

Introduction, practical guide, and limitations

Andrew James

May 23, 2022

Slides on github: amjames/pybind11-tech-share

Motivations and goals of the project

Motivations and goals of the project Binding Functions

Motivations and goals of the project

Binding Functions

Binding C++ classes

Motivations and goals of the project

Binding Functions

Binding C++ classes

Overloads

Motivations and goals of the project

Binding Functions

Binding C++ classes

Overloads

NumPy/ Buffer Protocol

Motivations and goals of the project

Binding Functions

Binding C++ classes

Overloads

NumPy/ Buffer Protocol

Embedding the Python Interpreter

Motivations and goals of the project

Binding Functions

Binding C++ classes

Overloads

NumPy/ Buffer Protocol

Embedding the Python Interpreter

What is going on in the background

What is pybind11



pybind11 is a lightweight header-only library that exposes C++ types in Python and vice versa, mainly to create Python bindings of existing C++ code.

- Generate C++ wrappers for native python types.
- Python PyObject wrappers for C++ types.

ctypes

The most low-level option. Your library must have an extern "C" interface. It is not easy to export types from your library, and you must use C data types at the call site. It does not scale well.

ctypes

The most low-level option. Your library must have an extern "C" interface. It is not easy to export types from your library, and you must use C data types at the call site. It does not scale well.

Cython

Generating bindings for a C++ library will require writing and maintaining a layer in the Cython language which is neither python, nor C. Shines when you want to accelerate parts of the codebase.

ctypes

The most low-level option. Your library must have an extern "C" interface. It is not easy to export types from your library, and you must use C data types at the call site. It does not scale well.

Cython

Generating bindings for a C++ library will require writing and maintaining a layer in the Cython language which is neither python, nor C. Shines when you want to accelerate parts of the codebase.

Boost.Python

Very similar to pybind11 in features and syntax. You write binding code in C++. Mostly, you are selectively adding features to the python API. The main drawback is boost itself.

Motivations



Think of this library as a tiny self-contained version of Boost.Python with everything stripped away that isn't relevant for binding generation.

This compact implementation was possible thanks to some of the new C++11 features...

Leveraging C++11 features to write a compact library for python bindings, hence the name pybind11.

Writing a simple module

cppimport is a neat little tool for playing around with simple pybind11 extensions. I have used for self contained examples of some concepts.

My environment:

conda create -n pybind-examples pybind11 ipython pip
conda activate pybind-examples
pip install cppimport

Writing a simple module

```
// cppimport
#include <pybind11/pybind11.h>
double add(double a, double b) {
  return a+b;
PYBIND11 MODULE(example1, m) {
  m.def("add", &add);
//
//<%
//setup pybind11(cfg)
// %>
//
```

A few notes:

- 1st line Indicates that the file should be importable.
- Entire example fits on a slide!
- I will be showing the important stuff moving forward, but all of the examples will be on github.

Writing a simple module

```
#include <pybind11/pybind11.h>

double add(double a, double b) {
  return a+b;
}

PYBIND11_MODULE(example1, m) {
  m.def("add", &add);
}
```

We can import this and run it directly:

```
>>> import cppimport.import_hook
>>> import example1
>>> example1.add(2.5, 3.2)
5.7
```

It will take a moment to (re)-compile after editing

When generating python bindings it is important to consider that the user of that API will be expecting certain things from a python library. Docstrings area great example, there is no need to support help() in C++.

```
>>> help(example1.add)
...
add(...) method of builtins.PyCapsule instance
add(arg0: float, arg1: float) -> float
```

```
>>> help(example1.add)
...
add(...) method of builtins.PyCapsule instance
   add(arg0: float, arg1: float) -> float
```

This is pretty good!

```
>>> help(example1.add)
...
add(...) method of builtins.PyCapsule instance
   add(arg0: float, arg1: float) -> float
```

This is pretty good!

We are missing argument names.

Docstrings - named arguments

```
namespace py = pybind11;
PYBIND11_MODULE(example1, m) {
   m.def("subtract", &subtract, "Computes a - b",
        py::arg("a"), py::arg("b"));
}
```

This alias is widely adopted convention

Similar to import numpy as np

Docstrings - named arguments

```
namespace py = pybind11;
PYBIND11_MODULE(example1, m) {
   m.def("subtract", &subtract, "Computes a - b",
        py::arg("a"), py::arg("b"));
}
```

We have now attached names to our arguments by annotating the function binding with some additional information.

