

Contemporary C++ in HPC codes

lakub Gałecki^{1,2}

October 17th 2023, UQ Computers & Fluids Discussion Group

¹ Department of Aerodynamics, Warsaw University of Technology

 $^{^{2}}$ Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw



Why C++?

- Lingua franca for HPC
- C interop
- Zero cost abstractions
- Rich ecosystem of existing (legacy) software
- Still growing & evolving!



Why C++?

- · Lingua franca for HPC
- C interop
- · Zero cost abstractions
- Rich ecosystem of existing (legacy) software
- Still growing & evolving!

Why should I care if I'm coding in X?

- · What facilities does X provide for the problems of interest?
- How can I improve X?



Why C++?

- · Lingua franca for HPC
- · Cinterop
- · Zero cost abstractions
- Rich ecosystem of existing (legacy) software
- Still growing & evolving!

Why should I care if I'm coding in X?

- · What facilities does X provide for the problems of interest?
- · How can I improve X?

Maintainability carries real-world cost



Why C++?

- · Lingua franca for HPC
- · Cinterop
- · Zero cost abstractions
- · Rich ecosystem of existing (legacy) software
- Still growing & evolving!

Why should I care if I'm coding in X?

- · What facilities does X provide for the problems of interest?
- How can I improve X?

Maintainability carries real-world cost

Naming: modern C++ = $\{11, 14\}$, contemporary C++ ≥ 17



- · Background & motivation
- Examples:
 - · Miscellanea
 - · MPI wrapper
 - · Physics injection



- · Background & motivation
- Examples:
 - · Miscellanea
 - MPI wrapper
 - · Physics injection

Please interrupt if you have questions/comments



- · Background & motivation
- Examples:
 - Miscellanea
 - MPI wrapper
 - · Physics injection

Please interrupt if you have questions/comments

Click to see code: 🧟

The least-squares finite element

method, L3STER

The least-squares principle



Consider the abstract boundary-value problem:

$$\begin{cases} \mathscr{A}(u) = f & \text{in } \Omega \\ \mathscr{B}(u) = g & \text{on } \partial \Omega \end{cases}$$

We define the least-squares functional:

$$\mathscr{J}\left(u;f,g\right) = \frac{1}{2}\left(\left\|\mathscr{A}\left(u\right) - f\right\|_{\Omega,0}^{2} + \left\|\mathscr{B}\left(u\right) - g\right\|_{\partial\Omega,0}^{2}\right)$$

The least-squares principle

Find $\mathbf{u} \in \mathcal{V}$ such that

$$\forall_{\tilde{u}\in\mathscr{V}}\mathscr{J}\left(u;f,g\right)\leq\mathscr{J}\left(\tilde{u};f,g\right)$$



Restrict \mathscr{V} to a finite-dimensional $\mathscr{V}_h = \operatorname{span} \{\Phi_1, ..., \Phi_N\}$

$$u = \sum_{i=1}^{N} u_i \Phi_i$$
, where $u_i \in \mathbb{R}$

We now have a convex algebraic minimization problem, where the stationary point must satisfy

$$\left[\int_{\Omega} \left(\mathscr{A}^{\mathsf{T}} \Phi_{i} \right) \mathscr{A} \Phi_{j} + \int_{\partial \Omega} \left(\mathscr{B}^{\mathsf{T}} \Phi_{i} \right) \mathscr{B} \Phi_{j} \right] \left[u_{j} \right] \\
= \left[\int_{\Omega} \left(\mathscr{A}^{\mathsf{T}} \Phi_{i} \right) f + \int_{\partial \Omega} \left(\mathscr{B}^{\mathsf{T}} \Phi_{i} \right) g \right]$$

Alternative view: weak formulation where the derivatives of the result of the operator w.r.t. the unknowns serve as the trial functions



The domain contribution to the local matrix is given by

$$\mathbf{K} = \left[\int_{\Omega} \left(\mathscr{A}^{\mathsf{T}} \Phi_{i} \right) \mathscr{A} \Phi_{j} \right] \approx \sum_{q=1}^{Q} w_{q} \left[\left(\mathscr{A}^{\mathsf{T}} \Phi_{i} \left(\mathbf{x}_{q} \right) \right) \mathscr{A} \Phi_{j} \left(\mathbf{x}_{q} \right) \right] = \mathbf{A}^{\mathsf{T}} \mathbf{W}_{d} \mathbf{A}$$

