

# SYMPY: SYMBOLIC COMPUTING IN PYTHON

AARON MEURER\*, MATEUSZ PAPROCKI†, ONDŘEJ ČERTÍK‡, MATTHEW ROCKLIN§,  
AMIT KUMAR¶, SERGIU IVANOV||, JASON K. MOORE#, SARTAJ SINGH††, THILINA  
RATHNAYAKE‡‡ SEAN VIG§§ BRIAN E. GRANGER¶¶ RICHARD P. MULLER|||  
FRANCESCO BONAZZI## HARSH GUPTA¹, SHIVAM VATS², FREDRIK JOHANSSON³,  
FABIAN PEDREGOSA⁴, ASHUTOSH SABOO⁵, ISURU FERNANDO⁶, SUMITH⁷, ROBERT  
CIMRMAN⁸, AND ANTHONY SCOPATZ⁹

**1. Introduction.** SymPy is a full featured computer algebra system (CAS) written in the Python programming language. It is free and open source software, being licensed under the 3-clause BSD license. 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 allows SymPy to rapidly respond to the needs of the community of users and developers.

Python is a dynamically typed programming language that has a focus on ease of use and readability. Due in part to this focus, it has become a popular language for scientific computing and data science, with a broad ecosystem of libraries [38]. SymPy is itself used by many libraries and tools to support research within a variety

---

\*University of South Carolina, Columbia, SC 29201 ([asmeurer@gmail.com](mailto:asmeurer@gmail.com)).

†Continuum Analytics, Inc., Austin, TX 78701 ([mattpap@gmail.com](mailto:mattpap@gmail.com)).

‡Los Alamos National Laboratory, Los Alamos, NM 87545 ([certik@lanl.gov](mailto:certik@lanl.gov)).

§Continuum Analytics, Inc., Austin, TX 78701 ([mrocklin@gmail.com](mailto:mrocklin@gmail.com)).

¶Delhi Technological University, Shahbad Daultpur, Bawana Road, New Delhi 110042, India ([dtu.amit@gmail.com](mailto:dtu.amit@gmail.com)).

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

#University of California, Davis, Davis, CA 95616 ([jkm@ucdavis.edu](mailto:jkm@ucdavis.edu)).

††Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh 221005, India ([singhsartaj94@gmail.com](mailto:singhsartaj94@gmail.com)).

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

§§University of Illinois at Urbana-Champaign, Urbana, IL 61801 ([sean.v.775@gmail.com](mailto:sean.v.775@gmail.com)).

¶¶California Polytechnic State University, San Luis Obispo, CA 93407 ([ellisonbg@gmail.com](mailto:ellisonbg@gmail.com)).

|||Center for Computing Research, Sandia National Laboratories, Albuquerque, NM 87185 ([rmuller@sandia.gov](mailto:rmuller@sandia.gov)).

##Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, 14476 Potsdam, Germany ([francesco.bonazzi@mpikg.mpg.de](mailto:francesco.bonazzi@mpikg.mpg.de)).

¹Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India ([har-gup@protonmail.com](mailto:har-gup@protonmail.com)).

²Indian Institute of Technology Kharagpur, Kharagpur, West Bengal 721302, India ([shivamvats.iitkgp@gmail.com](mailto:shivamvats.iitkgp@gmail.com)).

³INRIA Bordeaux-Sud-Ouest - LFANT project-team, 200 Avenue de la Vieille Tour, 33405 Talence, France ([fredrik.johansson@gmail.com](mailto:fredrik.johansson@gmail.com)).

⁴INRIA - SIERRA project-team, 2 Rue Simone IFF, 75012 Paris, France ([f@bianp.net](mailto:f@bianp.net)).

⁵Birla Institute of Technology and Science, Pilani, K.K. Birla Goa Campus, NH 17B Bypass Road, Zuarinagar, Sancoale, Goa 403726, India ([ashutosh.saboo@gmail.com](mailto:ashutosh.saboo@gmail.com)).

⁶University of Moratuwa, Bandaranayake Mawatha, Katubedda, Moratuwa 10400, Sri Lanka ([isuru.11@cse.mrt.ac.lk](mailto:isuru.11@cse.mrt.ac.lk)).

⁷Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra 400076, India ([sumith@cse.iitb.ac.in](mailto:sumith@cse.iitb.ac.in)).

⁸New Technologies - Research Centre, University of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic ([cimrman3@ntc.zcu.cz](mailto:cimrman3@ntc.zcu.cz)).

⁹University of South Carolina, Columbia, SC 29201 ([scopatz@cec.sc.edu](mailto:scopatz@cec.sc.edu)).

domains, such as Sage [48] (pure mathematics), yt [52] (astronomy and astrophysics), PyDy [25] (multibody dynamics), and SfePy [19] (finite elements).

Unlike many CASs, SymPy does not invent its own programming language. Python itself is used both for the internal implementation and the end user interaction. 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 is designed with a strong focus on usability as a library. Extensibility is important in its application program interface (API) design, and 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 is in an interactive environment or programmatic use as part of a larger system.

As a library, SymPy does not have a built-in graphical user interface (GUI). However, SymPy exposes a rich interactive display system, including registering printers with Jupyter [40] frontends, including the Notebook and Qt Console, which will render SymPy expressions using MathJax [18] or L<sup>A</sup>T<sub>E</sub>X.

The remainder of this paper discusses key components of the SymPy software. Section 2 discusses the architecture of SymPy. Section 3 enumerates the features of SymPy and takes a closer look at some of the important ones. Following that, section 4 looks at the numerical features of SymPy and its dependency library, mpmath. Section 5 looks at the domain specific physics submodules for doing classical mechanics and quantum mechanics. Finally, section 6 concludes the paper and discusses future work.

## 2. Architecture.

**2.1. Basic Usage.** Because SymPy is built on Python, it requires that all variable names be defined prior to use. The following statement imports all SymPy functions into the global Python namespace. From here on, all examples in this paper assume that this statement has been run.

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

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 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 mathematical syntax. Note that in Python, exponentiation is represented by `**`. For instance, the following Python code creates the expression  $(x^2 - 2x + 3)/y$ .

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

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

Importantly, SymPy expressions are immutable. This simplifies the design of SymPy by allowing expression interning. It also enables expressions to be hashed and stored in Python dictionaries, thereby permitting caching and other features.

**2.2. The Core.** A computer algebra system (CAS) represents mathematical expressions as data structures. For example the mathematical expression  $x + y$  is represented as a tree with three nodes,  $+$ ,  $x$ , and  $y$ , where  $x$  and  $y$  are ordered children of  $+$ . As users manipulate mathematical expressions with traditional mathematical syntax, the CAS manipulates the underlying data structures. Automated optimizations and computations such as integration, simplification, etc. are all functions that consume and produce expression trees.

In SymPy every symbolic expression is an instance of a Python `Basic` class, a superclass of all SymPy types providing common methods to all SymPy tree-elements, such as traversals. The children of a node in the tree are held in the `args` attribute. A terminal or leaf node in the expression tree has empty `args`.

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

```
>>> expr = x*y + 2
```

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

```
>>> type(expr)
```

```
<class 'sympy.core.add.Add'>
```

```
>>> expr.args
```

```
(2, x*y)
```

Traversing further into the expression tree grants the full expression. For example, the first child node, given by `expr.args[0]`, is 2. Its class is `Integer`, and it has empty `args` tuple, indicating that it is a leaf node.

```
>>> expr.args[0]
```

```
2
```

```
>>> type(expr.args[0])
```

```
<class 'sympy.core.numbers.Integer'>
```

```
>>> expr.args[0].args
```

```
()
```

A useful way to view an expression tree is with the `srepr` function. This returns a string representation of an expression as valid Python code with all the nested class constructor calls to create the given expression.

```
>>> srepr(expr)
```

```
"Add(Mul(Symbol('x'), Symbol('y')), Integer(2))"
```

Every SymPy expression satisfies a key identity invariant:

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

This means that expressions are rebuildable from their `args`.<sup>1</sup> We note that in SymPy, the `==` operator represents exact structural equality, not mathematical equality. This allows one to test if any two expressions are equal to one another as expression trees.

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

**2.3. Assumptions.** SymPy performs logical inference through its assumptions system. The assumptions system allows users to specify that symbols have cer-

---

<sup>1</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.

tain common mathematical properties, such as being positive, imaginary, or integral. 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 ( $t \geq 0$ ). If  $t$  is real, the identity  $\sqrt{t^2} = |t|$  holds. 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 symbols are 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.

```
>>> t = Symbol('t', positive=True)
>>> sqrt(t**2)
t
```

Some of the common assumptions that SymPy allows are `positive`, `negative`, `real`, `nonpositive`, `nonnegative`, `integer`, and `commutative`.<sup>2</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. It is 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 builtin objects `True`, `False`, and `None`. `None` represents the “unknown” case. This could mean that the given assumption could be either true or false under the given information, for instance, `Symbol('x', real=True).is_positive` will give `None` because a real symbol might be positive or it might not. It could also mean not enough is implemented to compute the given fact. For instance, `(pi + E).is_irrational` gives `None`, because SymPy does not know how to determine if  $\pi + e$  is rational or irrational, indeed, it is an open problem in mathematics.

Basic implications between the facts are used to deduce assumptions. For instance, the assumptions system knows that being an integer implies being rational, so `Symbol('x', integer=True).is_rational` returns `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`.

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

---

<sup>2</sup>If  $A$  and  $B$  are Symbols created with `commutative=False` then SymPy will keep  $A \cdot B$  and  $B \cdot A$  distinct.

160 to SymPy developer.

161 The typical way to create a custom SymPy object is to subclass an existing SymPy  
162 class, usually `Basic`, `Expr`, or `Function`. All SymPy classes used for expression trees<sup>3</sup>  
163 should be subclasses of the base class `Basic`, which defines some basic methods for  
164 symbolic expression trees. `Expr` is the subclass for mathematical expressions that  
165 can be added and multiplied together. Instances of `Expr` typically represent complex  
166 numbers, but may also include other “rings” like matrix expressions. Not all SymPy  
167 classes are subclasses of `Expr`. For instance, logic expressions such as `And(x, y)` are  
168 subclasses of `Basic` but not of `Expr`.

169 The `Function` class is a subclass of `Expr` which makes it easier to define mathe-  
170 matical functions called with arguments. This includes named functions like `sin(x)`  
171 and `log(x)` as well as undefined functions like `f(x)`. Subclasses of `Function` should  
172 define a class method `eval`, which returns values for which the function should be  
173 automatically evaluated, and `None` for arguments that should not be automatically  
174 evaluated.

175 Many SymPy functions perform various evaluations down the expression tree.  
176 Classes define their behavior in such functions by defining a relevant `_eval_*` method.  
177 For instance, an object can indicate to the `diff` function how to take the derivative  
178 of itself by defining the `_eval_derivative(self, x)` method, which may in turn call  
179 `diff` on its args. The most common `_eval_*` methods relate to the assumptions.  
180 `_eval_is_assumption` defines the assumptions for *assumption*.

