Facets of SciPy

Conferences

Collection of Tools

Community
A view of SciPy History

Thursday, July 1, 2010
## Python origins


<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9.0</td>
<td>Feb. 1991</td>
</tr>
<tr>
<td>0.9.4</td>
<td>Dec. 1991</td>
</tr>
<tr>
<td>0.9.6</td>
<td>Apr. 1992</td>
</tr>
<tr>
<td>0.9.8</td>
<td>Jan. 1993</td>
</tr>
<tr>
<td>1.0.0</td>
<td>Jan. 1994</td>
</tr>
<tr>
<td>1.2</td>
<td>Apr. 1995</td>
</tr>
<tr>
<td>1.4</td>
<td>Oct. 1996</td>
</tr>
<tr>
<td>1.5.2</td>
<td>Apr. 1999</td>
</tr>
</tbody>
</table>
Jim Fulton

- Created first Matrix object
- Released it to the net in 1994
- Started getting feedback on Matrix-SIG
- Convinced Guido to change some syntax
  - a[0,1] instead of a[(0,1)]
  - a[::2] instead of only a[:]
  - Ellipsis operator
- Complex numbers
Jim Hugunin (~1995)
Contributed complex numbers to Python
Created ScientificPython (a collection of packages for doing a variety of things --- 26 Jul 1999)
Extremely helpful on mailing list in early discussions.
Wrote tutorial to Numeric C-API
Paul Dubois

- Took over as lead NumPy developer when Jim Hugunin went on to write Jython.
- Funded first documentation for Numeric Python.
- Created Computing in Science and Engineering special edition devoted to Python.
David Ascher

- Wrote initial documentation for Numeric Python with funding from LLNL.
- Active on mailing lists and in discussions until about 2001.
- Became CEO of Mozilla Messaging
How did I end up involved with NumPy and SciPy?
Early beginnings (1971 - 1989)

\[
\sum_{k=1}^{N} k = \frac{N (N + 1)}{2}
\]
Maxwell’s Equations (1990 -

\[ \nabla \times E = -\frac{\partial B}{\partial t} \]
\[ \nabla \cdot D = \rho \]
\[ \nabla \times H = \mathbf{J} + \frac{\partial D}{\partial t} \]
\[ \nabla \cdot B = 0 \]
Electrical Engineering

Electromagnetics

... so I could get a real job!

Probability Theory

Signal Processing
Master’s Degree
$\rho_0 (2\pi f)^2 U_i (a, f) = [C_{ijkl} (a, f) U_{k,l} (a, f)]_{,j}$
Finding Derivatives of 5-d data

\[ U_X (a, f) \quad U_Y (a, f) \quad U_Z (a, f) \]
Finding Derivatives of 5-d data

\[ \Xi = \nabla \times \mathbf{U} \]

\[ \Xi_X(a, f) \]

\[ \Xi_Y(a, f) \]

\[ \Xi_Z(a, f) \]
Found Python and Numeric in 1997

I was a fairly proficient MATLAB user, but it was not memory efficient enough.

- Loved the expressive syntax of Python
- Loved the fact that slicing didn’t make copies
- Loved the existing multiple data-types
- Loved how much more flexible it was to extend than MATLAB was
- Loved that I could read the source code and extend it
First problem: Efficient Data Input

“It’s All About the Data”

TableIO
April 1998

Reference Counting Essay
http://www.python.org/doc/essays/refcnt/
May 1998

NumPyIO
June 1998

Michael A. Miller

Guido van Rossum
Early pieces of SciPy

**fftw wrappers**
June 1998

**cephesmodule**
November 1998

**stats.py**
December 1998

Gary Strangman
1999 : Early SciPy emerges


In response on 15 Jan, 1999, I posted to matrix-sig a list of routines I felt needed to be present and began wrapping / writing in earnest. On 6 April 1999, I announced I would be creating this uber-package which eventually became SciPy.

<table>
<thead>
<tr>
<th>Gaussian quadrature</th>
<th>5 Jan 1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>cephes 1.0</td>
<td>30 Jan 1999</td>
</tr>
<tr>
<td>sigtools 0.40</td>
<td>23 Feb 1999</td>
</tr>
<tr>
<td>Numeric docs</td>
<td>March 1999</td>
</tr>
<tr>
<td>cephes 1.1</td>
<td>9 Mar 1999</td>
</tr>
<tr>
<td>multipack 0.3</td>
<td>13 Apr 1999</td>
</tr>
<tr>
<td>Helper routines</td>
<td>14 Apr 1999</td>
</tr>
<tr>
<td>multipack 0.6 (leastsq, ode, fsolve, quad)</td>
<td>29 Apr 1999</td>
</tr>
<tr>
<td>sparse plan described</td>
<td>30 May 1999</td>
</tr>
<tr>
<td>multipack 0.7</td>
<td>14 Jun 1999</td>
</tr>
<tr>
<td>SparsePy 0.1</td>
<td>5 Nov 1999</td>
</tr>
<tr>
<td>cephes 1.2 (vectorize)</td>
<td>29 Dec 1999</td>
</tr>
</tbody>
</table>

Plotting??