- ✓ There's no weak formulation!
- ✓ BCs are trivial to impose (replace \mathscr{A} with \mathscr{B})
- √ The system is SPD regardless of
 ℳ
- ✓ Error norm is built in
- X A must be first-order and linear
- X Quadrature order is usually higher than for the Galerkin method, e.g. for 3D convection-diffusion $Q = 8p^3$



Linear first-order operators can be expressed as

$$\mathscr{A} = A_0 + \sum_{i=1}^{dim} A_i \frac{\partial}{\partial x_i}$$

where $A_i: \Omega \times (0,T] \to \mathbb{R}^{E \times U}$, E being the number of equations and U the number of unknowns (E and U may not be equal).

All we need to define a set of physics is to provide these matrices!

Different physics in different domains can be handled automatically.



- · Least-Squares Scalable SpecTral Element fRamework
- Written in C++ 20
- · General PDE solver no equations included in the core
- High premium placed on passing parameters (mostly kernel dimensions) at compile-time
- · Low entry barrier: equations + mesh = solution

Miscellaneous C++ bits for HPC

std::span



Span = pointer + size
Non-owning view over a contiguous memory region
Convenient API:

- · Compile-time or dynamically sized
- Sub-spans
- Range constructor

std::span



- Span = pointer + size
- Non-owning view over a contiguous memory region
- Convenient API:
 - · Compile-time or dynamically sized
 - Sub-spans
 - · Range constructor

```
auto v = std::vector{10, 9, 8, 7, 6, 5, 4, 3, 2, 1};
auto v_span = std::span{v};
auto drop3 = v_span.subspan(3);
auto drop3_take6 = drop3.subspan(0, 6);
std::print("{}\n", drop3_take6);
std::sort(drop3_take6.begin(), drop3_take6.end());
std::print("{}\n", v);
```



std::span



- Span = pointer + size
- Non-owning view over a contiguous memory region

Convenient API:

- · Compile-time or dynamically sized
- Sub-spans
- · Range constructor

```
[7, 6, 5, 4, 3, 2]
[10, 9, 8, 2, 3, 4, 5, 6, 7, 1]
```





- · Multi-dimensional span
- Non-owning view over a contiguous memory region, with multidimensional access information encoded
- API inspired by Kokkos
- Allows strided access and various memory layouts
- Can serve as common interface for matrix types from different libraries
- Finally multi-argument operator[]!
- · Sorry, still no support on Godbolt :(



```
template<typename T>
class atomic_ref;
```

Atomic reference type

Objects can be safely concurrently accessed from multiple threads, provided all accesses are via std::atomic_ref

Supports the usual set of operations and memory orders

Now full support for floating-point types!

std::atomic_ref



```
void atomic_increment(double& val, double inc) {
  std::atomic_ref{val}.fetch_add(inc /*, mem_order */);
}
```

std::atomic_ref

lock

ine

ret

.LBB0 1



cmpxchg qword ptr [rdi], rcx

void atomic_increment(double& val, double inc) {



std::atomic_ref



```
void atomic_increment(double& val, double inc) {
    std::atomic_ref{val}.fetch_add(inc /*, mem_order */);
 }
 atomic_increment(double&, double):
        mov
              rax, gword ptr [rdi]
 .LBB0 1:
               xmm1. rax
        mova
        addsd
               xmm1, xmm0
        movq
            rcx, xmm1
        lock
                     cmpxchg qword ptr [rdi], rcx
        ine
               .LBB0 1
```



Note the use of CTAD...

ret

MPI in contemporary C++





- · Communicator ownership semantics:
 - Own
 - View
 - Duplicate



- · Communicator ownership semantics:
 - · Own
 - View
 - Duplicate
- Comm buffer type deduction no more void*



- · Communicator ownership semantics:
 - · Own
 - View
 - Duplicate
- Comm buffer type deduction no more void*
- RAII for MPI requests (and other things)



- · Communicator ownership semantics:
 - Own
 - View
 - Duplicate
- Comm buffer type deduction no more void*
- RAII for MPI requests (and other things)
- Request tied with payload to automate memory management



- · Communicator ownership semantics:
 - Own
 - View
 - Duplicate
- Comm buffer type deduction no more void*
- RAII for MPI requests (and other things)
- Request tied with payload to automate memory management
- · Range-enabled MPI calls