Docstrings - More Metadata

```
namespace py = pybind11;
PYBIND11_MODULE(example1, m) {
   m.def("subtract", &subtract, "Computes a - b",
        py::arg("a") = 1, py::arg("b") = 2);
}
```

We can also specify the default values for the arguments.

The number of arguments is statically checked, however the types are not!

Docstrings - More Metadata

We may also use the _a suffix (C++11 literals)

Now the docstring will look more complete

```
>>> help(example2.subtract)
...
subtract(...) method of builtins.PyCapsule instance
    subtract(a: float = 1, b: float = 2) -> float
Computes a - b
```

```
Using py::class_:
```

- Creates a binding for a C++ class or struct.
- The py:class has methods available for binding additional information to the class
- Very similar to py::module (Without additional boilerplate, so no macro is required)

```
Using py::class_:
```

- Creates a binding for a C++ class or struct.
- The py:class_ has methods available for binding additional information to the class
- Very similar to py::module (Without additional boilerplate, so no macro is required)

```
struct Pet {
    std::string name;
};
```

```
namespace py = pybind11;
PYBIND11_MODULE(example3, m) {
   py::class_<Pet>(m, "Pet")
     .def(py::init<const std::string&>());
}
```

Using py::class_:

- Creates a binding for a C++ class or struct.
- The py:class has methods available for binding additional information to the class
- Very similar to py::module (Without additional boilerplate, so no macro is required)

```
struct Pet {
    std::string name;
};
```

```
namespace py = pybind11;
PYBIND11_MODULE(example3, m) {
   py::class_<Pet>(m, "Pet")
     .def(py::init<const std::string&>());
}
```

- The py::init wrapper is used to bind constructors for a class
- The template parameters should correspond to a constructor signature
- Without template parameters a callable can be provided which returns the type by value or the appropriate "holder"

We use the py::class_object to define the python visible interface for the class.

Anything we want to be available in python, must be explicitly declared in the binding

We use the py::class object to define the python visible interface for the class.

Anything we want to be available in python, must be explicitly declared in the binding

```
import example3
p = example3.Pet("Spot")
print(p.name) # AttributeError!
```

We use the py::class_object to define the python visible interface for the class.

Anything we want to be available in python, must be explicitly declared in the binding

```
import example3
p = example3.Pet("Spot")
print(p.name) # AttributeError!
```

We haven not added an attribute name to the python interface for the Pet class.

```
namespace py = pybind11;
PYBIND11_MODULE(example3, m) {
   py::class_<Pet>(m, "Pet")
      .def(py::init<const std::string&>());
}
```

Binding a C++ class - Attributes

```
namespace py = pybind11;
PYBIND11_MODULE(example3, m) {
   py::class_<Pet>(m, "Pet")
      .def(py::init<const std::string&>());
}
```

Binding a C++ class - Attributes

```
namespace py = pybind11;
PYBIND11_MODULE(example3, m) {
   py::class_<Pet>(m, "Pet")
     .def(py::init<const std::string&>());
}
```

We have a few different options:

Attributes:

```
.def_readonly("name", &Pet::name)
.def_readwrite("name", &Pet::name)
```

Binding a C++ class - Attributes

```
namespace py = pybind11;
PYBIND11_MODULE(example3, m) {
   py::class_<Pet>(m, "Pet")
     .def(py::init<const std::string&>());
}
```

We have a few different options:

Attributes:

```
.def_readonly("name", &Pet::name)
.def_readwrite("name", &Pet::name)
```

Or property:

```
.def_property("name", /*getter*/, /*setter*/)
.def_property_readonly("name", /*getter*/)
.def_property("name", nullptr, /*setter*/) //Write-only property
```

Binding a C++ class - Attributes

If we didn't have a getter/setter already the attribute route is a good one.

We can also use a lambda.

```
.def_property("name",
    /*getter*/[](const Pet& self) { return self.name; },
    /*setter*/[](Pet& self, std::string value) { self.name = value; })
```

Lambdas are particularly useful when you want to do something more pythonic, but don't want to introduce py::types into your library code

```
enum FeatureTypes{...};
std::tuple<bool, bool, bool,...> check_features(...);
```

...