181 As an example of the notions presented in this section, Listing 1 presents a  
182 stripped down version of the gamma function  $\Gamma(x)$  from SymPy, which evaluates  
183 itself on positive integer arguments, has the positive and real assumptions defined,  
184 can be rewritten in terms of factorial with `gamma(x).rewrite(factorial)`, and can  
185 be differentiated. `fdiff` is a convenience method for subclasses of `Function`. `fdiff`  
186 returns the derivative of the function without considering the chain rule. `self.func`  
187 is used throughout instead of referencing `gamma` explicitly so that potential subclasses  
188 of `gamma` can reuse the methods.

Listing 1: A stripped down version of `sympy.gamma`.

```
189 from sympy import Integer, Function, floor, factorial, polygamma
190
191 class gamma(Function)
192     @classmethod
193     def eval(cls, arg):
194         if isinstance(arg, Integer) and arg.is_positive:
195             return factorial(arg - 1)
196
197     def _eval_is_positive(self):
198         x = self.args[0]
199         if x.is_positive:
200             return True
201         elif x.is_noninteger:
202             return floor(x).is_even
203
204     def _eval_is_real(self):
```

---

<sup>3</sup>Some internal classes, such as those used in the polynomial module, do not follow this rule for efficiency reasons.

```

205     x = self.args[0]
206     # noninteger means real and not integer
207     if x.is_positive or x.is_noninteger:
208         return True
209
210     def _eval_rewrite_as_factorial(self, z):
211         return factorial(z - 1)
212
213     def fdiff(self, argindex=1):
214         from sympy.core.function import ArgumentIndexError
215         if argindex == 1:
216             return self.func(self.args[0])*polygamma(0, self.args[0])
217         else:
218             raise ArgumentIndexError(self, argindex)

```

219 The actual gamma function defined in SymPy has many more capabilities, such as  
220 evaluation at rational points and series expansion.

221 **3. Features.** SymPy has an extensive feature set that is too encompassing to  
222 cover each topic in-depth in this paper. Here though, calculus and other bedrock  
223 areas receive their own subsections. Additionally, Table 1 gives a compact listing of  
224 all major capabilities present in the SymPy codebase. This grants a sampling from  
225 the breadth of topics and application domains that SymPy services. Unless stated  
226 otherwise, all features noted in Table 1 are symbolic in nature. Numeric features are  
227 discussed in Section 4.

Table 1: SymPy Features and Descriptions

Feature	Description
Calculus	Algorithms for computing derivatives, integrals, and limits.
Category Theory	Representation of objects, morphisms, and diagrams. Tools for drawing diagrams with Xy-pic.
Code Generation	Enables generation of compilable and executable code in a variety of different programming languages directly from expressions. Target languages include C, Fortran, Julia, JavaScript, Mathematica, Matlab and Octave, Python, and Theano.
Combinatorics & Group Theory	Implements permutations, combinations, partitions, subsets, various permutation groups (such as polyhedral, Rubik, symmetric, and others), Gray codes [37], and Prufer sequences [13].
Concrete Math	Summation, products, tools for determining whether summation and product expressions are convergent, absolutely convergent, hypergeometric, and other properties. May also compute Gosper’s normal form [42] for two univariate polynomials.

Cryptography	Represents block and stream ciphers, including shift, Affine, substitution, Vigenere's, Hill's, bifid, RSA, Kid RSA, linear-feedback shift registers, and Elgamal encryption
Differential Geometry	Classes to represent manifolds, metrics, tensor products, and coordinate systems in Riemannian and pseudo-Riemannian geometries [49].
Geometry	Allows the creation 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 two lines.
Lie Algebras	Represents Lie algebras and root systems.
Logic	Boolean expression, equivalence testing, satisfiability, and normal forms.
Matrices	Tools for creating matrices of symbols and expressions. This is capable of both sparse and dense representations and performing symbolic linear algebraic operations (e.g., inversion and factorization).
Matrix Expressions	Matrices with symbolic dimensions (unspecified entries). Block matrices.
Number Theory	Prime number generation, primality testing, integer factorization, continued fractions, Egyptian fractions, modular arithmetic, quadratic residues, partitions, binomial and multinomial coefficients, prime number tools, and integer factorization.
Plotting	Hooks for visualizing expressions via matplotlib [30] or as text drawings when lacking a graphical back-end. 2D function plotting, 3D function plotting, and 2D implicit function plotting are supported.
Polynomials	Computes polynomial algebras over various coefficient domains. Functionality ranges from the simple (e.g., polynomial division) to the advanced (e.g., Gröbner bases [10] and multivariate factorization over algebraic number domains).
Printing	Functions for printing SymPy expressions in the terminal with ASCII or Unicode characters and converting SymPy expressions to L <sup>A</sup> T <sub>E</sub> X and MathML.
Quantum Mechanics	Quantum states, bra-ket notation, operators, basis sets, representations, tensor products, inner products, outer products, commutators, anticommutators.
Series	Implements series expansion, sequences, and limit of sequences. This includes Taylor, Laurent, and Puiseux series as well as special series, such as Fourier and formal power series.
Sets	Representations of empty, finite, and infinite sets. This includes special sets such as for all natural, integer, and complex numbers. Operations on sets such as union, intersection, Cartesian product, and building sets from other sets.



Simplification	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	Functions for symbolically solving equations algebraically, systems of equations, both linear and non-linear, inequalities, ordinary differential equations, partial differential equations, Diophantine equations, and recurrence relations.
Special Functions	Implements 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	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 [44].
Tensors	Symbolic manipulation of indexed objects.
Vectors	Provides basic vector math and differential calculus with respect to 3D Cartesian coordinate systems.

**3.1. Simplification.** The generic way to simplify an expression is by calling the `simplify` function. It must be emphasized that simplification is not an unambiguously defined mathematical operation [17]. The `simplify` function applies several simplification routines along with some heuristics to make the output expression as “simple” as possible.

It is often preferable to apply more directed simplification functions. These apply very specific rules to the input expression and are often 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 some common simplification functions.

Table 2: Some SymPy Simplification Functions

<code>expand</code>	expand the expression <pre>&gt;&gt;&gt; expand((x + y)**3) x**3 + 3*x**2*y + 3*x*y**2 + y**3</pre>
<code>factor</code>	factor a polynomial into irreducibles <pre>&gt;&gt;&gt; factor(x**3 + 3*x**2*y + 3*x*y**2 + y**3) (x + y)**3</pre>



```

collect    collect polynomial coefficients
>>> collect(y*x**2 + 3*x**2 - x*y + x - 1, x)
x**2*(y + 3) + x*(-y + 1) - 1

cancel     rewrite a rational function as  $p/q$  with common factors canceled
>>> cancel((x**2 + 2*x + 1)/(x**2 - 1))
(x + 1)/(x - 1)

apart      compute the partial fraction decomposition of a rational function
>>> apart((x**3 + 4*x - 1)/(x**2 - 1))
x + 3/(x + 1) + 2/(x - 1)

trigsimp   simplify trigonometric expressions [23]
>>> trigsimp(cos(x)**2*tan(x) - sin(2*x))
-sin(2*x)/2

```

---

Substitutions are performed through the `.subs` method.

```

>>> (sin(x) + x**2 + 1).subs(x, y + 1)
(y + 1)**2 + sin(y + 1) + 1

```

**3.2. Calculus.** Integrals are calculated with the `integrate` function. SymPy implements a combination of the Risch algorithm [16], table lookups, a reimplementation of Manuel Bronstein’s “Poor Man’s Integrator” [15], and an algorithm for computing integrals based on Meijer G-functions. These allow SymPy to compute a wide variety of indefinite and definite integrals.

```

>>> integrate(sin(x), x)
-cos(x)
>>> integrate(sin(x), (x, 0, 1))
-cos(1) + 1

```

Derivatives are computed with the `diff` function. Derivatives are computed recursively using the various differentiation rules.

```

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

```

Summations and products are computed with `summation` and `product`, respectively. Summations are computed using a combination of Gosper’s algorithm, an algorithm that uses Meijer G-functions, and heuristics. Products are computed via some heuristics.

Limits are computed with the `limit` function. The limit module implements the Gruntz algorithm [27] for computing symbolic limits. For example, the following computes  $\lim_{x \rightarrow \infty} x \sin(\frac{1}{x}) = 1$  (note that  $\infty$  is `oo` in SymPy).

```

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

```

As a more complicated 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*E**((1-cos(x))/sin(x))-1)**(sinh(x)/atan(x)**2), x, 0)
E

```

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)

```

269 `Integral(x**x, x)`

270 **3.3. Polynomials.** SymPy implements a suite of algorithms for polynomial ma-  
 271 nipulation, which ranges from relatively simple algorithms for doing arithmetics of  
 272 polynomials, to advanced methods for factoring multivariate polynomials into irre-  
 273 reducibles, symbolically determining real and complex root isolation intervals, or com-  
 274 puting Gröbner bases.

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

282 SymPy implements dense and sparse polynomial representations. Both are used  
 283 in the univariate and multivariate cases. The dense representation is the default for  
 284 univariate polynomials. For multivariate polynomials, the choice of representation is  
 285 based on the application. The most common case for the sparse representation is  
 286 algorithms for computing Gröbner bases (Buchberger, F4, and F5). This is because  
 287 different monomial orderings can be expressed easily in this representation. How-  
 288 ever, algorithms for computing multivariate GCDs or factorizations, at least those  
 289 currently implemented in SymPy, are better expressed when the representation is  
 290 dense. The dense multivariate representation is specifically a recursively dense rep-  
 291 resentation, where polynomials in  $K[x_0, x_1, \dots, x_n]$  are viewed as a polynomials in  
 292  $K[x_0][x_1] \dots [x_n]$ . Note that despite this, the coefficient domain  $K$ , can be a multi-  
 293 variate polynomial domain as well. The dense recursive representation in Python gets  
 294 inefficient when the number of variables gets high.