What you stand to gain:

- · Communicator ownership semantics:
 - Own
 - View
 - Duplicate
- Comm buffer type deduction no more void*
- · RAII for MPI requests (and other things)
- Request tied with payload to automate memory management
- · Range-enabled MPI calls

What you stand to lose:



What you stand to gain:

- · Communicator ownership semantics:
 - Own
 - View
 - Duplicate
- Comm buffer type deduction no more void*
- RAII for MPI requests (and other things)
- Request tied with payload to automate memory management
- · Range-enabled MPI calls

What you stand to lose:

· API surface



Range:

- represents an iterable sequence
- exposes iterator/sentinel pair via .begin() / .end()



Range:

- represents an iterable sequence
- exposes iterator/sentinel pair via .begin() / .end()

View:

- view over part or all of a range
- $\boldsymbol{\cdot}$ possibly with additional transformations applied on top
- standard views are lazily evaluated
- composable via pipeline operator



Range:

- · represents an iterable sequence
- exposes iterator/sentinel pair via .begin() / .end()

View:

- · view over part or all of a range
- possibly with additional transformations applied on top
- · standard views are lazily evaluated
- composable via pipeline operator

Borrowed range:

- · range which does not own the underlying sequence
- · iterators to a borrowed range can safely outlive their parent range
- · a range type can be queried to check whether it's borrowed



Range:

- · represents an iterable sequence
- exposes iterator/sentinel pair via .begin() / .end()

View:

- · view over part or all of a range
- possibly with additional transformations applied on top
- · standard views are lazily evaluated
- composable via pipeline operator

Borrowed range:

- range which does not own the underlying sequence
- · iterators to a borrowed range can safely outlive their parent range
- \cdot a range type can be queried to check whether it's borrowed

Ranges facilities are transparent to the compiler – no run-time penalty



Range:

- represents an iterable sequence
- exposes iterator/sentinel pair via .begin() / .end()

View:

- · view over part or all of a range
- possibly with additional transformations applied on top
- standard views are lazily evaluated
- composable via pipeline operator

Borrowed range:

- range which does not own the underlying sequence
- · iterators to a borrowed range can safely outlive their parent range
- · a range type can be queried to check whether it's borrowed

Ranges facilities are transparent to the compiler – no run-time penalty

Slideware note: namespace sr = std::ranges, namespace sv = std::views



```
1 std::vector v{1, 3, 314, 7, 9, 42};
2 sr::sort(v);
3 auto even = v | sv::filter([](int i) { return i%2==0; });
4 std::print("Sorted even elements: {}\n", even);
```





```
1 std::vector v{1, 3, 314, 7, 9, 42};
2 sr::sort(v);
3 auto even = v | sv::filter([](int i) { return i%2==0; });
4 std::print("Sorted even elements: {}\n", even);
```



Sorted even elements: [42, 314]



```
1 std::vector v{1, 3, 314, 7, 9, 42};
2 sr::sort(v);
auto even = v | sv::filter([](int i) { return i%2==0; });
4 std::print("Sorted even elements: {}\n", even);
 Sorted even elements: [42, 314]
auto maxit = sr::max_element(std::vector{1, 2, 3, 4, 5});
2 std::print("Max element: {}\n", *maxit);
```





```
1 std::vector v{1, 3, 314, 7, 9, 42};
2 sr::sort(v);
auto even = v | sv::filter([](int i) { return i%2==0; });
4 std::print("Sorted even elements: {}\n", even);
 Sorted even elements: [42, 314]
auto maxit = sr::max_element(std::vector{1, 2, 3, 4, 5});
2 std::print("Max element: {}\n", *maxit);
```



```
error: no match for 'operator*'
(operand type is 'std::ranges::dangling')
```

Range-enabled MPI wrapper



The arguments of various MPI calls get simplified:

- \cdot Communicator \longrightarrow base object
- (pointer, size, type) \longrightarrow range
- status \longrightarrow ignored(?)
- \cdot request \longrightarrow returned from call

Range-enabled MPI wrapper



The arguments of various MPI calls get simplified:

- \cdot Communicator \longrightarrow base object
- (pointer, size, type) \longrightarrow range
- status → ignored(?)
- \cdot request \longrightarrow returned from call

Passing an owning range to non-blocking calls:

- Compiler error, or
- Payload gets bundled with the request

Range-enabled MPI wrapper



The arguments of various MPI calls get simplified:

