = symbols('x y z')
```

creates three symbols representing variables named  $x$ ,  $y$ , and  $z$ . In this particular instance, these symbols are all assigned to Python variables of the same name. However, the user is free to assign them to different Python variables, while representing the same symbol, such as `a`, `b`, `c = symbols('x y z')`. In order to minimize potential confusion, though, all examples in this paper will assume that the symbols  $x$ ,  $y$ , and  $z$  have been assigned to Python variables identical to their symbolic names.

Expressions are created from symbols using Python's mathematical syntax. For instance, the following Python code creates the expression  $(x^2 - 2x + 3)/y$ . Note that the expression remains unevaluated: it is represented symbolically.

```
>>> (x**2 - 2*x + 3)/y
(x**2 - 2*x + 3)/y
```

### 2.2 List of Features

Although SymPy's extensive feature set cannot be covered in depth in this paper, bedrock areas, that is, those areas that are used throughout the library, are discussed in their own subsections below. Additionally, Table 1 gives a compact listing of all major capabilities present in the SymPy codebase. This grants a sampling from the breadth of topics and application domains that SymPy services. Unless stated otherwise, all features noted in Table 1 are symbolic in nature. Numeric features are discussed in Section 3.

**Table 1.** SymPy Features and Descriptions

---

<sup>2</sup>`import *` has been used here to aid the readability of the paper, but is best to avoid such wildcard import statements in production code, as they make it unclear which names are present in the namespace. Furthermore, imported names could clash with already existing imports from another package. For example, SymPy, the standard Python `math` library, and NumPy all define the `exp` function, but only the SymPy one will work with SymPy symbolic expressions.

<sup>3</sup> The three greater-than signs denote the user input for the Python interactive session, with the result, if there is one, shown on the next line.

| Feature (submodules)   | Description   |
|--|---|
| Calculus ( <code>sympy.core</code> ,<br><code>sympy.calculus</code> ,<br><code>sympy.integrals</code> ,<br><code>sympy.series</code> ) | Algorithms for computing derivatives, integrals, and limits.  |
| Category Theory<br>( <code>sympy.categories</code> )   | Representation of objects, morphisms, and diagrams. Tools for drawing diagrams with Xy-pic [47].  |
| Code Generation<br>( <code>sympy.printing</code> ,<br><code>sympy.codegen</code> )   | Generation of compilable and executable code in a variety of different programming languages from expressions directly. Target languages include C, Fortran, Julia, JavaScript, Mathematica, MATLAB and Octave, Python, and Theano.   |
| Combinatorics & Group Theory<br>( <code>sympy.combinatorics</code> )   | Permutations, combinations, partitions, subsets, various permutation groups (such as polyhedral, Rubik, symmetric, and others), Gray codes [36], and Prufer sequences [4].  |
| Concrete Math<br>( <code>sympy.concrete</code> )   | Summation, products, tools for determining whether summation and product expressions are convergent, absolutely convergent, hypergeometric, and for determining other properties; computation of Gosper's normal form [42] for two univariate polynomials.                          |
| Cryptography<br>( <code>sympy.crypto</code> )  | Block and stream ciphers, including shift, Affine, substitution, Vigenère's, Hill's, bifid, RSA, Kid RSA, linear-feedback shift registers, and Elgamal encryption.  |
| Differential Geometry<br>( <code>sympy.diffgeom</code> )   | Representations of manifolds, metrics, tensor products, and coordinate systems in Riemannian and pseudo-Riemannian geometries [51].   |
| Geometry ( <code>sympy.geometry</code> )   | Representations of 2D geometrical entities, such as lines and circles. Enables queries on these entities, such as asking the area of an ellipse, checking for collinearity of a set of points, or finding the intersection between objects.   |
| Lie Algebras<br>( <code>sympy.liealgebras</code> )   | Representations of Lie algebras and root systems.   |
| Logic ( <code>sympy.logic</code> )   | Boolean expressions, equivalence testing, satisfiability, and normal forms.   |
| Matrices ( <code>sympy.matrices</code> )   | Tools for creating matrices of symbols and expressions. Both sparse and dense representations, as well as symbolic linear algebraic operations (e.g., inversion and factorization), are supported.  |
| Matrix Expressions<br>( <code>sympy.matrices.expressions</code> )  | Matrices with symbolic dimensions (unspecified entries). Block matrices.  |
| Number Theory<br>( <code>sympy.ntheory</code> )  | Prime number generation, primality testing, integer factorization, continued fractions, Egyptian fractions, modular arithmetic, quadratic residues, partitions, binomial and multinomial coefficients, prime number tools, hexadecimal digits of $\pi$ , and integer factorization. |
| Plotting ( <code>sympy.plotting</code> )   | Hooks for visualizing expressions via matplotlib [25] or as text drawings when lacking a graphical back-end. 2D function plotting, 3D function plotting, and 2D implicit function plotting are supported.   |
| Polynomials ( <code>sympy.polys</code> )   | Polynomial algebras over various coefficient domains. Functionality ranges from simple operations (e.g., polynomial division) to advanced computations (e.g., Gröbner bases [1] and multivariate factorization over algebraic number domains).                                      |
| Printing ( <code>sympy.printing</code> )   | Functions for printing SymPy expressions in the terminal with ASCII or Unicode characters and converting SymPy expressions to $\text{\LaTeX}$ and MathML.   |

|  |   |
|--|---|
| Quantum Mechanics<br>( <code>sympy.physics.quantum</code> )                    | Quantum states, bra-ket notation, operators, basis sets, representations, tensor products, inner products, outer products, commutators, anticommutators, and specific quantum system implementations.   |
| Series ( <code>sympy.series</code> )   | Series expansion, sequences, and limits of sequences. This includes Taylor, Laurent, and Puiseux series as well as special series, such as Fourier and formal power series.   |
| Sets ( <code>sympy.sets</code> )   | Representations of empty, finite, and infinite sets (including special sets such as the natural, integer, and complex numbers). Operations on sets such as union, intersection, Cartesian product, and building sets from other sets are supported.   |
| Simplification<br>( <code>sympy.simplify</code> )                              | Functions for manipulating and simplifying expressions. Includes algorithms for simplifying hypergeometric functions, trigonometric expressions, rational functions, combinatorial functions, square root denesting, and common subexpression elimination.  |
| Solvers ( <code>sympy.solvers</code> )   | Functions for symbolically solving equations, systems of equations, both linear and non-linear, inequalities, ordinary differential equations, partial differential equations, Diophantine equations, and recurrence relations.   |
| Special Functions<br>( <code>sympy.functions</code> )                          | Implementations of a number of well known special functions, including Dirac delta, Gamma, Beta, Gauss error functions, Fresnel integrals, Exponential integrals, Logarithmic integrals, Trigonometric integrals, Bessel, Hankel, Airy, B-spline, Riemann Zeta, Dirichlet eta, polylogarithm, Lerch transcendent, hypergeometric, elliptic integrals, Mathieu, Jacobi polynomials, Gegenbauer polynomial, Chebyshev polynomial, Legendre polynomial, Hermite polynomial, Laguerre polynomial, and spherical harmonic functions. |
| Statistics ( <code>sympy.stats</code> )  | Support for a random variable type as well as the ability to declare this variable from prebuilt distribution functions such as Normal, Exponential, Coin, Die, and other custom distributions [46].  |
| Tensors ( <code>sympy.tensor</code> )<br>Vectors ( <code>sympy.vector</code> ) | Symbolic manipulation of indexed objects.<br>Basic operations on vectors and differential calculus with respect to 3D Cartesian coordinate systems.   |

## 2.3 Assumptions

The assumptions system allows users to specify that symbols have certain common mathematical properties, such as being positive, imaginary, or integer. SymPy is careful to never perform simplifications on an expression unless the assumptions allow them. For instance, the identity  $\sqrt{t^2} = t$  holds if  $t$  is nonnegative ( $t \geq 0$ ). However, for general complex  $t$ , no such identity holds.

By default, SymPy performs all calculations assuming that symbols are complex valued. This assumption makes it easier to treat mathematical problems in full generality.