295 Here are some examples of the `sympy.polys` submodule. Factorization:

```
296 >>> t = symbols("t")
297 >>> f = (2115*x**4*y + 45*x**3*z**3*t**2 - 45*x**3*t**2 - 423*x*y**4 -
298 ...      47*x*y**3 + 141*x*y*z**3 + 94*x*y*z*t - 9*y**3*z**3*t**2 +
299 ...      9*y**3*t**2 - y**2*z**3*t**2 + y**2*t**2 + 3*z**6*t**2 +
300 ...      2*z**4*t**3 - 3*z**3*t**2 - 2*z*t**3)
301 >>> factor(f)
302 (t**2*z**3 - t**2 + 47*x*y)*(2*t*z + 45*x**3 - 9*y**3 - y**2 + 3*z**3)
303 Gröbner bases:
304 >>> x0, x1, x2 = symbols('x:3')
305 >>> I = [x0 + 2*x1 + 2*x2 - 1,
306 ...      x0**2 + 2*x1**2 + 2*x2**2 - x0,
307 ...      2*x0*x1 + 2*x1*x2 - x1]
308 >>> groebner(I, order='lex')
309 GroebnerBasis([7*x0 - 420*x2**3 + 158*x2**2 + 8*x2 - 7,
310 7*x1 + 210*x2**3 - 79*x2**2 + 3*x2,
311 84*x2**4 - 40*x2**3 + x2**2 + x2], x0, x1, x2, domain='ZZ', order='lex')
312 Root isolation:
313 >>> f = 7*z**4 - 19*z**3 + 20*z**2 + 17*z + 20
314 >>> intervals(f, all=True, eps=0.001)
315 ([],
316 [((-425/1024 - 625*I/1024, -1485/3584 - 2185*I/3584), 1),
317 ((-425/1024 + 2185*I/3584, -1485/3584 + 625*I/1024), 1),
```

```

318 ((3175/1792 - 2605*I/1792, 1815/1024 - 10415*I/7168), 1),
319 ((3175/1792 + 10415*I/7168, 1815/1024 + 2605*I/1792), 1))

```

320 **3.4. Printers.** SymPy has a rich collection of expression printers for displaying  
 321 expressions to the user. By default, an interactive Python session will render the `str`  
 322 form of an expression, which has been used in all the examples in this paper so far.  
 323 The `str` form of an expression is valid Python and roughly matches what a user would  
 324 type to enter the expression.

```

325 >>> phi0 = Symbol('phi0')
326 >>> str(Integral(sqrt(phi0), phi0))
327 'Integral(sqrt(phi0), phi0)'

```

328 Expressions can be printed with 2D, monospace fonts via `pprint`. This uses  
 329 Unicode characters to render mathematical symbols such as integral signs, square  
 330 roots, and parentheses. Greek letters and subscripts in symbol names that have  
 331 Unicode code points associated with them are also rendered automatically.

```

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

```

```

332  ∫ √ φ₀ + 1 d(φ₀)

```

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

```

335 >>> pprint(Integral(sqrt(phi0 + 1), phi0), use_unicode=False)
336 /
337 |
338 | _____
339 | \ / phi0 + 1 d(phi0)
340 |
341 /

```

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

```

343 >>> print(latex(Integral(sqrt(phi0 + 1), phi0)))
344 \int \sqrt{\phi_{0} + 1}\, d\phi_{0}

```

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

350 Other printers such as MathML are also available. SymPy uses an extensible  
 351 printer subsystem for customizing the printing for any given printer, and for custom  
 352 objects to define their printing behavior for any printer. The code generation capabil-  
 353 ities of SymPy use this subsystem to convert expressions into code in various target  
 354 programming languages.

355 **3.5. Solvers.** SymPy has a module of equation solvers for symbolic equations.  
 356 There are two functions for solving algebraic equations in SymPy: `solve` and `solveset`.  
 357 `solveset` has several design changes with respect to the older `solve` function. This dis-  
 358 tinction is present in order to resolve the usability issues with the previous `solve` func-  
 359 tion API while maintaining backward compatibility with earlier versions of SymPy.  
 360 `solveset` only requires the necessary input information from the user. The function  
 361 signatures of `solve` and `solveset` are  
 362 `solve(f, *symbols, **flags)`

```

363 solveset(f, symbol, domain=S.Complexes)
364 The domain parameter is typically either S.Complexes (the default) or S.Reals, which
365 causes it to only return real solutions.
366 Additionally, solve has an inconsistent output API for various types of inputs. For
367 instance, depending on the input, sometimes it returns a Python list and sometimes it
368 returns a Python dictionary. On the other hand, the solveset has a canonical output
369 API. solveset always returns a SymPy set object.
370 Both functions implicitly assume that expressions are equal to 0. For instance,
371 solveset(x - 1, x) solves  $x - 1 = 0$  for  $x$ .
372 Single solution:
373 >>> solveset(x - 1, x)
374 {1}
375 Finite solution set, quadratic equation:
376 >>> solveset(x**2 - pi**2, x)
377 {-pi, pi}
378 No solution:
379 >>> solveset(1, x)
380 EmptySet()
381 Interval solution:
382 >>> solveset(x**2 - 3 > 0, x, domain=S.Reals)
383 (-oo, -sqrt(3)) U (sqrt(3), oo)
384 Infinitely many solutions:
385 >>> solveset(sin(x) - 1, x, domain=S.Reals)
386 ImageSet(Lambda(_n, 2*_n*pi + pi/2), Integers())
387 >>> solveset(x - x, x, domain=S.Reals)
388 (-oo, oo)
389 >>> solveset(x - x, x, domain=S.Complexes)
390 S.Complexes
391 Linear systems are solved with linsolve. Finite and infinite solution for deter-
392 mined, under determined, and over determined problems are supported.
393 >>> A = Matrix([[1, 2, 3], [4, 5, 6], [7, 8, 10]])
394 >>> b = Matrix([3, 6, 9])
395 >>> linsolve((A, b), x, y, z)
396 {(-1, 2, 0)}
397 >>> linsolve(Matrix([[1, 1, 1, 1], [1, 1, 2, 3]]), (x, y, z))
398 {(-y - 1, y, 2)}
399 solveset is under active development as a planned replacement for solve. There
400 are certain features which are implemented in solve that are not yet implemented in
401 solveset. Notably, these include nonlinear multivariate system and transcendental
402 equations. Below are examples of solve applied to these two cases.
403 Nonlinear (multivariate) system of equations (the intersection of a circle and a parabola):
404 >>> solve([x**2 + y**2 - 16, 4*x - y**2 + 6], x, y)
405 [(-2 + sqrt(14), -sqrt(-2 + 4*sqrt(14))),
406  (-2 + sqrt(14), sqrt(-2 + 4*sqrt(14))),
407  (-sqrt(14) - 2, -I*sqrt(2 + 4*sqrt(14))),
408  (-sqrt(14) - 2, I*sqrt(2 + 4*sqrt(14)))]
409 Transcendental equations:
410 >>> solve((x + log(x))**2 - 5*(x + log(x)) + 6, x)
411 [LambertW(exp(2)), LambertW(exp(3))]
412 >>> solve(x**3 + exp(x))

```

```
413 [-3*LambertW((-1)**(2/3)/3)]
```

414 **3.6. Matrices.** SymPy supports matrices with symbolic expressions as elements.■

```
415 >>> x, y = symbols('x y')
416 >>> A = Matrix(2, 2, [x, x + y, y, x])
417 >>> A
418 Matrix([
419 [x, x + y],
420 [y, x]])
```

421 All SymPy matrix types perform linear algebra including matrix addition, multi-  
422 plication, exponentiation, computing determinants, solving linear systems, and com-  
423 puting inverses using LU decomposition, LDL decomposition, Gauss-Jordan elimina-  
424 tion, Cholesky decomposition, Moore-Penrose pseudoinverse, and adjugate matrix.

425 All operations are computed symbolically. For instance, eigenvalues are computed  
426 by generating the characteristic polynomial using the Berkowitz algorithm and then  
427 solving it using polynomial routines. Diagonalizable matrices can be diagonalized first  
428 to compute the eigenvalues.

```
429 >>> A.eigenvals()
430 {x - sqrt(y*(x + y)): 1, x + sqrt(y*(x + y)): 1}
```

431 Internally these matrices store the elements as a list, making it a dense repre-  
432 sentation. For storing sparse matrices, the `SparseMatrix` class can be used. Sparse  
433 matrices store the elements in a dictionary of keys (DoK) format.

434 SymPy also supports matrices with symbolic dimension values. `MatrixSymbol`  
435 represents a matrix with dimensions  $m \times n$ , where  $m$  and  $n$  can be symbolic. Ma-  
436 trix addition and multiplication, scalar operations, matrix inverse, and transpose are  
437 stored symbolically as matrix expressions.

```
438 >>> m, n, p = symbols("m, n, p", integer=True)
439 >>> R = MatrixSymbol("R", m, n)
440 >>> S = MatrixSymbol("S", n, p)
441 >>> T = MatrixSymbol("T", m, p)
442 >>> U = R*S + 2*T
443 >>> U.shape
444 (m, p)
445 >>> U[0, 1]
446 2*T[0, 1] + Sum(R[0, _k]*S[_k, 1], (_k, 0, n - 1))
```

447 Block matrices are also implemented in SymPy. `BlockMatrix` elements can be  
448 any matrix expression which includes explicit matrices, matrix symbols, and block  
449 matrices. All functionalities of matrix expressions are also present in `BlockMatrix`.

```
450 >>> n, m, l = symbols('n m l')
451 >>> X = MatrixSymbol('X', n, n)
452 >>> Y = MatrixSymbol('Y', m, m)
453 >>> Z = MatrixSymbol('Z', n, m)
454 >>> B = BlockMatrix([[X, Z], [ZeroMatrix(m, n), Y]])
455 >>> B
456 Matrix([
457 [X, Z],
458 [0, Y]])
459 >>> B[0, 0]
460 X[0, 0]
461 >>> B.shape
```

(m + n, m + n)

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

**4. Numerics.** The `Float` class represents an arbitrary-precision binary floating-point by storing its value and precision (in bits). An operation between two `Float` inputs is rounded to the larger of the two precisions. Since Python floating-point literals automatically evaluate to `double` (53-bit) precision, strings should be used to input precise decimal values:

```
>>> Float(1.1)
1.1000000000000000
>>> Float(1.1, 30) # precision equivalent to 30 digits
1.100000000000000008881784197001
>>> Float("1.1", 30)
1.1000000000000000000000000000000000
```

The preferred way to evaluate an expression numerically is with the `evalf` method, which internally estimates the number of accurate bits of the floating-point approximation for each sub-expression, and adaptively increases the working precision until the estimated accuracy of the final result matches the sought number of decimal digits.

The internal error tracking does not provide rigorous error bounds (in the sense of interval arithmetic) and cannot be used to accurately track uncertainty in measurement data; the sole purpose is to mitigate loss of accuracy that typically occurs when converting symbolic expressions to numerical values. A well-known, notorious case is that of catastrophic cancellation. This is illustrated by the following example (the input 25 specifies that 25 digits are sought):

```
>>> cos(exp(-100)).evalf(25) - 1
0
>>> (cos(exp(-100)) - 1).evalf(25)
-6.919482633683687653243407e-88
```

The `evalf` method works with complex numbers and supports more complicated expressions, such as special functions, infinite series and integrals.

SymPy does not track the accuracy of approximate numbers outside of `evalf`. The familiar dangers of floating-point arithmetic apply [26], and symbolic expressions containing floating-point numbers should be treated with reasonable caution. This approach is similar to Maple and Maxima.

By contrast, Mathematica uses a form of significance arithmetic [46] for approximate numbers. This offers further protection against numerical errors, but leads to non-obvious semantics while still not being mathematically rigorous (for a critique of significance arithmetic, see Fateman [20]). SymPy's `evalf` internals are non-rigorous in the same sense, but have no bearing on the semantics of floating-point numbers in the rest of the system.