- Communicator → base object
- (pointer, size, type) \longrightarrow range
- status → ignored(?)
- \cdot request \longrightarrow returned from call

Passing an owning range to non-blocking calls:

- · Compiler error, or
- · Payload gets bundled with the request

Error handling via exceptions or std::expected



Inputs:

- std::vector of in-neighbors (ranks)
- std::flat_map of out-neighbors (ranks) to corresponding message – vector of doubles



Inputs:

- std::vector of in-neighbors (ranks)
- std::flat_map of out-neighbors (ranks) to corresponding message – vector of doubles

Algorithm:

- 1. Post nonblocking sends
- 2. Probe for incoming message sizes
- 3. Gather incoming messages into single std::vector





```
std::vector<double> halo(MPI Comm
                                                                           comm.
                           const std::vector<int>&
                                                                           in nbrs.
                            const std::flat map<int, std::vector<double>>& out) {
    std::vector<MPI_Request> reqs;
    send(comm, regs, out);
    const auto in sizes = probe(comm, in nbrs);
    auto
               retval = receive(comm, regs, in_nbrs, in_sizes);
    MPI Waitall(static cast<int>(regs.size()), regs, MPI STATUSES IGNORE);
    return retval;
10 }
  std::vector<double> halo(const MPIComm&
                                                                           comm,
                            const std::vector<int>&
                                                                           in nbrs.
                            const std::flat_map<int, std::vector<double>>& out) {
    std::vector<MPIComm::Request> reas:
    send(comm, regs, out);
    const auto in_sizes = probe(comm, in_nbrs);
                retval = receive(comm, regs, in nbrs, in sizes):
    auto
    // Waitall(regs);
   return retval;
10 }
```





```
void send(MPI Comm
                                                            comm,
            std::vector<MPI Request>&
                                                            regs,
            const std::flat_map<int, std::vector<double>>& out) {
4
    for(const auto& [id, data] : out) {
      MPI Request req;
      MPI_Isend(data.data(), data.size(), MPI_DOUBLE, id, 0, comm, &req);
      regs.push back(reg);
9
  void send(const MPIComm&
                                                            comm,
            std::vector<MPIComm::Request>&
                                                            regs,
            const std::flat map<int, std::vector<double>>& out) {
    for(const auto& [id, data] : out) {
      reqs.push_back(comm.Isend(data, id, 0));
```



```
void send(MPI Comm
                                                            comm,
            std::vector<MPI Request>&
                                                            reas.
            const std::flat_map<int, std::vector<double>>& out) {
    for(const auto& [id, data] : out) {
4
      MPI Request req;
      MPI Isend(data.data(), data.size(), MPI DOUBLE, id. 0, comm. &reg):
      regs.push back(reg);
  void send(const MPIComm&
                                                            comm,
            std::vector<MPIComm::Request>&
                                                            regs,
            const std::flat map<int, std::vector<double>>& out) {
    for(const auto& [id, data] : out) {
      regs.push back(comm.Isend(data. id. 0)):
  void send(const MPIComm&
                                                            comm.
            std::vector<MPIComm::Request>&
                                                            reas.
            const std::flat map<int, std::vector<double>>& out) {
4
    sr::transform(out, std::back inserter(regs), [8](const auto& pair){
      return comm. Isend(pair.second, pair.first, 0):
    });
```









```
1 void receive(MPI Comm
                                          comm,
                std::vector<MPI Request>& regs,
                const std::vector<int>&
                                         in nbrs.
4
                                         in sizes) {
                const std::vector<int>&
     std::vector<double> retval(std::reduce(in sizes.begin(), in sizes.end()));
     for(int i = 0. offset = 0: i != in nbrs.size(): ++i) {
      MPI Request req:
      MPI Irecv(retval.data()+offset, in sizes[i], MPI DOUBLE, in nbrs[i], 0, comm, &req);
      reas.push back(rea):
      offset += sz:
     return retval;
13 }
```

Physics injection &

Passing compile-time parameters



What we mean:

- Equations
- Boundary conditions
- · Simulation structure, if impactful at compile-time
- Dimensions, if they meaningfully affect computation (e.g. stencils)



What we mean:

- Equations
- Boundary conditions
- · Simulation structure, if impactful at compile-time
- Dimensions, if they meaningfully affect computation (e.g. stencils)