```
>>> t = Symbol('t')
>>> sqrt(t**2)
sqrt(t**2)
```

By assuming the most general case, that  $t$  is complex by default, SymPy avoids performing mathematically invalid operations. However, in many cases users will wish to simplify expressions containing terms like  $\sqrt{t^2}$ .

Assumptions are set on `Symbol` objects when they are created. For instance `Symbol('t', positive=True)` will create a symbol named  $t$  that is assumed to be positive.

```

154 >>> t = Symbol('t', positive=True)
155 >>> sqrt(t**2)
156 t

```

Some of the common assumptions that SymPy allows are **positive**, **negative**, **real**, **nonpositive**, **integer**, **prime** and **commutative**.<sup>4</sup> Assumptions on any object can be checked with the `is_assumption` attributes, like `t.is_positive`.

Assumptions are only needed to restrict a domain so that certain simplifications can be performed. They are not required to make the domain match the input of a function. For instance, one can create the object  $\sum_{n=0}^m f(n)$  as `Sum(f(n), (n, 0, m))` without setting `integer=True` when creating the Symbol object `n`.

The assumptions system additionally has deductive capabilities. The assumptions use a three-valued logic using the Python built in objects `True`, `False`, and `None`. Note that `False` is returned if the SymPy object doesn't or can't have the assumption. For example, both `I.is_real` and `I.is_prime` return `False` for the imaginary unit `I`.

`None` represents the “unknown” case. This could mean that given assumptions do not unambiguously specify the truth of an attribute. For instance, `Symbol('x', real=True).is_positive` will give `None` because a real symbol might be positive or negative. `None` could also mean that not enough is known or implemented to compute the given fact. For instance, `(pi + E).is_irrational` gives `None`—indeed, the rationality of  $\pi + e$  is an open problem in mathematics [31].

Basic implications between the facts are used to deduce assumptions. Deductions are made using the Rete algorithm [13].<sup>5</sup> For instance, the assumptions system knows that being an integer implies being rational.

```

176 >>> i = Symbol('i', integer=True)
177 >>> i.is_rational
178 True

```

Furthermore, expressions compute the assumptions on themselves based on the assumptions of their arguments. For instance, if `x` and `y` are both created with `positive=True`, then `(x + y).is_positive` will be `True` (whereas `(x - y).is_positive` will be `None`).

## 2.4 Simplification

The generic way to simplify an expression is by calling the `simplify` function. It must be emphasized that simplification is not a rigorously defined mathematical operation [34]. The `simplify` function applies several simplification routines along with heuristics to make the output expression “simple”.<sup>6</sup>

It is often preferable to apply more directed simplification functions. These apply very specific rules to the input expression and are typically able to make guarantees about the output. For instance, the `factor` function, given a polynomial with rational coefficients in several variables, is guaranteed to produce a factorization into irreducible factors. Table 2 lists common simplification functions.

Examples for these simplification functions can be found in the supplement.

<sup>4</sup>SymPy assumes that two expressions  $A$  and  $B$  commute with each other multiplicatively, that is,  $A \cdot B = B \cdot A$ , unless they both have `commutative=False`. Many algorithms in SymPy require special consideration to work correctly with noncommutative products.

<sup>5</sup>For historical reasons, this algorithm is distinct from the `sympy.logic` submodule, which is discussed in the supplementary material. SymPy also has an experimental assumptions system which stores facts separate from objects, and uses `sympy.logic` and a SAT solver for deduction. We will not discuss this system here.

<sup>6</sup>The `measure` parameter of the `simplify` function lets the user specify the Python function used to determine how complex an expression is. The default measure function returns the total number of operations in the expression.



**Table 2.** Some SymPy Simplification Functions

|                          |   |
|--------------------------|---|
| <code>expand</code>      | expand the expression   |
| <code>factor</code>      | factor a polynomial into irreducibles                             |
| <code>collect</code>     | collect polynomial coefficients                                   |
| <code>cancel</code>      | rewrite a rational function as $p/q$ with common factors canceled |
| <code>apart</code>       | compute the partial fraction decomposition of a rational function |
| <code>trigsimp</code>    | simplify trigonometric expressions [18]                           |
| <code>hyperexpand</code> | expand hypergeometric functions [44, 45]                          |

## 2.5 Calculus

SymPy provides all the basic operations of calculus, such as calculating limits, derivatives, integrals, or summations.

Limits are computed with the `limit` function, using the Gruntz algorithm [22] for computing symbolic limits and heuristics (a description of the Gruntz algorithm may be found in the supplement). For example, the following computes  $\lim_{x \rightarrow \infty} x \sin(\frac{1}{x}) = 1$ . Note that SymPy denotes  $\infty$  as `oo`.

```
>>> limit(x*sin(1/x), x, oo)
1
```

As a more complex example, SymPy computes

$$\lim_{x \rightarrow 0} \left( 2e^{\frac{1 - \cos(x)}{\sin(x)}} - 1 \right)^{\frac{\sinh(x)}{\operatorname{atan}^2(x)}} = e.$$

```
>>> limit((2*exp((1-cos(x))/sin(x))-1)**(sinh(x)/atan(x)**2), x, 0)
E
```

Derivatives are computed with the `diff` function, which recursively uses the various differentiation rules.

```
>>> diff(sin(x)*exp(x), x)
exp(x)*sin(x) + exp(x)*cos(x)
```

Integrals are calculated with the `integrate` function. SymPy implements a combination of the Risch algorithm [7], table lookups, a reimplement of Manuel Bronstein’s “Poor Man’s Integrator” [6], and an algorithm for computing integrals based on Meijer G-functions [44, 45]. These allow SymPy to compute a wide variety of indefinite and definite integrals. The Meijer G-function algorithm and the Risch algorithm are respectively demonstrated below by the computation of

$$\int_0^\infty e^{-st} \log(t) dt = -\frac{\log(s) + \gamma}{s}$$

and

$$\int \frac{-2x^2(\log(x) + 1)e^{x^2} + (e^{x^2} + 1)^2}{x(e^{x^2} + 1)^2(\log(x) + 1)} dx = \log(\log(x) + 1) + \frac{1}{e^{x^2} + 1}.$$

```
>>> s, t = symbols('s t', positive=True)
>>> integrate(exp(-s*t)*log(t), (t, 0, oo)).simplify()
-(log(s) + EulerGamma)/s
>>> integrate((-2*x**2*(log(x) + 1)*exp(x**2) +
... (exp(x**2) + 1)**2)/(x*(exp(x**2) + 1)**2*(log(x) + 1)), x)
log(log(x) + 1) + 1/(exp(x**2) + 1)
```

Summations are computed with the `summation` function, which uses a combination of Gosper's algorithm [21], an algorithm that uses Meijer G-functions [44, 45], and heuristics. Products are computed with `product` function via a suite of heuristics.