**4.1. The mpmath library.** The implementation of arbitrary-precision floating-point arithmetic is supplied by the `mpmath` library, which originally was developed as a SymPy module but subsequently has been moved to a standalone pure Python package. The basic datatypes in `mpmath` are `mpf` and `mpc`, which respectively act as multiprecision substitutes for Python's `float` and `complex`. The floating-point precision is controlled by a global context:

```
>>> import mpmath
```





occur (the perturbation size is automatically decreased until the result is detected to converge numerically).

Due to this generic approach, particular combinations of hypergeometric functions can be specified easily. The implementation of the Meijer G-function takes only a few dozen lines of code, yet covers the whole input domain in a robust way. The Meijer G-function instance  $G_{1,3}^{3,0}(0; \frac{1}{2}, -1, -\frac{3}{2}|x)$  is a good test case [51]; past versions of both Maple and Mathematica produced incorrect numerical values for large  $x > 0$ . Here, mpmath automatically removes the internal singularity and compensates for cancellations (amounting to 656 bits of precision when  $x = 10000$ ), giving correct values:

```
>>> mpmath.mp.dps = 15
>>> mpmath.meijerg([[], [0]], [[-0.5, -1, -1.5], []], 10000)
2.4392576907199564e-94
```

Equivalently, with SymPy's interface this function can be evaluated as:

```
>>> meijerg([[], [0]], [[-S(1)/2, -1, -S(3)/2], []], 10000).evalf()
2.43925769071996e-94
```

We highlight the generalized hypergeometric functions and the Meijer G-function, due to those functions' frequent appearance in closed forms for integrals and sums (see section 3.2). Via mpmath, SymPy has relatively good support for evaluating sums and integrals numerically, using two complementary approaches: direct numerical evaluation, or first computing a symbolic closed form involving special functions.

**5. Domain Specific Submodules.** SymPy includes several packages that allow users to solve domain specific problems. For example, a comprehensive physics package is included that is useful for solving problems in mechanics, optics, and quantum mechanics along with support for manipulating physical quantities with units.

## 5.1. Classical Mechanics.

**5.1.1. Vector Algebra.** The `sympy.physics.vector` package 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, cross products, etc. Both of these objects can 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 force and torques. The dyadics are essentially reference frame aware  $3 \times 3$  tensors. 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  $\frac{\pi}{3}$  rad, 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  $\frac{\pi}{2}$  rad.

```
>>> from sympy import pi
>>> from sympy.physics.vector import ReferenceFrame
>>> A = ReferenceFrame('A')
>>> B = ReferenceFrame('B')
>>> C = ReferenceFrame('C')
```

```

605 >>> B.orient(A, 'body', (pi, pi / 3, pi / 4), 'xzx')
606 >>> C.orient(B, 'axis', (pi / 2, B.x))
607 >>> v = 1 * A.x + 2 * B.z + 3 * C.y
608 >>> v
609 A.x + 2*B.z + 3*C.y
610 >>> v.express(A)
611 A.x + 5*sqrt(3)/2*A.y + 5/2*A.z

```

612 **5.1.2. Mechanics.** The `sympy.physics.mechanics` package utilizes the `sympy.`  
613 `physics.vector` package to populate time aware particle and rigid body objects to  
614 fully describe the kinematics and kinetics of a rigid multi-body system. These objects  
615 store all of the information needed to derive the ordinary differential or differential  
616 algebraic equations that govern the motion of the system, i.e., the equations of mo-  
617 tion. These equations of motion abide by Newton’s laws of motion and can handle any  
618 arbitrary kinematic constraints or complex loads. The package offers two automated  
619 methods for formulating the equations of motion based on Lagrangian Dynamics [32]  
620 and Kane’s Method [31]. Lastly, there are automated linearization routines for con-  
621 strained dynamical systems based on [41].

622 **5.2. Quantum Mechanics.** The `sympy.physics.quantum` package has extensive  
623 capabilities for performing symbolic quantum mechanics, using Python objects to rep-  
624 resent the different mathematical objects relevant in quantum theory [45]: states (bras  
625 and kets), operators (unitary, hermitian, etc.), and basis sets, as well as operations  
626 on these objects such as representations, tensor products, inner products, outer prod-  
627 ucts, commutators, anticommutators, etc. The base objects are designed in the most  
628 general way possible to enable any particular quantum system to be implemented  
629 by subclassing the base operators and defining the relevant class methods to provide  
630 system specific logic.

631 For example, symbolic quantum operators and states may be defined and perform  
632 a full range of operations with them:

```

633 >>> from sympy.physics.quantum import Commutator, Dagger, Operator
634 >>> from sympy.physics.quantum import Ket, qapply
635 >>> A = Operator('A')
636 >>> B = Operator('B')
637 >>> C = Operator('C')
638 >>> D = Operator('D')
639 >>> a = Ket('a')
640 >>> comm = Commutator(A, B)
641 >>> comm
642 [A,B]
643 >>> qapply(Dagger(comm*a)).doit()
644 -<a|*(Dagger(A)*Dagger(B) - Dagger(B)*Dagger(A))
645 Commutators can be expanded using common commutator identities:
646 >>> Commutator(C+B, A*D).expand(commutator=True)
647 -[A,B]*D - [A,C]*D + A*[B,D] + A*[C,D]

```

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

- 650 • Many of the exactly solvable quantum systems, including simple harmonic  
651 oscillator states and raising/lowering operators, infinite square well states,  
652 and 3D position and momentum operators and states.
- 653 • Second quantized formalism of non-relativistic many-body quantum mechan-

```

654     ics [22].
655     • Quantum angular momentum [53]. Spin operators and their eigenstates can
656       be represented in any basis and for any quantum numbers. A rotation opera-
657       tor representing the Wigner-D matrix, which may be defined symbolically or
658       numerically, is also implemented to rotate spin eigenstates. Functionality for
659       coupling and uncoupling of arbitrary spin eigenstates is provided, including
660       symbolic representations of Clebsch-Gordon coefficients and Wigner symbols.
661     • Quantum information and computing [36]. Multidimensional qubit states,
662       and a full set of one- and two-qubit gates are provided and can be represented
663       symbolically or as matrices/vectors. With these building blocks it is possible
664       to implement a number of basic quantum algorithms including the quantum
665       Fourier transform, quantum error correction, quantum teleportation, Grover's
666       algorithm, dense coding, etc.
667     Here are a few short examples of the quantum information and computing capa-
668     bilities in sympy.physics.quantum. Start with a simple 4 qubit state and flip one of
669     the qubits:
670     >>> from sympy.physics.quantum.qubit import Qubit
671     >>> q = Qubit('0101')
672     >>> q
673     |0101>
674     >>> q.flip(1)
675     |0111>
676     Qubit states can also be used in adjoint operations, tensor products, inner/outer
677     products:
678     >>> Dagger(q)
679     <0101|
680     >>> ip = Dagger(q)*q
681     >>> ip
682     <0101|0101>
683     >>> ip.doit()
684     1
685     Quantum gates (unitary operators) can be applied to transform these states and then
686     classical measurements can be performed on the results:
687     >>> from sympy.physics.quantum.qubit import Qubit, measure_all
688     >>> from sympy.physics.quantum.gate import H, X, Y, Z
689     >>> from sympy.physics.quantum.qapply import qapply
690     >>> c = H(0)*H(1)*Qubit('00')
691     >>> c
692     H(0)*H(1)*|00>
693     >>> q = qapply(c)
694     >>> measure_all(q)
695     [(|00>, 1/4), (|01>, 1/4), (|10>, 1/4), (|11>, 1/4)]
696     Lastly, the following example creates a 3-qubit quantum fourier transform, decom-
697     posing it into one- and two-qubit gates, and then generating a circuit plot for the
698     sequence of gates (see Figure 1).
699     >>> from sympy.physics.quantum.qft import QFT
700     >>> from sympy.physics.quantum.circuitplot import circuit_plot
701     >>> fourier = QFT(0,3).decompose()
702     >>> fourier
703     SWAP(0,2)*H(0)*C((0),S(1))*H(1)*C((0),T(2))*C((1),S(2))*H(2)

```



Fig. 1: The circuit diagram for a 3-qubit quantum fourier transform generated by SymPy.

```
>>> c = circuit_plot(fourier, nqubits=3)
```

**6. 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. It is written in the general purpose Python programming language. This allows SymPy to be used in a first-class way with other Python projects, including the scientific Python stack. SymPy is designed to be used in an extensible way and, unlike many other CASs, both as an end-user application and as a library.

SymPy expressions are immutable trees of Python objects. SymPy uses Python both as the internal language and the user language. This permits users to access to the same methods that the library implements in order to extend it for their needs. Additionally, SymPy has a powerful assumptions system for declaring and deducing mathematical properties on expressions.

SymPy has submodules for many areas of mathematics. This includes functions for simplifying expressions, performing common calculus operations, pretty printing expressions, solving equations, and representing symbolic matrices. Other included areas are discrete math, concrete math, plotting, geometry, statistics, polynomials, sets, series, vectors, combinatorics, group theory, code generation, tensors, Lie algebras, cryptography, and special functions. Additionally, SymPy contains submodules targeting certain specific domains, such as classical mechanics and quantum mechanics. This breadth of domains has been engendered by a strong and vibrant user community. Anecdotally, these users likely chose SymPy because of its ease of access.

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

Work is being done on an assumptions subsystem, distinct from the one discussed in section 2.3. The new system stores assumption predicates separate from objects, and uses a SAT solver to do inference.

**7. Acknowledgements.** 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.

The Google Summer of Code, an international annual program, in which Google awards stipends to all students who successfully complete a requested free and open-source software coding project during the summer.