Gist
XPLOT
DISLIN
Gnuplot

Helping with f2py
Early contributors in 1999

Pearu Peterson
- Hosting of first Multipack CVS repository (June 1999)
- Amazing makefiles
- Interface to FITPACK
- Wrote f2py as he watched my brute-force approach (July 1999)

Janko Hauser
- (IPP) Early IPython interactive environment (27 Apr 1999)
- Matlab file reader (24 Apr 1999)

Robert Kern
- Created windows binaries of multipack, cephesmodule, fftw, and signaltools (June 1999 while still in high school!)
Early 2000: Numeric needs

- Memory mapped arrays
- Rank-0 arrays or scalars
- Handling indirect indexing: \( a[[10,5,7]] \)
- Handling masked indexing: \( a[[\text{True, False, False}]] \)
- More attributes to N-d arrays
- “Record arrays”
Finally... Plotting

matplotlib

John Hunter
2001
IPython (building on IPP)

Fernando Perez
Dec 10, 2001
Teaching

Research

$??$
SciPy work at BYU from students

- Iterative solvers
- Special Function tests
- Some additions to Numeric

No time for Numeric overhaul contemplated in early 2000!
STSCI leads out with Numarray

Perry Greenfield
J. Todd Miller
Rick White
Paul Barrett

Thursday, July 1, 2010
Numarray released in 2003

- too slow for small arrays
- incomplete implementation for ufuncs
- minimal Numeric code re-use

- lots of very nice things, though (e.g. memory maps, fast code for large arrays, better sorting algorithms)
Split in the community

Numeric
SciPy

Numarray
ndimage
others
ndimage
NumPy
Version 1.0 October 2006

Key contributions from:
  Numarray
  Numeric
  Chuck Harris
  Robert Kern
  David Cooke
  Pierre GM

started January 2005
Q: When is NumPy going to be part of the core Python?

A: Data structure is in Python as PEP 3118
Elastography reconstruction

Ehman,
Scanning Impedance Imaging
Transition to Industry

Traits

Mayavi
TVTK
VTK

Chaco
Envisage
Enable
Kiva

Visualization
Optimization
CFD EM Integration
CFD Simulator
EM Simulator
Army Research Labs (ARL)

- Electromagnetics Simulation
- Python/C/Fortan90

- Fine grained Parallelism (MPI based)
- Desktop Application
- Cluster Execution
Fluid Characterization -- Microrheology
VMS – Virtual Mixing System

- Design Drawings
- Computational Fluid Dynamics
- Parallel Simulation
- Data Visualization
High Speed Signal Integrity Testing
Community effort

- Chuck Harris
- Pauli Virtanen
- David Cournapeau
- Stefan van der Walt
- Jarrod Millman
- Josef Perktold
- Anne Archibald
- Dag Sverre Seljebotn
- Robert Kern
- Matthew Brett
- Warren Weckesser
- Ralf Gommers
- Joe Harrington --- Documentation effort
- Andrew Straw --- www.scipy.org

many, many others --- forgive me!
In 2009

North America
In 2009

Europe
### SciPy Unique Visits – 2009 (~400K total)

(first half of)

<table>
<thead>
<tr>
<th>Universities</th>
<th>Location</th>
<th>Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT</td>
<td></td>
<td>2965</td>
</tr>
<tr>
<td>Harvard</td>
<td></td>
<td>1553</td>
</tr>
<tr>
<td>Stanford</td>
<td></td>
<td>1534</td>
</tr>
<tr>
<td>Cambridge</td>
<td></td>
<td>1214</td>
</tr>
<tr>
<td>CalTech</td>
<td></td>
<td>1121</td>
</tr>
<tr>
<td>Cornell</td>
<td></td>
<td>1109</td>
</tr>
<tr>
<td>U Wash</td>
<td></td>
<td>1011</td>
</tr>
<tr>
<td>Berkeley</td>
<td></td>
<td>958</td>
</tr>
<tr>
<td>Oxford</td>
<td></td>
<td>811</td>
</tr>
<tr>
<td>Univ. BC</td>
<td></td>
<td>795</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Labs</th>
<th>Location</th>
<th>Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASA</td>
<td></td>
<td>2378</td>
</tr>
<tr>
<td>LANL</td>
<td></td>
<td>508</td>
</tr>
<tr>
<td>LLNL</td>
<td></td>
<td>420</td>
</tr>
<tr>
<td>LBNL</td>
<td></td>
<td>368</td>
</tr>
<tr>
<td>ORNL</td>
<td></td>
<td>205</td>
</tr>
<tr>
<td>Brookhaven</td>
<td></td>
<td>172</td>
</tr>
<tr>
<td>Argonne</td>
<td></td>
<td>158</td>
</tr>
<tr>
<td>PNL</td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