```
>>> i, n = symbols('i n')
>>> summation(2**i, (i, 0, n - 1))
2**n - 1
>>> summation(i*factorial(i), (i, 1, n))
n*factorial(n) + factorial(n) - 1
```

Series expansions are computed with the `series` function. This example computes the power series of  $\sin(x)$  around  $x = 0$  up to  $x^6$ .

```
>>> series(sin(x), x, 0, 6)
x - x**3/6 + x**5/120 + O(x**6)
```

The supplementary material discusses series expansions methods in more depth.

Integrals, derivatives, summations, products, and limits that cannot be computed return unevaluated objects. These can also be created directly if the user chooses.

```
>>> integrate(x**x, x)
Integral(x**x, x)
>>> Sum(2**i, (i, 0, n - 1))
Sum(2**i, (i, 0, n - 1))
```

## 2.6 Polynomials

SymPy implements a suite of algorithms for polynomial manipulation, which ranges from relatively simple algorithms for doing arithmetic of polynomials, to advanced methods for factoring multivariate polynomials into irreducibles, symbolically determining real and complex root isolation intervals, or computing Gröbner bases.

Polynomial manipulation is useful in its own right. Within SymPy, though, it is mostly used indirectly as a tool in other areas of the library. In fact, many mathematical problems in symbolic computing are first expressed using entities from the symbolic core, preprocessed, and then transformed into a problem in the polynomial algebra, where generic and efficient algorithms are used to solve the problem. The solutions to the original problem are subsequently recovered from the results. This is a common scheme in symbolic integration or summation algorithms.

SymPy implements dense and sparse polynomial representations.<sup>7</sup> Both are used in the univariate and multivariate cases. The dense representation is the default for univariate polynomials. For multivariate polynomials, the choice of representation is based on the application. The most common case for the sparse representation is algorithms for computing Gröbner bases (Buchberger, F4, and F5) [8, 14, 15]. This is because different monomial orderings can be expressed easily in this representation. However, algorithms for computing multivariate GCDs or factorizations, at least those currently implemented in SymPy [38], are better expressed when the representation is dense. The dense multivariate representation is specifically a recursively-dense representation, where polynomials in  $K[x_0, x_1, \dots, x_n]$  are viewed as a polynomials in  $K[x_0][x_1] \dots [x_n]$ . Note that despite this, the coefficient domain  $K$ , can be a multivariate polynomial domain as well. The dense recursive representation in Python gets inefficient as the number of variables increases.

Some examples for the `sympy.polys` submodule can be found in the supplement.

## 2.7 Printers

SymPy has a rich collection of expression printers. By default, an interactive Python session will render the `str` form of an expression, which has been used in all the examples in this paper so

---

<sup>7</sup>In a dense representation, the coefficients for all terms up to the degree of each variable are stored in memory. In a sparse representation, only the nonzero coefficients are stored.



261 far. The `str` form of an expression is valid Python and roughly matches what a user would type  
 262 to enter the expression.<sup>8</sup>

```
263 >>> phi0 = Symbol('phi0')
264 >>> str(Integral(sqrt(phi0), phi0))
265 'Integral(sqrt(phi0), phi0)'
```

266 A two-dimensional (2D) textual representation of the expression can be printed with  
 267 monospace fonts via `pprint`. Unicode characters are used for rendering mathematical sym-  
 268 bols such as integral signs, square roots, and parentheses. Greek letters and subscripts in symbol  
 269 names that have Unicode code points associated are also rendered automatically.

```
>>> pprint(Integral(sqrt(phi0 + 1), phi0))

$$\int \sqrt{\varphi_0 + 1} \, d(\varphi_0)$$

```

270  
 271 Alternately, the `use_unicode=False` flag can be set, which causes the expression to be printed  
 272 using only ASCII characters.

```
273 >>> pprint(Integral(sqrt(phi0 + 1), phi0), use_unicode=False)
274 /
275 |
276 | _____
277 | \ / phi0 + 1 d(phi0)
278 |
279 /
```

280 The function `latex` returns a  $\text{\LaTeX}$  representation of an expression.

```
281 >>> print(latex(Integral(sqrt(phi0 + 1), phi0)))
282 \int \sqrt{\phi_{0} + 1}\, d\phi_{0}
```

283 Users are encouraged to run the `init_printing` function at the beginning of interactive  
 284 sessions, which automatically enables the best pretty printing supported by their environment.  
 285 In the Jupyter Notebook or Qt Console [40], the  $\text{\LaTeX}$  printer is used to render expressions  
 286 using MathJax or  $\text{\LaTeX}$ , if it is installed on the system. The 2D text representation is used  
 287 otherwise.

288 Other printers such as MathML are also available. SymPy uses an extensible printer subsystem,  
 289 which allows extending any given printer, and also allows custom objects to define their printing  
 290 behavior for any printer. The code generation functionality of SymPy relies on this subsystem  
 291 to convert expressions into code in various target programming languages.

## 292 2.8 Solvers

293 SymPy has equation solvers that can handle ordinary differential equations, recurrence relation-  
 294 ships, Diophantine equations<sup>9</sup>, and algebraic equations. There is also rudimentary support for  
 295 simple partial differential equations.

296 There are two functions for solving algebraic equations in SymPy: `solve` and `solveset`.  
 297 `solveset` has several design changes with respect to the older `solve` function. This distinction  
 298 is present in order to resolve the usability issues with the previous `solve` function API while  
 299 maintaining backward compatibility with earlier versions of SymPy. `solveset` only requires  
 300 essential input information from the user. The function signatures of `solve` and `solveset` are

<sup>8</sup>Many Python libraries distinguish the `str` form of an object, which is meant to be human-readable, and the `repr` form, which is meant to be valid Python that recreates the object. In SymPy, `str(expr) == repr(expr)`. In other words, the string representation of an expression is designed to be compact, human-readable, and valid Python code that could be used to recreate the expression. As noted in section 5.1, the `srepr` function prints the exact, verbose form of an expression.

<sup>9</sup>See the supplementary material for an in depth discussion on the Diophantine submodule.

```

301 solve(f, *symbols, **flags)
302 solveset(f, symbol, domain=S.Complexes)

```

303 The `domain` parameter can be any set from the `sympy.sets` module (see the supplementary  
304 material for details on `sympy.sets`), but is typically either `S.Complexes` (the default) or `S.Reals`;  
305 the latter causes `solveset` to only return real solutions.

306 An important difference between the two functions is that the output API of `solve` varies  
307 with input (sometimes returning a Python list and sometimes a Python dictionary) whereas  
308 `solveset` always returns a SymPy set object.

309 Both functions implicitly assume that expressions are equal to 0. For instance, `solveset(x -`  
310 `1, x)` solves  $x - 1 = 0$  for  $x$ .

311 `solveset` is under active development as a planned replacement for `solve`. There are certain  
312 features which are implemented in `solve` that are not yet implemented in `solveset`, including  
313 multivariate systems, and some transcendental equations.

314 Some examples for `solveset` and `solve` can be found in the supplement.

## 315 2.9 Matrices

316 Besides being an important feature in its own right, computations on matrices with symbolic  
317 entries are important for many algorithms within SymPy. The following code shows some basic  
318 usage of the `Matrix` class.