## 8. References.

## REFERENCES

- [1] <https://github.com/sympy/sympy/blob/master/doc/src/modules/polys/ringseries.rst>.
- [2] <https://github.com/scolobb/sympy/tree/ct4-commutativity>.
- [3] <https://scolobb.wordpress.com/>.
- [4] <https://reference.wolfram.com/language/ref/Flat.html>.
- [5] <https://reference.wolfram.com/language/ref/Orderless.html>.
- [6] <https://reference.wolfram.com/language/ref/OneIdentity.html>.
- [7] <https://reference.wolfram.com/language/tutorial/FlatAndOrderlessFunctions.html>.
- [8] *The software engineering of the wolfram system*, 2016, <https://reference.wolfram.com/language/tutorial/TheSoftwareEngineeringOfTheWolframSystem.html>.
- [9] M. ABRAMOWITZ AND I. A. STEGUN, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, NY, USA, ninth printing ed., 1964, <http://www.math.ucla.edu/~cbm/aands/>.
- [10] W. W. ADAMS AND P. LOUSTAUNAU, *An introduction to Gröbner bases*, no. 3, American Mathematical Soc., 1994.
- [11] D. H. BAILEY, K. JEYABALAN, AND X. S. LI, *A comparison of three high-precision quadrature schemes*, *Experimental Mathematics*, 14 (2005), pp. 317–329.
- [12] C. M. BENDER AND S. A. ORSZAG, *Advanced Mathematical Methods for Scientists and Engineers*, Springer, 1st ed., October 1999.
- [13] N. BIGGS, E. K. LLOYD, AND R. J. WILSON, *Graph Theory, 1736-1936*, Oxford University Press, 1976.
- [14] R. P. BRENT AND P. ZIMMERMANN, *Modern Computer Arithmetic*, Cambridge University Press, version 0.5.1 ed.
- [15] M. BRONSTEIN, *Poor Man's Integrator*, <http://www-sop.inria.fr/cafe/Manuel.Bronstein/pmint>.
- [16] M. BRONSTEIN, *Symbolic Integration I: Transcendental Functions*, Springer-Verlag, New York, NY, USA, 2005.
- [17] J. CARETTE, *Understanding Expression Simplification*, in ISSAC '04: Proceedings of the 2004 International Symposium on Symbolic and Algebraic Computation, New York, NY, USA, 2004, ACM Press, pp. 72–79, <http://dx.doi.org/http://doi.acm.org/10.1145/1005285.1005298>.
- [18] D. CERVONE, *Mathjax: a platform for mathematics on the web*, *Notices of the AMS*, 59 (2012), pp. 312–316.
- [19] R. CIMRMAN, *SfePy - write your own FE application*, in Proceedings of the 6th European Conference on Python in Science (EuroSciPy 2013), P. de Buyl and N. Varoquaux, eds., 2014, pp. 65–70. <http://arxiv.org/abs/1404.6391>.
- [20] R. J. FATEMAN, *A review of Mathematica*, *Journal of Symbolic Computation*, 13 (1992), pp. 545–579, [http://dx.doi.org/DOI:10.1016/S0747-7171\(10\)80011-2](http://dx.doi.org/DOI:10.1016/S0747-7171(10)80011-2).
- [21] H. R. P. FERGUSON, D. H. BAILEY, AND S. ARNO, *Analysis of PSLQ, an integer relation finding algorithm*, *Mathematics of Computation*, 68 (1999), pp. 351–369.
- [22] A. FETTER AND J. WALECKA, *Quantum Theory of Many-Particle Systems*, Dover Publications, 2003.
- [23] H. FU, X. ZHONG, AND Z. ZENG, *Automated and Readable Simplification of Trigonometric Expressions*, *Mathematical and Computer Modelling*, 55 (2006), pp. 1169–1177.
- [24] Y. C. FUNG, *A first course in continuum mechanics*, Pearson, third edition ed., 1993.
- [25] G. GEDE, D. L. PETERSON, A. S. NANJANGUD, J. K. MOORE, AND M. HUBBARD, *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, American Society of Mechanical Engineers, 2013, pp. V07BT10A051–V07BT10A051.
- [26] D. GOLDBERG, *What every computer scientist should know about floating-point arithmetic*, *ACM Computing Surveys (CSUR)*, 23 (1991), pp. 5–48.
- [27] D. GRUNTZ, *On Computing Limits in a Symbolic Manipulation System*, PhD thesis, Swiss Federal Institute of Technology, Zürich, Switzerland, 1996.
- [28] D. GRUNTZ AND W. KOEPF, *Formal power series*, (1993).
- [29] C. V. HORSEN, *GMPY*. <https://pypi.python.org/pypi/gmpy2>, 2015.

- [30] J. D. HUNTER, *Matplotlib: A 2d graphics environment*, Computing In Science & Engineering, 9 (2007), pp. 90–95.
- [31] T. R. KANE AND D. A. LEVINSON, *Dynamics, Theory and Applications*, McGraw Hill, 1985.
- [32] J. LAGRANGE, *Mécanique analytique*, no. v. 1 in *Mécanique analytique*, Ve Courcier, 1811.
- [33] L. R. U. MANSSUR, R. PORTUGAL, AND B. F. SVAITER, *Group-theoretic approach for symbolic tensor manipulation*, Int. J. Mod. Phys. C, 13 (2002), <http://dx.doi.org/http://dx.doi.org/10.1142/S0129183102004571>.
- [34] J. MARTÍN-GARCÍA, *xact, efficient tensor computer algebra*, 2002-2016, <http://metric.iem.csic.es/Martin-Garcia/xAct/>.
- [35] M. MOSKEWICZ, C. MADIGAN, AND S. MALIK, *Method and system for efficient implementation of boolean satisfiability*, Aug. 26 2008, <http://www.google.co.in/patents/US7418369>. US Patent 7,418,369.
- [36] M. NIELSEN AND I. CHUANG, *Quantum Computation and Quantum Information*, Cambridge University Press, 2011.
- [37] A. NIJENHUIS AND H. S. WILF, *Combinatorial Algorithms: For Computers and Calculators*, Academic Press, New York, NY, USA, second ed., 1978.
- [38] T. E. OLIPHANT, *Python for scientific computing*, Computing in Science & Engineering, 9 (2007), pp. 10–20.
- [39] K. PEETERS, *Cadabra: a field-theory motivated symbolic computer algebra system*, Computer Physics Communications, (2007).
- [40] F. PÉREZ AND B. E. GRANGER, *Ipython: a system for interactive scientific computing*, Computing in Science & Engineering, 9 (2007), pp. 21–29.
- [41] D. L. PETERSON, G. GEDE, AND M. HUBBARD, *Symbolic linearization of equations of motion of constrained multibody systems*, Multibody System Dynamics, 33 (2014), pp. 143–161, <http://dx.doi.org/10.1007/s11044-014-9436-5>.
- [42] M. PETKOVŠEK, H. S. WILF, AND D. ZEILBERGER, *A = bak peters*, Wellesley, MA, (1996).
- [43] E. RAYMOND, *The cathedral and the bazaar*, Knowledge, Technology & Policy, 12 (1999), pp. 23–49.
- [44] M. ROCKLIN AND A. R. TERREL, *Symbolic statistics with sympy*, Computing in Science and Engineering, 14 (2012), <http://dx.doi.org/10.1109/MCSE.2012.56>.
- [45] J. SAKURAI AND J. NAPOLITANO, *Modern Quantum Mechanics*, Addison-Wesley, 2010.
- [46] M. SOFRONIOU AND G. SPALETTA, *Precise numerical computation*, Journal of Logic and Algebraic Programming, 64 (2005), pp. 113–134.
- [47] P. SOLIN, K. SEGETH, AND I. DOLEZEL, *Higher-Order Finite Element Methods*, Chapman & Hall / CRC Press, 2003.
- [48] W. STEIN AND D. JOYNER, *SAGE: System for Algebra and Geometry Experimentation*, Communications in Computer Algebra, 39 (2005).
- [49] G. J. SUSSMAN AND J. WISDOM, *Functional Differential Geometry*, Massachusetts Institute of Technology Press, 2013.
- [50] H. TAKAHASI AND M. MORI, *Double exponential formulas for numerical integration*, Publications of the Research Institute for Mathematical Sciences, 9 (1974), pp. 721–741.
- [51] V. T. TOTH, *Maple and meijer’s g-function: a numerical instability and a cure*. <http://www.vttoth.com/CMS/index.php/technical-notes/67>, 2007.
- [52] M. J. TURK, B. D. SMITH, J. S. OISHI, S. SKORY, S. W. SKILLMAN, T. ABEL, AND M. L. NORMAN, *yt: A Multi-code Analysis Toolkit for Astrophysical Simulation Data*, The Astrophysical Journal Supplement Series, 192 (2011), pp. 9–, <http://dx.doi.org/10.1088/0067-0049/192/1/9>, [arXiv:1011.3514](https://arxiv.org/abs/1011.3514).
- [53] R. ZARE, *Angular Momentum: Understanding Spatial Aspects in Chemistry and Physics*, Wiley, 1991.
- [54] O. ZIENKIEWICZ, R. TAYLOR, AND J. ZHU, *The Finite Element Method: Its Basis and Fundamentals*, Butterworth-Heinemann, seventh edition ed., 2013, <http://dx.doi.org/http://dx.doi.org/10.1016/B978-1-85617-633-0.00019-8>.

## 9. Supplement.

**9.1. Limits: The Gruntz Algorithm.** SymPy calculates limits using the Gruntz algorithm, as described in [27]. The basic idea is as follows: any limit can be converted to a limit  $\lim_{x \rightarrow \infty} f(x)$  by substitutions like  $x \rightarrow \frac{1}{x}$ . Then the most varying subexpression  $\omega$  (that converges to zero as  $x \rightarrow \infty$  the fastest from all subexpressions) is identified in  $f(x)$ , and  $f(x)$  is expanded into a series with respect to  $\omega$ . Any positive powers of  $\omega$  converge to zero. If there are negative powers of  $\omega$ , then the

limit is infinite. The constant term (independent of  $\omega$ , but could depend on  $x$ ) then determines the limit (one might need to recursively apply the Gruntz algorithm on this term to determine the limit).

To determine the most varying subexpression, the comparability classes must first be defined, by calculating  $L$ :

$$(1) \quad L \equiv \lim_{x \rightarrow \infty} \frac{\log |f(x)|}{\log |g(x)|}$$

And then operations  $<$ ,  $>$  and  $\sim$  are defined as follows:  $f > g$  when  $L = \pm\infty$  (it is said that  $f$  is more rapidly varying than  $g$ , i.e.,  $f$  goes to  $\infty$  or  $0$  faster than  $g$ ,  $f$  is greater than any power of  $g$ ),  $f < g$  when  $L = 0$  ( $f$  is less rapidly varying than  $g$ ) and  $f \sim g$  when  $L \neq 0, \pm\infty$  (both  $f$  and  $g$  are bounded from above and below by suitable integral powers of the other). Here are some examples of comparability classes:

$$\begin{aligned} 2 &< x < e^x < e^{x^2} < e^{e^x} \\ 2 &\sim 3 \sim -5 \\ x &\sim x^2 \sim x^3 \sim \frac{1}{x} \sim x^m \sim -x \\ e^x &\sim e^{-x} \sim e^{2x} \sim e^{x+e^{-x}} \\ f(x) &\sim \frac{1}{f(x)} \end{aligned}$$

The Gruntz algorithm is now illustrated on the following example:

$$(2) \quad f(x) = e^{x+2e^{-x}} - e^x + \frac{1}{x}.$$

The goal is to calculate  $\lim_{x \rightarrow \infty} f(x)$ . First the set of most rapidly varying subexpressions is determined, the so called *mrsv set*. For (2), the following mrsv set  $\{e^x, e^{-x}, e^{x+2e^{-x}}\}$  is obtained. These are all subexpressions of (2) and they all belong to the same comparability class. This calculation can be done using SymPy as follows:

```
>>> from sympy.series.gruntz import mrsv
>>> mrsv(exp(x+2*exp(-x))-exp(x) + 1/x, x)[0].keys()
dict_keys([exp(x + 2*exp(-x)), exp(x), exp(-x)])
```

Next any item  $\omega$  is taken from mrsv that converges to zero for  $x \rightarrow \infty$ . The item  $\omega = e^{-x}$  is obtained. If such a term is not present in the mrsv set (i.e., all terms converge to infinity instead of zero), the relation  $f(x) \sim \frac{1}{f(x)}$  can be used.