Lambdas are particularly useful when you want to do something more pythonic, but don't want to introduce py::types into your library code

```
enum FeatureTypes{...};
std::tuple<bool, bool, bool,...> check_features(...);
...

//In binding code
using namespace pybind11::literals;
m.def("check_features", [](...) {
   auto feature_tuple = check_features(...);
   return py::dict("feature_a"_a = std::get<FeatureTypes::A>(feature_tuple),...);
});
```

Lambdas are particularly useful when you want to do something more pythonic, but don't want to introduce py::types into your library code

```
enum FeatureTypes{...};
std::tuple<bool, bool, bool,...> check_features(...);

...

//In binding code
using namespace pybind11::literals;
m.def("check_features", [](...) {
   auto feature_tuple = check_features(...);
   return py::dict("feature_a"_a = std::get<FeatureTypes::A>(feature_tuple),...);
   });
```

Using lambdas at the binding layer to translate c++ patterns to pythonic variants is a popular practice.

If a function has multiple overloads we will have some trouble with the basic pattern for generating a binding

```
struct Reader {
    // reads everything
    size_t read()
    // reads from begin to the end
    size_t read(size_t begin)
    // reads from offset begin to offset end
    size_t read(size_t begin, size_t end)
};
```

If a function has multiple overloads we will have some trouble with the basic pattern for generating a binding

```
struct Reader {
    // reads everything
    size_t read()
    // reads from begin to the end
    size_t read(size_t begin)
    // reads from offset begin to offset end
    size_t read(size_t begin, size_t end)
};
```

```
py::class_<Reader>(m, "Reader")
   .def("read", &Reader::read)
```

If a function has multiple overloads we will have some trouble with the basic pattern for generating a binding

```
struct Reader {
    // reads everything
    size_t read()
    // reads from begin to the end
    size_t read(size_t begin)
    // reads from offset begin to offset end
    size_t read(size_t begin, size_t end)
};
```

```
py::class_<Reader>(m, "Reader")
   .def("read", &Reader::read)
```

The compiler is not able to read your mind!

We can disambiguate by casting to a function pointer

```
py::class_<Reader>(m, "Reader")
   .def("read", static_cast<size_t (Reader::*)(size_t)>(&Rader::read))
```

We can disambiguate by casting to a function pointer

```
py::class_<Reader>(m, "Reader")
   .def("read", static_cast<size_t (Reader::*)(size_t)>(&Rader::read))
```

Or there is a handy helper (C++14)

```
py::class_<Reader>(m, "Reader")
   .def("read", py::overload_cast<size_t>(&Reader::read))
```

Note: The py::init wrapper we use to bind constructors takes care of this for us

It would not be very useful if we could only bind a single overload of a function.

It would not be very useful if we could only bind a single overload of a function.

We could manually differentiate by binding to different names:

```
py::class_<Reader>(m, "Reader")
    .def("read_all", py::overload_cast<>(&Reader::read))
    .def("read_from", py::overload_cast<size_t>(&Reader::read))
    .def("read_between", py::overload_cast<size_t,size_t>(&Reader::read))
```

It would not be very useful if we could only bind a single overload of a function.

We could manually differentiate by binding to different names:

```
py::class_<Reader>(m, "Reader")
    .def("read_all", py::overload_cast<>(&Reader::read))
    .def("read_from", py::overload_cast<size_t>(&Reader::read))
    .def("read_between", py::overload_cast<size_t,size_t>(&Reader::read))
```

It is not desirable to have our python and C++ api diverge like this.

It would not be very useful if we could only bind a single overload of a function.

We could manually differentiate by binding to different names:

```
py::class_<Reader>(m, "Reader")
    .def("read_all", py::overload_cast<>(&Reader::read))
    .def("read_from", py::overload_cast<size_t>(&Reader::read))
    .def("read_between", py::overload_cast<size_t,size_t>(&Reader::read))
```

It is not desirable to have our python and C++ api diverge like this.