```

319 >>> A = Matrix([[x, x + y], [y, x]])
320 >>> A
321 Matrix([
322 [x, x + y],
323 [y,      x]])

```

324 SymPy matrices support common symbolic linear algebra manipulations, including matrix  
325 addition, multiplication, exponentiation, computing determinants, solving linear systems, and  
326 computing inverses using LU decomposition, LDL decomposition, Gauss-Jordan elimination,  
327 Cholesky decomposition, Moore-Penrose pseudoinverse, singular values, and adjugate matrix.

328 All operations are performed symbolically. For instance, eigenvalues are computed by  
329 generating the characteristic polynomial using the Berkowitz algorithm and then solving it using  
330 polynomial routines.

```

331 >>> A.eigenvals()
332 {x - sqrt(y*(x + y)): 1, x + sqrt(y*(x + y)): 1}

```

333 Internally these matrices store the elements as Lists of Lists (LIL) [27], meaning the matrix  
334 is stored as a list of lists of entries (effectively, the input format used to create the matrix `A`  
335 above), making it a dense representation.<sup>10</sup> For storing sparse matrices, the `SparseMatrix` class  
336 can be used. Sparse matrices store their elements in Dictionary of Keys (DOK) format, meaning  
337 entries are stored as (`row`, `column`) pairs mapping to the elements.

338 SymPy also supports matrices with symbolic dimension values. `MatrixSymbol` represents  
339 a matrix with dimensions  $m \times n$ , where  $m$  and  $n$  can be symbolic. Matrix addition and  
340 multiplication, scalar operations, matrix inverse, and transpose are stored symbolically as matrix  
341 expressions.

342 Block matrices are also implemented in SymPy. `BlockMatrix` elements can be any matrix ex-  
343 pression, including explicit matrices, matrix symbols, and other block matrices. All functionalities  
344 of matrix expressions are also present in `BlockMatrix`.

345 When symbolic matrices are combined with the assumptions submodule for logical inference,  
346 they provide powerful reasoning over invertibility, semi-definiteness, orthogonality, etc., which  
347 are valuable in the construction of numerical linear algebra systems.

348 More examples for `Matrix` and `BlockMatrix` may be found in the supplement.

<sup>10</sup>Similar to the polynomials submodule, dense here means that all entries are stored in memory, contrasted with a sparse representation where only nonzero entries are stored.

### 349 3 NUMERICS

350 While SymPy primarily focuses on symbolics, it is impossible to have a complete symbolic system  
351 without the ability to numerically evaluate expressions. Many operations directly use numerical  
352 evaluation, such as plotting a function, or solving an equation numerically. Beyond this, certain  
353 purely symbolic operations require numerical evaluation to effectively compute. For instance,  
354 determining the truth value of  $e + 1 > \pi$  is most conveniently done by numerically evaluating  
355 both sides of the inequality and checking which is larger.

#### 356 3.1 Floating-Point Numbers

357 Floating-point numbers in SymPy are implemented by the `Float` class, which represents an  
358 arbitrary-precision binary floating-point number by storing its value and precision (in bits).  
359 This representation is distinct from the Python built-in `float` type, which is a wrapper around  
360 machine `double` types and uses a fixed precision (53-bit).

361 Because Python `float` literals are limited in precision, strings should be used to input precise  
362 decimal values:

```
363 >>> Float(1.1)
364 1.1000000000000000
365 >>> Float(1.1, 30) # precision equivalent to 30 digits
366 1.100000000000000008881784197001
367 >>> Float("1.1", 30)
368 1.100000000000000000000000000000
```

369 The `evalf` method converts a constant symbolic expression to a `Float` with the specified precision,  
370 here 25 digits:

```
371 >>> (pi + 1).evalf(25)
372 4.141592653589793238462643
```

373 `Float` numbers do not track their accuracy, and should be used with caution within symbolic  
374 expressions since familiar dangers of floating-point arithmetic apply [20]. A notorious case is  
375 that of catastrophic cancellation:

```
376 >>> cos(exp(-100)).evalf(25) - 1
377 0
```

378 Applying the `evalf` method to the whole expression solves this problem. Internally, `evalf`  
379 estimates the number of accurate bits of the floating-point approximation for each sub-expression,  
380 and adaptively increases the working precision until the estimated accuracy of the final result  
381 matches the sought number of decimal digits:

```
382 >>> (cos(exp(-100)) - 1).evalf(25)
383 -6.919482633683687653243407e-88
```

384 The `evalf` method works with complex numbers and supports more complicated expressions, such  
385 as special functions, infinite series, and integrals. The internal error tracking does not provide  
386 rigorous error bounds (in the sense of interval arithmetic) and cannot be used to accurately track  
387 uncertainty in measurement data; the sole purpose is to mitigate loss of accuracy that typically  
388 occurs when converting symbolic expressions to numerical values.

#### 389 3.2 The mpmath Library

390 The implementation of arbitrary-precision floating-point arithmetic is supplied by the mpmath  
391 library [26]. Originally, it was developed as a SymPy submodule but has subsequently been  
392 moved to a standalone pure-Python package. The basic datatypes in mpmath are `mpf` and  
393 `mpc`, which respectively act as multiprecision substitutes for Python's `float` and `complex`. The  
394 floating-point precision is controlled by a global context:



## 4.1 Classical Mechanics

One of the core domains that SymPy supports is the physics of classical mechanics. This is in turn separated into two distinct components: vector algebra and mechanics.

### 4.1.1 Vector Algebra

The `sympy.physics.vector` submodule provides reference frame-, time-, and space-aware vector and dyadic objects that allow for three-dimensional operations such as addition, subtraction, scalar multiplication, inner and outer products, and cross products. The vector and dyadic objects can both be written in very compact notation that make it easy to express the vectors and dyadics in terms of multiple reference frames with arbitrarily defined relative orientations. The vectors are used to specify the positions, velocities, and accelerations of points; orientations, angular velocities, and angular accelerations of reference frames; and forces and torques. The dyadics are essentially reference frame-aware  $3 \times 3$  tensors [52]. The vector and dyadic objects can be used for any one-, two-, or three-dimensional vector algebra, and they provide a strong framework for building physics and engineering tools.

The following Python code demonstrates how a vector is created using the orthogonal unit vectors of three reference frames that are oriented with respect to each other, and the result of expressing the vector in the  $A$  frame. The  $B$  frame is oriented with respect to the  $A$  frame using Z-X-Z Euler Angles of magnitude  $\pi$ ,  $\frac{\pi}{2}$ , and  $\frac{\pi}{3}$ , respectively, whereas the  $C$  frame is oriented with respect to the  $B$  frame through a simple rotation about the  $B$  frame's  $X$  unit vector through  $\frac{\pi}{2}$ .