Next step is to rewrite the mrsv in terms of  $\omega$ :  $\{\frac{1}{\omega}, \omega, \frac{1}{\omega}e^{2\omega}\}$ . Then the original subexpressions are substituted back into  $f(x)$  and expanded with respect to  $\omega$ :

$$(3) \quad f(x) = \frac{1}{x} - \frac{1}{\omega} + \frac{1}{\omega}e^{2\omega} = 2 + \frac{1}{x} + 2\omega + O(\omega^2)$$

Since  $\omega$  is from the mrsv set, then in the limit  $x \rightarrow \infty$  it is  $\omega \rightarrow 0$  and so  $2\omega + O(\omega^2) \rightarrow 0$  in (3):

$$(4) \quad f(x) = \frac{1}{x} - \frac{1}{\omega} + \frac{1}{\omega}e^{2\omega} = 2 + \frac{1}{x} + 2\omega + O(\omega^2) \rightarrow 2 + \frac{1}{x}$$

Since the result  $(2 + \frac{1}{x})$  still depends on  $x$ , the above procedure is iterated on the result until just a number (independent of  $x$ ) is obtained, which is the final limit. In the above case the limit is 2, as can be verified by SymPy:



```

881 >>> limit(exp(x+2*exp(-x))-exp(x) + 1/x, x, oo)
882 2

```

883 In general, when  $f(x)$  is expanded in terms of  $\omega$ , it is obtained:

$$884 \quad (5) \quad f(x) = \underbrace{O\left(\frac{1}{\omega^3}\right)}_{\infty} + \underbrace{\frac{C_{-2}(x)}{\omega^2}}_{\infty} + \underbrace{\frac{C_{-1}(x)}{\omega}}_{\infty} + C_0(x) + \underbrace{C_1(x)\omega}_0 + \underbrace{O(\omega^2)}_0$$

885 The positive powers of  $\omega$  are zero. If there are any negative powers of  $\omega$ , then the  
886 result of the limit is infinity, otherwise the limit is equal to  $\lim_{x \rightarrow \infty} C_0(x)$ . The expression  
887  $C_0(x)$  is simpler than  $f(x)$  and so the algorithm always converges. A proof of this, as  
888 well as further details are given in Gruntz's Ph.D. thesis [27].

## 889 9.2. Series.

890 **9.2.1. Series Expansion.** SymPy is able to calculate the symbolic series expansion  
891 of an arbitrary series or expression involving elementary and special functions and  
892 multiple variables. For this it has two different implementations- the `series` method  
893 and Ring Series.

894 The first approach stores a series as an object of the `Basic` class. Each function  
895 has its specific implementation of its expansion which is able to evaluate the Puiseux  
896 series expansion about a specified point. For example, consider a Taylor expansion  
897 about 0:

```

898 >>> from sympy import symbols, series
899 >>> x, y = symbols('x, y')
900 >>> series(sin(x+y) + cos(x*y), x, 0, 2)
901 1 + sin(y) + x*cos(y) + 0(x**2)

```

902 The newer and much faster[1] approach called Ring Series makes use of the ob-  
903 servation that a truncated Taylor series, is in fact a polynomial. Ring Series uses the  
904 efficient representation and operations of sparse polynomials. The choice of sparse  
905 polynomials is deliberate as it performs well in a wider range of cases than a dense  
906 representation. Ring Series gives the user the freedom to choose the type of coeffi-  
907 cients he wants to have in his series, allowing the use of faster operations on certain  
908 types.

909 For this, several low level methods for expansion of trigonometric, hyperbolic and  
910 other elementary functions like inverse of a series, calculating  $n$ th root, etc, are imple-  
911 mented using variants of the Newton Method [14]. All these support Puiseux series  
912 expansion. The following example demonstrates the use of an elementary function  
913 that calculates the Taylor expansion of the sine of a series.

```

914 >>> from sympy import ring
915 >>> from sympy.polys.ring_series import rs_sin
916 >>> R, t = ring('t', QQ)
917 >>> rs_sin(t**2 + t, t, 5)
918 -1/2*t**4 - 1/6*t**3 + t**2 + t

```

919 The function `sympy.polys.rs_series` makes use of these elementary functions to  
920 expand an arbitrary SymPy expression. It does so by following a recursive strategy  
921 of expanding the lower most functions first and then composing them recursively to  
922 calculate the desired expansion. Currently, it only supports expansion about 0 and  
923 is under active development. Ring Series is several times faster than the default  
924 implementation with the speed difference increasing with the size of the series. The  
925 `sympy.polys.rs_series` takes as input any SymPy expression and hence there is no

```

926 need to explicitly create a polynomial ring. An example:
927 >>> from sympy.polys.ring_series import rs_series
928 >>> from sympy.abc import a, b
929 >>> from sympy import sin, cos
930 >>> rs_series(sin(a + b), a, 4)
931 -1/2*(sin(b))*a**2 + (sin(b)) - 1/6*a**3*(cos(b)) + a*(cos(b))

```

**9.2.2. Formal Power Series.** SymPy can be used for computing the Formal Power Series of a function. The implementation is based on the algorithm described in the paper on Formal Power Series [28]. The advantage of this approach is that an explicit formula for the coefficients of the series expansion is generated rather than just computing a few terms.

The following example shows how to use `fps`:

```

937 >>> f = fps(sin(x), x, x0=0)
938 >>> f.truncate(6)
939 x - x**3/6 + x**5/120 + O(x**6)
940 >>> f[15]
941 -x**15/1307674368000

```

**9.2.3. Fourier Series.** SymPy provides functionality to compute Fourier series of a function using the `fourier_series` function. Under the hood, this function computes  $a_0$ ,  $a_n$ ,  $b_n$  coefficients using standard integration formulas.

Here's an example on how to compute Fourier series in SymPy:

```

947 >>> L = symbols('L')
948 >>> expr = 2 * (Heaviside(x/L) - Heaviside(x/L - 1)) - 1
949 >>> f = fourier_series(expr, (x, 0, 2*L))
950 >>> f.truncate(3)
951 4*sin(pi*x/L)/pi + 4*sin(3*pi*x/L)/(3*pi) + 4*sin(5*pi*x/L)/(5*pi)

```

**9.3. Logic.** SymPy supports construction and manipulation of boolean expressions through the `logic` module. SymPy symbols can be used as propositional variables and also be substituted as `True` or `False`. A good number of manipulation features for boolean expressions have been implemented in the `logic` module.

**9.3.1. Constructing boolean expressions.** A boolean variable can be declared as a SymPy symbol. Python operators `&`, `|` and `~` are overridden when using SymPy objects to use the SymPy functionality for logical `And`, `Or`, and `negate`. Other logic functions are also integrated into SymPy, including `Xor` and `Implies`, which are constructed with `^` and `>>`, respectively. The above are just a shorthand, expressions can also be constructed by directly creating the relevant objects: `And()`, `Or()`, `Not()`, `Xor()`, `Nand()`, `Nor()`, etc.

```

963 >>> from sympy import *
964 >>> x, y, z = symbols('x y z')
965 >>> e = (x & y) | z
966 >>> e.subs({x: True, y: True, z: False})
967 True

```

**9.3.2. CNF and DNF.** Any boolean expression can be converted to conjunctive normal form, disjunctive normal form, and negation normal form. The API also exposes methods to check if a boolean expression is in any of the above mentioned forms.

```

972 >>> from sympy.logic.boolalg import is_dnf, is_cnf

```

```

973 >>> x, y, z = symbols('x y z')
974 >>> to_cnf((x & y) | z)
975 And(Or(x, z), Or(y, z))
976 >>> to_dnf(x & (y | z))
977 Or(And(x, y), And(x, z))
978 >>> is_cnf((x | y) & z)
979 True
980 >>> is_dnf((x & y) | z)
981 True

```

982 **9.3.3. Simplification and Equivalence.** The module supports simplification  
983 of given boolean expression by making deductions from the expression. Equivalence  
984 of two logical expressions can also be checked. In the case of equivalence, it is possible  
985 to return the mapping of variables in two expressions so as to represent the same  
986 logical behaviour.

```

987 >>> from sympy import *
988 >>> a, b, c, x, y, z = symbols('a b c x y z')
989 >>> e = a & (~a | ~b) & (a | c)
990 >>> simplify(e)
991 And(Not(b), a)
992 >>> e1 = a & (b | c)
993 >>> e2 = (x & y) | (x & z)
994 >>> bool_map(e1, e2)
995 (And(Or(b, c), a), {a: x, b: y, c: z})

```

996 **9.3.4. SAT solving.** The module also supports satisfiability (SAT) checking of  
997 a given boolean expression. If satisfiable, it is possible to return a model for which the  
998 expression is satisfiable. The API also supports returning all possible models. The  
999 SAT solver has a clause learning DPLL algorithm implemented with a watch literal  
1000 scheme and VSIDS heuristic[35].

```

1001 >>> from sympy import *
1002 >>> a, b, c = symbols('a b c')
1003 >>> satisfiable(a & (~a | b) & (~b | c) & ~c)
1004 False
1005 >>> satisfiable(a & (~a | b) & (~b | c) & c)
1006 {a: True, b: True, c: True}

```

1007 **9.4. Diophantine Equations.** Diophantine equations play a central and an im-  
1008 portant role in number theory. A Diophantine equation has the form,  $f(x_1, x_2, \dots, x_n) =$   
1009 0 where  $n \geq 2$  and  $x_1, x_2, \dots, x_n$  are integer variables. If we can find  $n$  integers  
1010  $a_1, a_2, \dots, a_n$  such that  $x_1 = a_1, x_2 = a_2, \dots, x_n = a_n$  satisfies the above equation, we  
1011 say that the equation is solvable.