What we don't mean:

- · Physical parameters, e.g., viscosity, heat capacity, ...
- Simulation structure, if no compile-time optimization occurs



2 distinct approaches:



2 distinct approaches:

- · Code generation:
 - Flexibility
 - · Requires auxiliary tooling
 - · Hostile to IDEs
 - e.g.: TCLB



2 distinct approaches:

- · Code generation:
 - Flexibility
 - Requires auxiliary tooling
 - Hostile to IDEs
 - · e.g.: TCLB
- · Meta-programming:
 - · Self-contained
 - Transparent to the IDE
 - · Requires some knowledge of C++ from the user
 - Templates
 - · More templates!
 - Unless...?



We've always had aggregate initialization:

```
struct Params {
  int nx = 1, ny = 1, nz = 1;
  double nu = 1e-3;
  int n_rhs = 1, n_depends = 0;
};
Params p{2, 4, 1, 1e-3, 1, 2};
```



We've always had aggregate initialization:

```
struct Params {
  int nx = 1, ny = 1, nz = 1;
  double nu = 1e-3;
  int n_rhs = 1, n_depends = 0;
};
Params p{2, 4, 1, 1e-3, 1, 2};
```

But now we have designated initialization:

```
Params p{.nx = 2, .ny = 4, .n_depends = 2};
```

- · Reader friendly
- No need to touch sensible defaults

Compile-time parameter structs



We can now use any literal class type as a template parameter. Combined with the following

std::integral_constant-esque utility

```
1 template<auto>
2 struct ConstexprWrapper {};
```



We can now use any literal class type as a template parameter. Combined with the following

std::integral_constant-esque utility

```
template<auto>
struct ConstexprWrapper {};
```

it gives us a beautiful, clear, and powerful interface for passing CT parameters:

```
constexpr auto params = Params {2, 4, 1, 1e-3, 1, 2};
/* 1 */ auto simulation = makeSim(ConstexprWrapper<params>{});
/* 2 */ auto simulation = makeSim(WRAP_CT_PARAM(params));
/* 3 */ auto simulation = makeSim<params>();
```



But we're not limited to simple structs:

25

Physics injection



Since we know how to elegantly pass compile-time parameters, we know how to set the size of operators, the discretization order, etc.

example from L3STER



Since we know how to elegantly pass compile-time parameters, we know how to set the size of operators, the discretization order, etc.

example from L3STER

The question remains – how do we provide the bodies of these operators?



Since we know how to elegantly pass compile-time parameters, we know how to set the size of operators, the discretization order, etc.

example from L3STER

The question remains – how do we provide the bodies of these operators?

Idea: rely on generic lambdas and structured bindings to provide a type-oblivious interface.

Physics injection



```
1 const auto diffusion3d_kernel = wrapDomainEquationKernel<diff3_params>(
     [](const auto& in. auto& out) {
          auto& [operators, rhs] = out;
          auto& [A0, Ax, Ay, Az] = operators;
4
           constexpr double k = 1.; // diffusivity
          constexpr double s = 1.; // source
          Ax(0.1) = -k:
          Av(0, 2) = -k:
          Az(0, 3) = -k;
           rhs[0] = s:
          A0(1.1) = -1.:
          Ax(1, 0) = 1.;
          A0(2, 2) = -1.;
14
          Av(2.0) = 1.:
          A0(3, 3) = -1.;
16
          Az(3, 0) = 1.;
          Av(4.3) = 1.:
          Az(4, 2) = -1.;
          Ax(5, 3) = -1.;
          Az(5.1) = 1.:
          Ax(6.2) = 1.:
          Av(6, 1) = -1.;
     }):
  sys->assembleProblem(diffusion3d_kernel, std::views::single(solid_id));
```

Conclusions



• Today's C++ is not the same language you held your nose at 20 years ago



- Today's C++ is not the same language you held your nose at 20 years ago
- It's (relatively) easy to pass compile-time information to an API and to the compiler



- Today's C++ is not the same language you held your nose at 20 years ago
- It's (relatively) easy to pass compile-time information to an API and to the compiler
- Most of the complexity can be hidden from the user



- Today's C++ is not the same language you held your nose at 20 years ago
- It's (relatively) easy to pass compile-time information to an API and to the compiler
- · Most of the complexity can be hidden from the user



Discussion

jakub.galecki.dokt@pw.edu.pl