```
>>> from sympy.physics.vector import ReferenceFrame
>>> A, B, C = symbols('A B C', cls=ReferenceFrame)
>>> B.orient(A, 'body', (pi, pi/3, pi/4), 'zxz')
>>> C.orient(B, 'axis', (pi/2, B.x))
>>> v = 1*A.x + 2*B.z + 3*C.y
>>> v
A.x + 2*B.z + 3*C.y
>>> v.express(A)
A.x + 5*sqrt(3)/2*A.y + 5/2*A.z
```

### 4.1.2 Mechanics

The `sympy.physics.mechanics` submodule utilizes the `sympy.physics.vector` submodule to populate time-aware particle and rigid-body objects to fully describe the kinematics and kinetics of a rigid multi-body system. These objects store all of the information needed to derive the ordinary differential or differential algebraic equations that govern the motion of the system, i.e., the equations of motion. These equations of motion abide by Newton's laws of motion and can handle arbitrary kinematic constraints or complex loads. The submodule offers two automated methods for formulating the equations of motion based on Lagrangian Dynamics [30] and Kane's Method [28]. Lastly, there are automated linearization routines for constrained dynamical systems [41].

## 4.2 Quantum Mechanics

The `sympy.physics.quantum` submodule has extensive capabilities to solve problems in quantum mechanics, using Python objects to represent the different mathematical objects relevant in quantum theory [49]: states (bras and kets), operators (unitary, Hermitian, etc.), and basis sets, as well as operations on these objects such as representations, tensor products, inner products, outer products, commutators, and anticommutators. The base objects are designed in the most general way possible to enable any particular quantum system to be implemented by subclassing the base operators and defining the relevant class methods to provide system-specific logic.

Symbolic quantum operators and states may be defined, and one can perform a full range of operations with them.

```
>>> from sympy.physics.quantum import Commutator, Dagger, Operator
>>> from sympy.physics.quantum import Ket, qapply
>>> A, B, C, D = symbols('A B C D', cls=Operator)
>>> a = Ket('a')
```

```

498 >>> comm = Commutator(A, B)
499 >>> comm
500 [A,B]
501 >>> qapply(Dagger(comm*a)).doit()
502 -<a|*(Dagger(A)*Dagger(B) - Dagger(B)*Dagger(A))

```

503 Commutators can be expanded using common commutator identities:

```

504 >>> Commutator(C+B, A*D).expand(commutator=True)
505 -[A,B]*D - [A,C]*D + A*[B,D] + A*[C,D]

```

506 On top of this set of base objects, a number of specific quantum systems have been implemented  
507 in a fully symbolic framework. These include:

- 508 • Many of the exactly solvable quantum systems, including simple harmonic oscillator states  
509 and raising/lowering operators, infinite square well states, and 3D position and momentum  
510 operators and states.
- 511 • Second quantized formalism of non-relativistic many-body quantum mechanics [16].
- 512 • Quantum angular momentum [64]. Spin operators and their eigenstates can be represented  
513 in any basis and for any quantum numbers. A rotation operator representing the Wigner-D  
514 matrix, which may be defined symbolically or numerically, is also implemented to rotate  
515 spin eigenstates. Functionality for coupling and uncoupling of arbitrary spin eigenstates is  
516 provided, including symbolic representations of Clebsch-Gordon coefficients and Wigner  
517 symbols.
- 518 • Quantum information and computing [35]. Multidimensional qubit states, and a full  
519 set of one- and two-qubit gates are provided and can be represented symbolically or as  
520 matrices/vectors. With these building blocks, it is possible to implement a number of basic  
521 quantum algorithms including the quantum Fourier transform, quantum error correction,  
522 quantum teleportation, Grover's algorithm, dense coding, etc. In addition, any quantum  
523 circuit may be plotted using the `circuit_plot` function (Figure 1).

524 Here are a few short examples of the quantum information and computing capabilities in  
525 `sympy.physics.quantum`. Start with a simple four-qubit state and flip the second qubit from the  
526 right using a Pauli-X gate:

```

527 >>> from sympy.physics.quantum.qubit import Qubit
528 >>> from sympy.physics.quantum.gate import XGate
529 >>> q = Qubit('0101')
530 >>> q
531 |0101>
532 >>> X = XGate(1)
533 >>> qapply(X*q)
534 |0111>

```

535 Qubit states can also be used in adjoint operations, tensor products, inner/outer products:

```

536 >>> Dagger(q)
537 <0101|
538 >>> ip = Dagger(q)*q
539 >>> ip
540 <0101|0101>
541 >>> ip.doit()
542 1

```

543 Quantum gates (unitary operators) can be applied to transform these states and then classical  
544 measurements can be performed on the results:



```

545 >>> from sympy.physics.quantum.qubit import measure_all
546 >>> from sympy.physics.quantum.gate import H, X, Y, Z
547 >>> c = H(0)*H(1)*Qubit('00')
548 >>> c
549 H(0)*H(1)*|00>
550 >>> q = qapply(c)
551 >>> measure_all(q)
552 [(|00>, 1/4), (|01>, 1/4), (|10>, 1/4), (|11>, 1/4)]

```



**Figure 1.** The circuit diagram for a three-qubit quantum Fourier transform generated by SymPy.

553 Lastly, the following example demonstrates creating a three-qubit quantum Fourier transform,  
 554 decomposing it into one- and two-qubit gates, and then generating a circuit plot for the sequence  
 555 of gates (see Figure 1).

```

556 >>> from sympy.physics.quantum.qft import QFT
557 >>> from sympy.physics.quantum.circuitplot import circuit_plot
558 >>> fourier = QFT(0,3).decompose()
559 >>> fourier
560 SWAP(0,2)*H(0)*C((0),S(1))*H(1)*C((0),T(2))*C((1),S(2))*H(2)
561 >>> c = circuit_plot(fourier, nqubits=3)

```

## 562 5 ARCHITECTURE

563 Software architecture is of central importance in any large software project because it establishes  
 564 predictable patterns of usage and development [50]. This section describes the essential structural  
 565 components of SymPy, provides justifications for the design decisions that have been made, and  
 566 gives example user-facing code as appropriate.

### 567 5.1 The Core

568 A computer algebra system stores mathematical expressions as data structures. For example,  
 569 the mathematical expression  $x + y$  is represented as a tree with three nodes,  $+$ ,  $x$ , and  $y$ ,  
 570 where  $x$  and  $y$  are ordered children of  $+$ . As users manipulate mathematical expressions  
 571 with traditional mathematical syntax, the CAS manipulates the underlying data structures.  
 572 Automated optimizations and computations such as integration, simplification, etc. are all  
 573 functions that consume and produce expression trees.

574 In SymPy every symbolic expression is an instance of a Python `Basic` class,<sup>11</sup> a superclass  
 575 of all SymPy types providing common methods to all SymPy tree-elements, such as traversals.

<sup>11</sup>Some internal classes, such as those used in the polynomial submodule, do not follow this rule for efficiency reasons.

576 The children of a node in the tree are held in the `args` attribute. A terminal or leaf node in the  
577 expression tree has empty `args`.

578 For example, consider the expression  $xy + 2$ :

```
579 >>> x, y = symbols('x y')
580 >>> expr = x*y + 2
```

581 By order of operations, the parent of the expression tree for `expr` is an addition, so it is of type  
582 `Add`. The child nodes of `expr` are 2 and `x*y`.