1012 Currently, the following five types of Diophantine equations can be solved using  
1013 SymPy's Diophantine module.

- 1014 • Linear Diophantine equations:  $a_1x_1 + a_2x_2 + \dots + a_nx_n = b$
- 1015 • General binary quadratic equation:  $ax^2 + bxy + cy^2 + dx + ey + f = 0$
- 1016 • Homogeneous ternary quadratic equation:  $ax^2 + by^2 + cz^2 + dxy + eyz + fzx = 0$
- 1017 • Extended Pythagorean equation:  $a_1x_1^2 + a_2x_2^2 + \dots + a_nx_n^2 = a_{n+1}x_{n+1}^2$
- 1018 • General sum of squares:  $x_1^2 + x_2^2 + \dots + x_n^2 = k$

1019 When an equation is fed into Diophantine module, it factors the equation (if  
1020 possible) and solves each factor separately. Then, all the results are combined to

```

1021 create the final solution set. The following examples illustrate some of the basic
1022 functionalities of the Diophantine module.
1023 >>> from sympy import symbols
1024 >>> x, y, z = symbols("x, y, z", integer=True)
1025
1026 >>> from sympy.solvers.diophantine import *
1027 >>> diophantine(2*x + 3*y - 5)
1028 set([(3*t_0 - 5, -2*t_0 + 5)])
1029
1030 >>> diophantine(2*x + 4*y - 3)
1031 set()
1032
1033 >>> diophantine(x**2 - 4*x*y + 8*y**2 - 3*x + 7*y - 5)
1034 set([(2, 1), (5, 1)])
1035
1036 >>> diophantine(x**2 - 4*x*y + 4*y**2 - 3*x + 7*y - 5)
1037 set([(-2*t**2 - 7*t + 10, -t**2 - 3*t + 5)])
1038
1039 >>> diophantine(3*x**2 + 4*y**2 - 5*z**2 + 4*x*y - 7*y*z + 7*z*x)
1040 set([(-16*p**2 + 28*p*q + 20*q**2,
1041 3*p**2 + 38*p*q - 25*q**2,
1042 4*p**2 - 24*p*q + 68*q**2)])
1043
1044 >>> from sympy.abc import a, b, c, d, e, f
1045 >>> diophantine(9*a**2 + 16*b**2 + c**2 + 49*d**2 + 4*e**2 - 25*f**2)
1046 set([(70*t1**2 + 70*t2**2 + 70*t3**2 + 70*t4**2 - 70*t5**2, 105*t1*t5,
1047 420*t2*t5, 60*t3*t5, 210*t4*t5,
1048 42*t1**2 + 42*t2**2 + 42*t3**2 + 42*t4**2 + 42*t5**2)])
1049
1050 >>> diophantine(a**2 + b**2 + c**2 + d**2 + e**2 + f**2 - 112)
1051 set([(8, 4, 4, 4, 0, 0)])

```

1052 **9.5. Sets.** SymPy supports representation of a wide variety of mathematical  
1053 sets. This is achieved by first defining abstract representations of atomic set classes  
1054 and then combining and transforming them using various set operations.

1055 Each of the set classes inherits from the base class `Set` and defines methods to  
1056 check membership and calculate unions, intersections, and set differences. When these  
1057 methods are not able to evaluate to atomic set classes, they are represented as abstract  
1058 unevaluated objects.

1059 SymPy has the following atomic set classes:

- 1060 • `EmptySet` represents the empty set  $\emptyset$ .
- 1061 • `UniversalSet` is an abstract “universal set” for which everything is a member.  
1062 The union of the universal set with any set gives the universal set and the  
1063 intersection gives the other set itself.
- 1064 • `FiniteSet` is functionally equivalent to Python’s built in `set` object. Its mem-  
1065 bers can be any SymPy object including other sets.
- 1066 • `Integers` represents the set of integers  $\mathbb{Z}$ .
- 1067 • `Naturals` represents the set of natural numbers  $\mathbb{N}$ , i.e., the set of positive  
1068 integers.
- 1069 • `Naturals0` represents the set of whole numbers  $\mathbb{N}^0$ , which are all the non-

negative integers.

- **Range** represents a range of integers. A range is defined by specifying a start value, an end value, and a step size. The enumeration of a **Range** object is functionally equivalent to Python's **range** except it supports infinite endpoints, allowing the representation of infinite ranges.
- **Interval** represents an interval of real numbers. It is specified by giving the start and end point and specifying if it is open or closed in the respective ends.

Other than unevaluated classes of **Union**, **Intersection**, and **Complement** operations, we have following set classes.

- **ProductSet** defines the Cartesian product of two or more sets. The product set is useful when representing higher dimensional spaces. For example, to represent a three-dimensional space, we simply take the Cartesian product of three real sets.
- **ImageSet** represents the image of a function when applied to a particular set. The image set of a function  $F$  with respect to a set  $S$  is  $\{F(x)|x \in S\}$ . SymPy uses image sets to represent sets of infinite solutions equations such as  $\sin(x) = 0$ .
- **ConditionSet** represents a subset of a set whose members satisfies a particular condition. The condition set of the set  $S$  with respect to the condition  $H$  is  $\{x|H(x), x \in S\}$ . SymPy uses condition sets to represent the set of solutions of equations and inequalities, where the equation or the inequality is the condition and the set is the domain being solved over.

A few other classes are implemented as special cases of the classes described above. The set of real numbers, **Reals**, is implemented as a special case of **Interval** over the interval  $(-\infty, \infty)$ . **ComplexRegion** is implemented as a special case of **ImageSet**. **ComplexRegion** supports both polar and rectangular representation of regions on the complex plane.

**9.6. Category Theory.** SymPy includes a basic version of the module for dealing with categories — abstract mathematical objects representing classes of structures as classes of objects (points) and morphisms (arrows) between the objects. This version of the module was designed with the following two goals in mind:

1. automatic typesetting of diagrams given by a collection of objects and of morphisms between them, and
2. specification and (semi-)automatic derivation of properties using commutative diagrams.

At of version 1.0, SymPy only implements the first goal, while a (very partially working) draft of implementation of the second goal is available at [2].

In order to achieve the two goals, the module **categories** defines several classes representing some of the essential concepts: objects, morphisms, categories, and diagrams. In category theory, the inner structure of objects is often discarded in the favour of studying the properties of morphisms, so the class **Object** is essentially a synonym of the class **Symbol**. There are several morphism classes which do not have a particular internal structure either, though an exception is **CompositeMorphism**, which essentially stores a list of morphisms.

To capture the properties of morphisms, the class **Diagram** is expected to be used. This class stores a family of morphisms, the corresponding source and target objects, and, possibly, some properties of the morphisms. Generally, no restrictions are imposed on what the properties may be — for example, one might use strings of

the form “forall”, “exists”, “unique”, etc. Furthermore, the morphisms of a diagram are grouped into *premises* and *conclusions*, in order to be able to represent logical implications of the form “for a collection of morphisms  $P$  with properties  $p : P \rightarrow \Omega$  (the premises), there exists a collection of morphisms  $C$  with properties  $c : C \rightarrow \Omega$  (the conclusions),” where  $\Omega$  is the universal collection of properties. Finally, the class **Category** includes a collection of diagrams which are deemed commutative and which therefore define the properties of this category.

Automatic typesetting of diagrams takes a **Diagram** and produces L<sup>A</sup>T<sub>E</sub>X code using the Xy-pic package. Typesetting is done in two stages: layout and generation of Xy-pic code. The layout stage is taken care of by the class **DiagramGrid**, which takes a **Diagram** and lays out the objects in a grid, trying to reduce the average length of the arrows in the final picture. By default, **DiagramGrid** uses a series of triangle-based heuristics to produce a rectangular grid. A linear layout can also be imposed. Furthermore, groups of objects can be given; in this case, the groups will be treated as atomic cells, and the member objects will be typeset independently of the other objects.

The second phase of diagram typesetting consists of actually drawing the picture and is carried out by the class **XypicDiagramDrawer**. An example of a diagram automatically typeset by **DiagramGrid** and **XypicDiagramDrawer** is given in Figure 2.

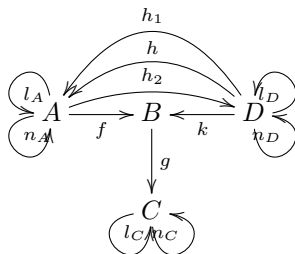


Fig. 2: An automatically typeset commutative diagram

As far as the second main goal of the module is concerned, a (non-working) draft of an implementation is at [2]. The principal idea consists of automatically deciding whether a diagram is commutative or not, given a collection of “axioms” — diagrams *known* to be commutative. The implementation is based on graph embeddings (injective maps): whenever an embedding of a commutative diagram into a given diagram is found, one concludes that the subdiagram is commutative. Deciding commutativity of the whole diagram is therefore based (theoretically) on finding a “cover” of the target diagram by embeddings of the axioms. The naïve implementation proved to be prohibitively slow; a better optimised version is therefore in order, as well as application of heuristics.

Contributions to automatic inference of commutativity of diagrams are welcome. The source code (both the one in master and in ct4-commutativity) is extensively documented. Even more extensive explanations (including some literary chatter) are given at [3].

**9.7. SymPy Gamma.** SymPy Gamma is a simple web application that runs on Google App Engine. It executes and displays the results of SymPy expressions as well as additional related computations, in a fashion similar to that of Wolfram|Alpha.

For instance, entering an integer will display its prime factors, digits in the base-10 expansion, and a factorization diagram. Entering a function will display its docstring; in general, entering an arbitrary expression will display its derivative, integral, series expansion, plot, and roots.

SymPy Gamma also has several additional features than just computing the results using SymPy.

- SymPy Gamma displays integration and differentiation steps in detail, which can be viewed in Figure 3:

Integral Steps:

integrate(tan(x), x)

Fullscreen

1. Rewrite the integrand:

$$\tan(x) = \frac{\sin(x)}{\cos(x)}$$

2. Let  $u = \cos(x)$ .

Then let  $du = -\sin(x)dx$  and substitute  $du$ :

$$\int -\frac{1}{u} du$$

A. The integral of a constant times a function is the constant times the integral of the function:

$$\int -\frac{1}{u} du = - \int \frac{1}{u} du$$

I. The integral of  $\frac{1}{u}$  is  $\log(u)$ .

So, the result is:  $-\log(u)$

Now substitute  $u$  back in:

$$-\log(\cos(x))$$

3. Add the constant of integration:

$$-\log(\cos(x)) + \text{constant}$$

The answer is:

$$-\log(\cos(x)) + \text{constant}$$

Fig. 3: Integral steps of  $\tan(x)$

- SymPy Gamma displays the factor tree diagrams for different numbers.
- SymPy Gamma saves user search queries, and offers many such similar features for free, which Wolfram|Alpha only offers to its paid users.

Every input query from the user on SymPy Gamma is first parsed by its own parser, which handles several different forms of function names, which SymPy as a library does not support. For instance, SymPy Gamma supports queries like `sin x`, whereas SymPy does not support this, and supports only `sin(x)`.

This parser converts the input query to the equivalent SymPy readable code, which is then eventually processed by SymPy, and the result is finally printed with the built-in LaTeX output and rendered on the SymPy Gamma web-application.



**9.8. SymPy Live.** SymPy Live is an online Python shell, which runs on Google App Engine, that executes SymPy code. It is integrated in the SymPy documentation examples, located at this [link](#).

This is accomplished by providing a HTML/JavaScript GUI for entering source code and visualization of output, and a server that evaluates the requested source code. It is an interactive AJAX shell that runs SymPy code using Python on the server.