Thankfully pybind allows us to bind multiple methods to the same name!

```
py::class_<Reader>(m, "reader")
    .def("read", py::overload_cast<>(&Reader::read))
    .def("read", py::overload_cast<size_t>(&Reader::read))
    .def("read", py::overload_cast<size_t,size_t>(&Reader::read))
```

```
py::class_<Reader>(m, "Reader")
    .def("read", py::overload_cast<>(&Reader::read))
    .def("read", py::overload_cast<size_t>(&Reader::read))
    .def("read", py::overload_cast<size_t,size_t>(&Reader::read))
```

How does that work?

```
py::class_<Reader>(m, "Reader")
    .def("read", py::overload_cast<>(&Reader::read))
    .def("read", py::overload_cast<size_t>(&Reader::read))
    .def("read", py::overload_cast<size_t,size_t>(&Reader::read))
```

How does that work?

When we make the first call to py::class_::def pybind is going to create an entry in the class __dict__ for the method read. It points to the wrapper built around the bound method.

```
py::class_<Reader>(m, "Reader")
    .def("read", py::overload_cast<>(&Reader::read))
    .def("read", py::overload_cast<size_t>(&Reader::read))
    .def("read", py::overload_cast<size_t,size_t>(&Reader::read))
```

How does that work?

When we make the first call to py::class_::def pybind is going to create an entry in the class __dict__ for the method read. It points to the wrapper built around the bound method.

When we .def the second and third overloads, the existence of the first is detected. This triggers the new method to be added to the end of the linked list of methods in the slot.

```
py::class_<Reader>(m, "Reader")
    .def("read", py::overload_cast<>(&Reader::read))
    .def("read", py::overload_cast<size_t>(&Reader::read))
    .def("read", py::overload_cast<size_t,size_t>(&Reader::read))
```

How does that work?

When we make the first call to py::class_::def pybind is going to create an entry in the class __dict__ for the method read. It points to the wrapper built around the bound method.

When we .def the second and third overloads, the existence of the first is detected. This triggers the new method to be added to the end of the linked list of methods in the slot.

There is some inspection of the metadata for the function to make sure the overload is compatible.

```
py::class_<Reader>(m, "Reader")
    .def("read", py::overload_cast<>(&Reader::read))
    .def("read", py::overload_cast<size_t>(&Reader::read))
    .def("read", py::overload_cast<size_t,size_t>(&Reader::read))
```

How does that work?

When we make the first call to py::class_::def pybind is going to create an entry in the class __dict__ for the method read. It points to the wrapper built around the bound method.

When we .def the second and third overloads, the existence of the first is detected. This triggers the new method to be added to the end of the linked list of methods in the slot.

There is some inspection of the metadata for the function to make sure the overload is compatible.

In general every callable (instance bound, class bound, or module bound) is set up this way. Even if no overloads are defined.

If you want to influence overload resolution:

- py::arg().noconvert() to prevent casting of an argument even in the relaxed pass
- The order of .def statements will change the order of the list and therefore the search.
- You can add py::prepend to the tags section of the def to place it at the beginning of the chain.

The first matching overload is always selected to be called.

There is no priority for minimizing the number of casts required for example.

Consider:

Consider:

When calling foo ([1,2,3]) in python, the layer wrapping the function foo applies a thin py::list wrapper around some PyListObject.

Consider:

When calling foo ([1,2,3]) in python, the layer wrapping the function foo applies a thin py::list wrapper around some PyListObject.

The return involves creating a $py::float_in c++$, but it is also a thin wrapper around a native python type and removed on the way out.

Consider:

When calling foo ([1,2,3]) in python, the layer wrapping the function foo applies a thin py::list wrapper around some PyListObject.

The return involves creating a $py::float_in c++$, but it is also a thin wrapper around a native python type and removed on the way out.

When the wrapping layer for baz sees the Python interface for a py::class_<CustomT> it does much the same, pulling a wrapper off to expose the native C++ object.

Consider:

When calling foo ([1,2,3]) in python, the layer wrapping the function foo applies a thin py::list wrapper around some PyListObject.

The return involves creating a $py::float_in c++$, but it is also a thin wrapper around a native python type and removed on the way out.

When the wrapping layer for baz sees the Python interface for a py::class_<CustomT> it does much the same, pulling a wrapper off to expose the native C++ object.

None of these operations are considered casts

Consider:

What happens with bar?

Consider:

What happens with bar?

The wrapping layer finds no appropriate overload in the strict search.

Consider:

What happens with bar?

The wrapping layer finds no appropriate overload in the strict search.

If the argument were a list or tuple and the contents were all numeric objects. It would be able to cast and make the call.

Consider:

What happens with bar?