680K total visits in first half of 2010!!
### Tech Sampling

<table>
<thead>
<tr>
<th>Location</th>
<th>Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seagate</td>
<td>1018</td>
</tr>
<tr>
<td>Intel</td>
<td>390</td>
</tr>
<tr>
<td>Microsoft</td>
<td>246</td>
</tr>
<tr>
<td>HP</td>
<td>241</td>
</tr>
<tr>
<td>IBM</td>
<td>146</td>
</tr>
<tr>
<td>Google</td>
<td>85</td>
</tr>
<tr>
<td>Freescale</td>
<td>64</td>
</tr>
</tbody>
</table>

### Industrial Sampling

<table>
<thead>
<tr>
<th>Location</th>
<th>Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boeing</td>
<td>286</td>
</tr>
<tr>
<td>Caterpillar</td>
<td>205</td>
</tr>
<tr>
<td>Airbus</td>
<td>158</td>
</tr>
<tr>
<td>P&amp;G</td>
<td>51</td>
</tr>
<tr>
<td>Rolls Royce</td>
<td>30</td>
</tr>
</tbody>
</table>

680K total visits in first half of 2010!!
### SciPy Unique Visits – 2009 (~400K total)

(First half of)

<table>
<thead>
<tr>
<th>Location</th>
<th>Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE Shaw</td>
<td>693</td>
</tr>
<tr>
<td>Citadel</td>
<td>373</td>
</tr>
<tr>
<td>JP Morgan</td>
<td>340</td>
</tr>
<tr>
<td>Softbank</td>
<td>315</td>
</tr>
<tr>
<td>Goldman Sachs</td>
<td>273</td>
</tr>
<tr>
<td>AQR Capital</td>
<td>171</td>
</tr>
<tr>
<td>Susquehanna</td>
<td>159</td>
</tr>
<tr>
<td>UBS</td>
<td>160</td>
</tr>
<tr>
<td>Tradelink</td>
<td>129</td>
</tr>
<tr>
<td>Deutsche Bank</td>
<td>78</td>
</tr>
<tr>
<td>Merrill Lynch</td>
<td>77</td>
</tr>
<tr>
<td>CitiGroup</td>
<td>62</td>
</tr>
<tr>
<td>Morgan Stanley</td>
<td>40</td>
</tr>
<tr>
<td>Renaissance</td>
<td>20</td>
</tr>
</tbody>
</table>

### Oil Sampling

<table>
<thead>
<tr>
<th>Location</th>
<th>Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shell</td>
<td>160</td>
</tr>
<tr>
<td>Stat Oil</td>
<td>155</td>
</tr>
<tr>
<td>Schlumberger</td>
<td>110</td>
</tr>
<tr>
<td>Exxon</td>
<td>76</td>
</tr>
<tr>
<td>ConocoPhillips</td>
<td>27</td>
</tr>
<tr>
<td>Aramco</td>
<td>27</td>
</tr>
</tbody>
</table>

680K total visits in first half of 2010!!
Perspectives
Why Python for Technical Computing

• syntax, syntax, syntax
  – “it doesn’t get in my way” --- Robert Kern
  – “it fits your brain” --- Guido
• multiple styles of programming
• meta-programming
• complex numbers
• first-class functions
• powerful array object
• Zen of Python ("practicality beats purity")
What needs to be improved

- build and distribution ecosystem
- modularity of SciPy (ease of contribution for occasional participant)
- easy creation / performance of native code (fast for loops matter)
- parallelization and concurrency strategies
- patterns for dealing with streamed and distributed data
- person-power (more people with more time)
NumPy enhancements

• date-time data-type
• group-by
• SQL execution context
• fields of context to put “fields” into namespace
• arithmetic on structured data-types
• some kind of “labeled arrays” ??
• indirect array ??
SciPy Improvements

• interpolate API improvements
• stats improvements
• better documentation
• general API improvements throughout code
• more examples
• more tests
• modularity
• migration to Python3
Need for Champions in SciPy

• Thriving community
• Relies on volunteer labor and people willing to step-up and contribute
  – code
  – docs
  – examples
  – mailing list participants
Extending the reach of NumPy / SciPy

- C/C++
- Fortran
- C/C++
- Fortran

- C# / .NET
- Java
- IronPython
- Jython
Potential Impact on community

- NumPy refactor
- Full time attention to NumPy and SciPy
- Possibility to extend NumPy to Jython and potentially even other languages
- Adoption by entire new community of users

- Timeline --- NumPy by September, SciPy by December
Features of IronPython (from .NET)

• Sophisticated Garbage collection
• Threading
• Faster for loops (JIT compilation)
• Access to large set of libraries
• Glue with .NET