Certain Features of SymPy Live:

- It supports the exact same syntax as SymPy, hence it can be used easily to test for outputs from various SymPy expressions.
- It can be run as a standalone app or in an existing app as an admin-only handler, and can also be used for system administration tasks, as an interactive way to try out APIs, or as a debugging aid during development.
- It can also be used to plot figures ([link](#)), and execute all kinds of expressions that SymPy can evaluate.
- SymPy Live also renders the output in LaTeX for pretty-printing the output.

**9.9. Comparison with Mathematica.** Wolfram Mathematica is a popular proprietary CAS. It features highly advanced algorithms. Mathematica has a core implemented in C++ [8] which interprets its own programming language (known as Wolfram language).

Analogous to Lisp's S-expressions, Mathematica uses its own style of M-expressions, which are arrays of either atoms or other M-expression. The first element of the expression identifies the type of the expression and is indexed by zero, whereas the first argument is indexed by one. Notice that SymPy expression arguments are stored in a Python tuple (that is, an immutable array), while the expression type is identified by the type of the object storing the expression.

Mathematica can associate attributes to its atoms. Attributes may define mathematical properties and behavior of the nodes associated to the atom. In SymPy, the usage of static class fields is roughly similar to Mathematica's attributes, though other programming patterns may also be used to achieve an equivalent behavior, such as class inheritance.

Unlike SymPy, Mathematica's expressions are mutable, that is one can change parts of the expression tree without the need of creating a new object. The mutability of Mathematica allows for a lazy updating of any references to that data structure.

Products in Mathematica are determined by some builtin node types, such as `Times`, `Dot`, and others. `Times` is a representation of the `*` operator, and is always meant to represent a commutative product operator. The other notable product is `Dot`, which represents the `.` operator. This product represents matrix multiplication, it is not commutative. In general, SymPy uses the same node for both scalar and matrix multiplication, the only exception being with abstract matrix symbols. Unlike Mathematica, SymPy determines commutativity with respect to multiplication from the factor's expression type. Mathematica puts the `Orderless` attribute on the expression type.

Regarding associative expressions, SymPy handles associativity by making associative expressions inherit the class `AssocOp`, while Mathematica specifies the `Flat` [4] attribute on the expression type.

Mathematica relies heavily on pattern matching — even the so-called equivalent of function declaration is in reality the definition of a pattern matching generating an expression tree transformation on input expressions. Mathematica's pattern matching

is sensitive to associative [4], commutative [5], and one-identity [6] properties of its expression tree nodes [7]. SymPy has various ways to perform pattern matching. All of them play a lesser role in the CAS than in Mathematica and are basically available as a tool to rewrite expressions. The differential equation solver in SymPy somewhat relies on pattern matching to identify the kind of differential equation, but it is envisaged to replace that strategy with analysis of Lie symmetries in the future. Mathematica's real advantage is the ability to add new overloading to the expression builder at runtime, or for specific subnodes. Consider for example:

```
In[1]:= Unprotect[Plus]
Out[1]= {Plus}

In[2]:= Sin[x_]^2 + Cos[y_]^2 := 1

In[3]:= x + Sin[t]^2 + y + Cos[t]^2
Out[3]= 1 + x + y
```

This expression in Mathematica defines a substitution rule that overloads the functionality of the `Plus` node (the node for additions in Mathematica). The trailing underscore after a symbol means that it is to be considered a wildcard. This example may not be practical, one may wish to keep this identity unevaluated. Nevertheless, it clearly illustrates the potential to define one's own immediate transformation rules. In SymPy, the operations constructing the addition node in the expression tree are Python class constructors and cannot be modified at runtime.<sup>4</sup> The way SymPy deals with extending the missing runtime overloadability functionality is by subclassing the node types. Subclasses may redefine the class constructor to yield the proper extended functionality.

Unlike SymPy, Mathematica does not support type inheritance or polymorphism [20]. SymPy relies heavily on class inheritance, but for the most part, class inheritance is used to make sure that SymPy objects inherit the proper methods and implement the basic hashing system. Associativity of expressions can be achieved by inheriting the class `AssocOp`, which may appear a more cumbersome operation than Mathematica's attribute setting.

Matrices in SymPy are types on their own. In Mathematica, nested lists are interpreted as matrices whenever the sublists have the same length. The main difference to SymPy is that ordinary operators and functions do not get generalized the same way as used in traditional mathematics. Using the standard multiplication in Mathematica performs an elementwise product, this is compatible with Mathematica's convention of commutativity of `Times` nodes. Matrix product is expressed by the `dot` operator, or the `Dot` node. The same is true for the other operators, and even functions, most notably calling the exponential function `Exp` on a matrix returns an elementwise exponentiation of its elements. The real matrix exponentiation is available through the `MatrixExp` function.

Unevaluated expressions in Mathematica can be achieved in various ways, most commonly with the `HoldForm` or `Hold` nodes, that block the evaluation of subnodes by the parser. Note that such a node cannot be expressed in Python, because of greedy evaluation. Whenever needed in SymPy, it is necessary to add the parameter

---

<sup>4</sup>In reality, Python supports monkey patching, nonetheless, it is a discouraged programming pattern.

1272 `evaluate=False` to all subnodes, or put the input expression in a string.

1273 In Mathematica, the operator `==` returns a boolean whenever it is able to immediately evaluate the truth of the equality, otherwise it returns an `Equal` expression.  
1274 In SymPy, `==` means structural equality and is always guaranteed to return a boolean expression. To express an equality in SymPy it is necessary to explicitly construct an object of the `Equality` class.

1275 SymPy, in accordance with Python and unlike the usual programming convention,  
1276 uses `**` to express the power operator, while Mathematica uses the more common `^`.

1280 **9.10. Other Projects that use SymPy.** There are several projects that use  
1281 SymPy as a library for implementing a part of their project, or even as a part of  
1282 back-end for their application as well.

1283 Some of them are listed below:

- 1284 • **Cadabra**: Cadabra is a symbolic computer algebra system (CAS) designed  
1285 specifically for the solution of problems encountered in field theory.
- 1286 • **Octave Symbolic**: The Octave-Forge Symbolic package adds symbolic calculation  
1287 features to GNU Octave. These include common Computer Algebra  
1288 System tools such as algebraic operations, calculus, equation solving, Fourier  
1289 and Laplace transforms, variable precision arithmetic and other features.
- 1290 • **SymPy.jl**: Provides a Julia interface to SymPy using PyCall.
- 1291 • **Mathics**: Mathics is a free, general-purpose online CAS featuring Mathematica  
1292 compatible syntax and functions. It is backed by highly extensible  
1293 Python code, relying on SymPy for most mathematical tasks.
- 1294 • **Mathpix**: An iOS App, that uses Artificial Intelligence to detect handwritten  
1295 math as input, and uses SymPy Gamma, to evaluate the math input and  
1296 generate the relevant steps to solve the problem.
- 1297 • **IKFast**: IKFast is a robot kinematics compiler provided by **OpenRAVE**.  
1298 It analytically solves robot inverse kinematics equations and generates optimized  
1299 C++ files. It uses SymPy for its internal symbolic mathematics.
- 1300 • **Sage**: A CAS, visioned to be a viable free open source alternative to Magma,  
1301 Maple, Mathematica and Matlab.
- 1302 • **SageMathCloud**: SageMathCloud is a web-based cloud computing and  
1303 course management platform for computational mathematics.
- 1304 • **PyDy**: Multibody Dynamics with Python.
- 1305 • **galgebra**: Geometric algebra (previously `sympy.galgebra`).
- 1306 • **yt**: Python package for analyzing and visualizing volumetric data (`yt.units`  
1307 uses SymPy).
- 1308 • **SfePy**: Simple finite elements in Python, see section 9.11.1.
- 1309 • **Quameon**: Quantum Monte Carlo in Python.
- 1310 • **Lcapy**: Experimental Python package for teaching linear circuit analysis.
- 1311 • **Quantum Programming in Python**: Quantum 1D Simple Harmonic Oscillator  
1312 and Quantum Mapping Gate.
- 1313 • **LaTeX Expression project**: Easy LaTeX typesetting of algebraic expressions  
1314 in symbolic form with automatic substitution and result computation.
- 1315 • **Symbolic statistical modeling**: Adding statistical operations to complex  
1316 physical models.

1317 **9.11. Project Details.** Below we provide particular examples of SymPy use in  
1318 some of the projects listed above.

**9.11.1. SfePy.** **SfePy** (Simple finite elements in Python), cf. [19]. is a Python package for solving partial differential equations (PDEs) in 1D, 2D and 3D by the finite element (FE) method [54]. SymPy is used within this package mostly for code generation and testing, namely:

- generation of the hierarchical FE basis module, involving generation and symbolic differentiation of 1D Legendre and Lobatto polynomials, constructing the FE basis polynomials [47] and generating the C code;
- generation of symbolic conversion formulas for various groups of elastic constants [24] – provide any two of the Young’s modulus, Poisson’s ratio, bulk modulus, Lamé’s first parameter, shear modulus (Lamé’s second parameter) or longitudinal wave modulus and get the other ones;
- simple physical unit conversions, generation of consistent unit sets;
- testing FE solutions using method of manufactured (analytical) solutions – the differential operator of a PDE is symbolically applied and a symbolic right-hand side is created, evaluated in quadrature points, and subsequently used to obtain a numerical solution that is then compared to the analytical one;
- testing accuracy of 1D, 2D and 3D numerical quadrature formulas (cf. [9]) by generating polynomials of suitable orders, integrating them, and comparing the results with those obtained by the numerical quadrature.

**9.12. Tensors.** Ongoing work to provide the capabilities of tensor computer algebra has so far produced the `tensor` module. It is composed of three separated sub-modules, whose purposes are quite different: `tensor.indexed` and `tensor.indexed_methods` support indexed symbols, `tensor.array` contains facilities to operator on symbolic  $N$ -dimensional arrays, and finally `tensor.tensor` is used to define abstract tensors. The abstract tensors subsection is inspired by xAct [34] and Cadabra [39]. Canonicalization based on the Butler-Portugal [33] algorithm is supported in SymPy. It is currently limited to polynomial tensor expressions.

**9.13. Numerical simplification.** The `nsimplify` function in SymPy (a wrapper of `identify` in `mpmath`) attempts to find a simple symbolic expression that evaluates to the same numerical value as the given input. It works by applying a few simple transformations (including square roots, reciprocals, logarithms and exponentials) to the input and, for each transformed value, using the PSLQ algorithm [21] to search for a matching algebraic number or optionally a linear combination of user-provided base constants (such as  $\pi$ ).

```
>>> t = 1 / (sin(pi/5)+sin(2*pi/5)+sin(3*pi/5)+sin(4*pi/5))*2
>>> nsimplify(t)
-2*sqrt(5)/5 + 1
>>> nsimplify(pi, tolerance=0.01)
22/7
>>> nsimplify(1.783919626661888, [pi], tolerance=1e-12)
pi/(-1/3 + 2*pi/3)
```