The wrapping layer finds no appropriate overload in the strict search.

If the argument were a list or tuple and the contents were all numeric objects. It would be able to cast and make the call.

The key difference is there is no *simple* conversion involving the addition/removal of a pybind wrapper.

Consider:

What happens with bar?

The wrapping layer finds no appropriate overload in the strict search.

If the argument were a list or tuple and the contents were all numeric objects. It would be able to cast and make the call.

The key difference is there is no *simple* conversion involving the addition/removal of a pybind wrapper.

This is important to consider when defining overloads. If bar had an overload accepting py::list the std::vector<double> version would never match if passed a list on the python side.

Holder types

All pybind wrapped types py::class_<T> uses a special holder type to manage references to the object. The default is to use std::unique_ptr<T>.

This will mean that the pybind11 wrapper is going to reference count for us and delete the object when it no longer referenced anywhere.

Sometimes, a codebase may rely on std::shared_ptr<T> heavily, and it is possible to specify this as holder type for the object.

```
class AlwaysShared {};
py::class_<AlwaysShared, std::shared_ptr<AlwaysShared>>(m, "AlwaysShared");
```

Now pybind11 will allow the smart pointer to do the reference counting.

Holder types

All pybind wrapped types py::class_<T> uses a special holder type to manage references to the object. The default is to use std::unique ptr<T>.

This will mean that the pybind11 wrapper is going to reference count for us and delete the object when it no longer referenced anywhere.

Sometimes, a codebase may rely on std::shared_ptr<T> heavily, and it is possible to specify this as holder type for the object.

```
class AlwaysShared {};
py::class_<AlwaysShared, std::shared_ptr<AlwaysShared>>(m, "AlwaysShared");
```

Now pybind11 will allow the smart pointer to do the reference counting.

Careful! Any bound function returning a raw pointer will improperly be captured in a *new* shared ptr.

Built in support for the Python Buffer protocol

```
class Matrix {
public:
    Matrix(size t rows, size t cols)
        : m rows (rows)
        , m cols(cols) {
        m data = new float[rows*cols];
    float *data() { return m data; }
    size t rows() const { return m rows; }
    size t cols() const { return m cols; }
private:
    size t m rows, m cols;
    float *m data;
};
```

```
py::class <Matrix>(m, "Matrix",
 py::buffer protocol()
).def buffer([](Matrix &m){
 return py::buffer info(
   /* Pointer to buffer */
   m.data(),
   /* Size of one scalar */
   sizeof(float),
   /* Python struct format descriptor */
   py::format descriptor<float>::format(),
    /* Number of dimensions */
    2,
    /* Buffer dimensions */
    { m.rows(), m.cols() },
    /* Strides (in bytes) */
    { sizeof(float) * m.cols(),
      sizeof(float) }
 );
});
```

This makes it possible construct a NumPy array from our Matrix object without an expensive copy of the underlying data.

We could design a constructor for our Maxtrix object which makes it a view on some NumPy owned memory.

This makes it possible construct a NumPy array from our Matrix object without an expensive copy of the underlying data.

We could design a constructor for our Maxtrix object which makes it a view on some NumPy owned memory.

There is also direct support for NumPy arrays py::array.

This makes it possible construct a NumPy array from our Matrix object without an expensive copy of the underlying data.

We could design a constructor for our Maxtrix object which makes it a view on some NumPy owned memory.

There is also direct support for NumPy arrays py::array.

A py::vectorize wrapper for transforming scalar ops into vectorized point wise ops over NumPy arrays

```
double add_3(double a, double b, double c) { return a + b + c }
m.def("vectorized_add_3", py::vectorize(add_3))"
```

Embedding the interpreter

```
# plots.py
def plot x squared(x):
    import matplotlib.pyplot as plt
    plt.plot(x, x^{**2})
    plt.show()
#include <pybind11/embed.h>
namespace py = pybind11;
int main()
    using namespace py::literals;
    py::scoped interpreter guard{};
    py::module np = py::module ::import("numpy")
    py::object x = np.attr("arange")(0, 100);
    py::module plots= py::module ::import("plots");
    plots.attr("plot x squared")(x);
    return 0;
```

Very easy to embed a scripting console in your application!

Thank you!

Andrew James

May 23, 2022

Slides on github: amjames/pybind11-tech-share