```
583 >>> type(expr)
584 <class 'sympy.core.add.Add'>
585 >>> expr.args
586 (2, x*y)
```

587 Descending further down into the expression tree yields the full expression. For example,  
588 the next child node (given by `expr.args[0]`) is 2. Its class is `Integer`, and it has an empty `args`  
589 tuple, indicating that it is a leaf node.

```
590 >>> expr.args[0]
591 2
592 >>> type(expr.args[0])
593 <class 'sympy.core.numbers.Integer'>
594 >>> expr.args[0].args
595 ()
```

596 Symbols or symbolic constants, like  $e$  or  $\pi$ , are examples of leaf nodes.

```
597 >>> exp(1)
598 E
599 >>> exp(1).args
600 ()
601 >>> x.args
602 ()
```

603 A useful way to view an expression tree is using the `srepr` function, which returns a string  
604 representation of an expression as valid Python code<sup>12</sup> with all the nested class constructor calls  
605 to create the given expression.

```
606 >>> srepr(expr)
607 "Add(Mul(Symbol('x'), Symbol('y')), Integer(2))"
```

608 Every SymPy expression satisfies a key identity invariant:

```
609 expr.func(*expr.args) == expr
```

610 This means that expressions are rebuildable from their `args`.<sup>13</sup> Note that in SymPy the `==`  
611 operator represents exact structural equality, not mathematical equality. This allows testing if  
612 any two expressions are equal to one another as expression trees. For example, even though  
613  $(x+1)^2$  and  $x^2 + 2x + 1$  are equal mathematically, SymPy gives

```
614 >>> (x + 1)**2 == x**2 + 2*x + 1
615 False
```

---

<sup>12</sup> The `dotprint` function from the `sympy.printing.dot` submodule prints output to dot format, which can be rendered with Graphviz to visualize expression trees graphically.

<sup>13</sup>`expr.func` is used instead of `type(expr)` to allow the function of an expression to be distinct from its actual Python class. In most cases the two are the same.

616 because they are different as expression trees (the former is a `Pow` object and the latter is an `Add`  
617 object).

618 Another important property of SymPy expressions is that they are immutable. This simplifies  
619 the design of SymPy by allowing expression interning. It also enables expressions to be hashed,  
620 which is used to implement caching in SymPy.

621 Python allows classes to override mathematical operators. The Python interpreter translates  
622 the above `x*y + 2` to, roughly, `(x.__mul__(y)).__add__(2)`. Both `x` and `y`, returned from the  
623 `symbols` function, are `Symbol` instances. The `2` in the expression is processed by Python as a  
624 literal, and is stored as Python's built in `int` type. When `2` is passed to the `__add__` method  
625 of `Symbol`, it is converted to the SymPy type `Integer(2)` before being stored in the resulting  
626 expression tree. In this way, SymPy expressions can be built in the natural way using Python  
627 operators and numeric literals.

## 628 5.2 Extensibility

629 While the core of SymPy is relatively small, it has been extended to a wide variety of domains  
630 by a broad range of contributors. This is due, in part, to the fact that the same language,  
631 Python, is used both for the internal implementation and the external usage by users. All of  
632 the extensibility capabilities available to users are also utilized by SymPy itself. This eases the  
633 transition pathway from SymPy user to SymPy developer.

634 The typical way to create a custom SymPy object is to subclass an existing SymPy class,  
635 usually `Basic`, `Expr`, or `Function`. As it was stated before, all SymPy classes used for expression  
636 trees should be subclasses of the base class `Basic`. `Expr` is the `Basic` subclass for mathematical  
637 objects that can be added and multiplied together. The most commonly seen classes in SymPy  
638 are subclasses of `Expr`, including `Add`, `Mul`, and `Symbol`. Instances of `Expr` typically represent  
639 complex numbers, but may also include other “rings”, like matrix expressions. Not all SymPy  
640 classes are subclasses of `Expr`. For instance, logic expressions, such as `And(x, y)`, are subclasses  
641 of `Basic` but not of `Expr`.

642 The `Function` class is a subclass of `Expr` which makes it easier to define mathematical functions  
643 called with arguments. This includes named functions like  $\sin(x)$  and  $\log(x)$  as well as undefined  
644 functions like  $f(x)$ . Subclasses of `Function` should define a class method `eval`, which returns a  
645 canonical form of the function application (usually an instance of some other class, i.e., a `Number`)  
646 or `None`, if for given arguments that function should not be automatically evaluated.

647 Many SymPy functions perform various evaluations down the expression tree. Classes  
648 define their behavior in such functions by defining a relevant `_eval *` method. For instance,  
649 an object can indicate to the `diff` function how to take the derivative of itself by defining the  
650 `_eval_derivative(self, x)` method, which may in turn call `diff` on its `args`. (Subclasses of  
651 `Function` should implement `fdiff` method instead, it returns the derivative of the function without  
652 considering the chain rule.) The most common `_eval *` methods relate to the assumptions:  
653 `_eval_is_assumption` is used to deduce *assumption* on the object.

654 As an example of the notions presented in this section, Listing 1 presents a minimal version of  
655 the gamma function  $\Gamma(x)$  from SymPy, which evaluates itself on positive integer arguments, has the  
656 real assumption defined, can be rewritten in terms of factorial with `gamma(x).rewrite(factorial)`,  
657 and can be differentiated. `self.func` is used throughout instead of referencing `gamma` explicitly  
658 so that potential subclasses of `gamma` can reuse the methods.

**Listing 1.** A minimal implementation of `sympy.gamma`.

```
659 from sympy import Integer, Function, floor, factorial, polygamma
660
661 class gamma(Function)
662     @classmethod
663     def eval(cls, arg):
664         if isinstance(arg, Integer) and arg.is_positive:
665             return factorial(arg - 1)
666
667     def _eval_is_real(self):
668         x = self.args[0]
```

```

669         # noninteger means real and not integer
670         if x.is_positive or x.is_noninteger:
671             return True
672
673     def _eval_rewrite_as_factorial(self, z):
674         return factorial(z - 1)
675
676     def fdiff(self, argindex=1):
677         from sympy.core.function import ArgumentIndexError
678         if argindex == 1:
679             return self.func(self.args[0])*polygamma(0, self.args[0])
680         else:
681             raise ArgumentIndexError(self, argindex)

```

682 The gamma function implemented in SymPy has many more capabilities than the above listing,  
683 such as evaluation at rational points and series expansion.

### 684 5.3 Performance

685 Due to being written in pure Python without the use of extension modules, SymPy's performance  
686 characteristics are generally poorer than its commercial competitors. For many applications,  
687 the performance of SymPy, as measured by clock cycles, memory usage, and memory layout, is  
688 sufficient. However, the boundaries for when SymPy's pure Python strategy becomes insufficient  
689 are when the user requires handling of very long expressions or many small expressions. Where  
690 this boundray lies depends on the system at hand, but tends to be within the range of  $10^4$ – $10^6$   
691 symbols for modern computers.

692 For this reason, a new project called SymEngine [59] has been started. The aim of this  
693 project is to develop a library with better performance characteristics for symbolic manipulation.  
694 SymEngine is a pure C++ library, which allows it fine grain control over the memory layout of  
695 expressions. SymEngine has thin wrappers to other languages (Python, Ruby, Julia, etc.) whose  
696 aim is to be the fastest symbolic manipulation library. Preliminary benchmarks suggest that  
697 SymEngine is as performant as its commercial and open source competitors.

698 SymPy has recently started to use SymEngine as an optional backend, initially in `sympy.`  
699 `physics.mechanics` only. Future work will involve allowing more algorithms in SymPy to use  
700 SymEngine as a backend.

## 701 6 PROJECTS THAT DEPEND ON SYMPY

702 There are several projects that depend on SymPy as a library for implementing a part of their  
703 functionality. A selection of these projects are listed in Table 3.

**Table 3.** Selected projects that depend on SymPy.

| Project name                                | Description  |
|---|--|
| <a href="#">SymPy Gamma</a>                 | An open source analog of Wolfram Alpha that uses SymPy [60]. There is more information about SymPy Gamma supplementary material.                             |
| <a href="#">Cadabra</a>                     | A CAS designed specifically for the resolution of problems encountered in field theory [39].   |
| <a href="#">GNU Octave Symbolic Package</a> | An implementation of a symbolic toolbox for Octave using SymPy [58].   |
| <a href="#">SymPy.jl</a>                    | A Julia interface to SymPy, provided using PyCall [61].  |
| <a href="#">Mathics</a>                     | A free, online CAS featuring Mathematica compatible syntax and functions [55].   |
| <a href="#">Mathpix</a>                     | An iOS App that detects handwritten math as input and uses SymPy Gamma to evaluate the math input and generate the relevant steps to solve the problem [33]. |

|                 |  |
|-----------------|--|
| <b>IKFast</b>   | A robot kinematics compiler provided by <a href="#">OpenRAVE</a> [12].   |
| <b>SageMath</b> | A free open-source mathematics software system, which builds on top of many existing open-source packages, including SymPy [57]. |
| <b>PyDy</b>     | Multibody Dynamics with Python [19].   |
| <b>galgebra</b> | A module for geometric algebra (previously <code>sympy.galgebra</code> ) [5].  |
| <b>yt</b>       | A Python package for analyzing and visualizing volumetric data [63].   |
| <b>SfePy</b>    | A Python package for solving partial differential equations (PDEs) in 1D, 2D, and 3D by the finite element (FE) method [65, 10]. |
| <b>Quameon</b>  | Quantum Monte Carlo in Python [56].  |
| <b>Lcapy</b>    | An experimental Python package for teaching linear circuit analysis [54].  |

---

## 7 CONCLUSION AND FUTURE WORK

SymPy is a robust computer algebra system that provides a wide spectrum of features both in traditional computer algebra and in a plethora of scientific disciplines. This allows SymPy to be used in a first-class way with other Python projects, including the scientific Python stack.

SymPy supports a wide array of mathematical facilities. This includes functions for assuming and deducing common mathematical facts, simplifying expressions, performing common calculus operations, manipulating polynomials, pretty printing expressions, solving equations, and representing symbolic matrices. Other supported facilities include discrete math, concrete math, plotting, geometry, statistics, sets, series, vectors, combinatorics, group theory, code generation, tensors, Lie algebras, cryptography, and special functions. SymPy has strong support for arbitrary precision numerics, backed by the mpmath package. Additionally, SymPy contains submodules targeting certain specific physics domains, such as classical mechanics and quantum mechanics. This breadth of domains has been engendered by a strong and vibrant user community. Anecdotally, many of these users chose SymPy because of its ease of access. SymPy a dependency of many external projects across a wide spectrum of domains.

SymPy expressions are immutable trees of Python objects. Unlike many other CAS's, SymPy is designed to be used in an extensible way: both as an end-user application and as a library. SymPy uses Python both as the internal language and the user language. This permits users to access to the same methods used by the library itself in order to extend it for their needs.

Some of the planned future work for SymPy includes work on improving code generation, improvements to the speed of SymPy using SymEngine, improving the assumptions system, and improving the solvers submodule.

## REFERENCES

- [1] Adams, W. W. and Loustaunau, P. (1994). *An introduction to Gröbner bases*. Number 3. American Mathematical Society.
- [2] Bailey, D. H., Jeyabalan, K., and Li, X. S. (2005). A comparison of three high-precision quadrature schemes. *Experimental Mathematics*, 14(3):317–329.
- [3] Bender, C. M. and Orszag, S. A. (1999). *Advanced Mathematical Methods for Scientists and Engineers*. Springer, 1st edition.
- [4] Biggs, N., Lloyd, E. K., and Wilson, R. J. (1976). *Graph Theory, 1736-1936*. Oxford University Press.
- [5] Bromborsky, A. (2016). *geometric algebra/calculus modules for SymPy*. <https://github.com/brombo/galgebra>.
- [6] Bronstein, M. (2005a). pmint—The Poor Man’s Integrator. <http://www-sop.inria.fr/cafe/Manuel.Bronstein/pmint>.
- [7] Bronstein, M. (2005b). *Symbolic Integration I: Transcendental Functions*. Springer-Verlag, New York, NY, USA.
- [8] Buchberger, B. (1965). *Ein Algorithmus zum Auffinden der Basis Elemente des Restklassenrings nach einem nulldimensionalen Polynomideal*. PhD thesis, University of Innsbruck, Innsbruck, Austria.
- [9] Cervone, D. (2012). Mathjax: a platform for mathematics on the web. *Notices of the AMS*, 59(2):312–316.

- [10] Cimrman, R. (2014). SfePy - write your own FE application. In de Buyl, P. and Varoquaux, N., editors, *Proceedings of the 6th European Conference on Python in Science (EuroSciPy 2013)*, pages 65–70. <http://arxiv.org/abs/1404.6391>.
- [11] Ciurana, E. (2009). Google app engine. *Developing with Google App Engine*, pages 1–10.
- [12] Diankov, R. (2010). Ikfast: The robot kinematics compiler. [http://openrave.org/docs/latest\\_stable/openravepy/ikfast/](http://openrave.org/docs/latest_stable/openravepy/ikfast/).
- [13] Doorenbos, R. B. (1995). *Production matching for large learning systems*. PhD thesis, University of Southern California.
- [14] Faugère, J. C. (1999). A New Efficient Algorithm for Computing Gröbner Bases (F4). *Journal of Pure and Applied Algebra*, 139(1-3):61–88.
- [15] Faugère, J. C. (2002). A New Efficient Algorithm for Computing Gröbner Bases Without Reduction To Zero (F5). In *ISSAC '02: Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation*, pages 75–83, New York, NY, USA. ACM Press.
- [16] Fetter, A. and Walecka, J. (2003). *Quantum Theory of Many-Particle Systems*. Dover Publications.
- [17] Fousse, L., Hanrot, G., Lefèvre, V., Pélissier, P., and Zimmermann, P. (2007). Mpmc: A multiple-precision binary floating-point library with correct rounding. *ACM Trans. Math. Softw.*, 33(2).
- [18] Fu, H., Zhong, X., and Zeng, Z. (2006). Automated and Readable Simplification of Trigonometric Expressions. *Mathematical and Computer Modelling*, 55(11-12):1169–1177.
- [19] Gede, G., Peterson, D. L., Nanjangud, A. S., Moore, J. K., and Hubbard, M. (2013). Constrained multibody dynamics with Python: From symbolic equation generation to publication. In *ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages V07BT10A051–V07BT10A051. American Society of Mechanical Engineers.
- [20] Goldberg, D. (1991). What every computer scientist should know about floating-point arithmetic. *ACM Computing Surveys (CSUR)*, 23(1):5–48.
- [21] Gosper, R. W. (1978). Decision procedure for indefinite hypergeometric summation. *Proceedings of the National Academy of Sciences*, 75(1):40–42.
- [22] Gruntz, D. (1996). *On Computing Limits in a Symbolic Manipulation System*. PhD thesis, Swiss Federal Institute of Technology, Zürich, Switzerland.
- [23] Horsen, C. V. (2015). GMPY. <https://pypi.python.org/pypi/gmpy2>.
- [24] Hudak, P. (1998). Domain specific languages. In Salas, P. H., editor, *Handbook of Programming Languages, Vol. III: Little Languages and Tools*, chapter 3, pages 39–60. MacMillan, Indianapolis.
- [25] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3):90–95.
- [26] Johansson, F. et al. (2014). mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.19). <http://mpmath.org/>.
- [27] Jones, E., Oliphant, T., Peterson, P., et al. (2001–). SciPy: Open source scientific tools for Python. [Online; accessed 2016-09-28].
- [28] Kane, T. R. and Levinson, D. A. (1985). *Dynamics, Theory and Applications*. McGraw Hill.
- [29] Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., et al. (2016). Jupyter notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing*, page 87. IOS Press.
- [30] Lagrange, J. (1811). *Mécanique analytique*. Number v. 1 in *Mécanique analytique*. Ve Courcier.
- [31] Lang, S. (1966). Introduction to transcendental numbers. *Reading, Mass.*
- [32] Lutz, M. (2013). *Learning Python*. O’Reilly Media, Inc.
- [33] Mathpix, Inc. (2016). *Mathpix — Solve and graph math using pictures*. <http://mathpix.com/>.
- [34] Moses, J. (1971). Algebraic Simplification: A Guide for the Perplexed. In *SYMSAC '71: Proceedings of the second ACM Symposium on Symbolic and Algebraic Computation*, pages 282–304, New York, NY, USA. ACM Press.



- [35] Nielsen, M. and Chuang, I. (2011). *Quantum Computation and Quantum Information*. Cambridge University Press.
- [36] Nijenhuis, A. and Wilf, H. S. (1978). *Combinatorial Algorithms: For Computers and Calculators*. Academic Press, New York, NY, USA, second edition.
- [37] Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science & Engineering*, 9(3):10–20.
- [38] Paprocki, M. (2010). Design and implementation issues of a computer algebra system in an interpreted, dynamically typed programming language. Master’s thesis, University of Technology of Wrocław, Poland.
- [39] Peeters, K. (2007). Cadabra: a field-theory motivated symbolic computer algebra system. *Computer Physics Communications*.
- [40] Pérez, F. and Granger, B. E. (2007). IPython: a system for interactive scientific computing. *Computing in Science & Engineering*, 9(3):21–29.
- [41] Peterson, D. L., Gede, G., and Hubbard, M. (2014). Symbolic linearization of equations of motion of constrained multibody systems. *Multibody System Dynamics*, 33(2):143–161.
- [42] Petkovšek, M., Wilf, H. S., and Zeilberger, D. (1996). A=BAK peters. Wellesley, MA.
- [43] Raymond, E. (1999). The cathedral and the bazaar. *Knowledge, Technology & Policy*, 12(3):23–49.
- [44] Roach, K. (1996). Hypergeometric function representations. In *ISSAC ’96: Proceedings of the 1996 International Symposium on Symbolic and Algebraic Computation*, pages 301–308, New York, NY, USA. ACM Press.
- [45] Roach, K. (1997). Meijer G function representations. In *ISSAC ’97: Proceedings of the 1997 international symposium on Symbolic and algebraic computation*, pages 205–211, New York, NY, USA. ACM.
- [46] Rocklin, M. and Terrel, A. R. (2012). Symbolic statistics with SymPy. *Computing in Science and Engineering*, 14.
- [47] Rose, K. H. et al. (1999). Xy-pic user’s guide.
- [48] Rosen, L. (2005). *Open source licensing*, volume 692. Prentice Hall.
- [49] Sakurai, J. and Napolitano, J. (2010). *Modern Quantum Mechanics*. Addison-Wesley.
- [50] Shaw, M. and Garlan, D. (1996). *Software Architecture: Perspectives on an Emerging Discipline*. Prentice Hall. Prentice Hall Ordering Information.
- [51] Sussman, G. J. and Wisdom, J. (2013). *Functional Differential Geometry*. Massachusetts Institute of Technology Press.
- [52] Tai, C.-T. (1997). *Generalized vector and dyadic analysis: applied mathematics in field theory*, volume 9. Wiley-IEEE Press.
- [53] Takahasi, H. and Mori, M. (1974). Double exponential formulas for numerical integration. *Publications of the Research Institute for Mathematical Sciences*, 9(3):721–741.
- [54] The Lcapy Developers (2016). *Lcapy, a Python package for linear circuit analysis*. <http://lcapy.elec.canterbury.ac.nz/>.
- [55] The Mathics Developers (2016). *Mathics, a free, general-purpose online computer algebra system featuring Mathematica-compatible syntax and functions*. <https://mathics.github.io/>.
- [56] The Quameon Developers (2016). *Quameon, Quantum Monte Carlo in Python*. <http://quameon.sourceforge.net/>.
- [57] The Sage Developers (2016). *SageMath, the Sage Mathematics Software System*. <http://www.sagemath.org>.
- [58] The Symbolic Package Developers (2016). *The Symbolic Package for GNU Octave*. <http://octave.sourceforge.net/symbolic>.
- [59] The SymPy Developers (2016a). *SymEngine, a fast symbolic manipulation library, written in C++*. <https://github.com/symengine/symengine>.
- [60] The SymPy Developers (2016b). *SymPy Gamma*. <http://www.sympygamma.com/>.
- [61] The SymPy.jl Developers (2016). *SymPy.jl, a package to bring Python’s Sympy functionality into Julia via PyCall*. <https://github.com/JuliaPy/SymPy.jl>.
- [62] Toth, V. T. (2007). Maple and Meijer’s G-function: a numerical instability and a cure. <http://www.vttoth.com/CMS/index.php/technical-notes/67>.
- [63] Turk, M. J., Smith, B. D., Oishi, J. S., Skory, S., Skillman, S. W., Abel, T., and Norman,

- 856 M. L. (2011). yt: A Multi-code Analysis Toolkit for Astrophysical Simulation Data. *The*  
857 *Astrophysical Journal Supplement Series*, 192:9.
- 858 [64] Zare, R. (1991). *Angular Momentum: Understanding Spatial Aspects in Chemistry and*  
859 *Physics*. Wiley.
- 860 [65] Zienkiewicz, O., Taylor, R., and Zhu, J. (2013). *The Finite Element Method: Its Basis and*  
861 *Fundamentals*. Butterworth-Heinemann, seventh edition edition.