List of definitions used in the pseudocode

Sílvia Casacuberta, Grace Tian, and Connor Wagaman

Summer 2021
Version as of September 6, 2021 (UTC)

We use the following guideline: if a term appears in the `\texttt{...}` LaTeXfont, then this term is defined in the pseudocode definitions document. Otherwise, it appears in the proof definitions document.

We provide the terms in alphabetical order within each section. "TODOs" should be included at the end of the corresponding section. On the other hand, "TODOs" which better specify an already-defined term should be included immediately following the definition of that term. Examples should never be part of the definition, but we encourage their use right after the definition of a term.

We also recommend linking to the Rust Standard Library when the term is defined there.

Contents

0.1 List of terms that have not yet been added

- The compatibility pairing and subdomains. Compatibility inheritance for subdomains and proof for subdomain traits. But this is awaiting the specific final implementation.
- Clarify why we specify vectors for `SizedDomain`
- Add definition of a valid measurement
- Double-check newly added definitions, and trait vs function (e.g., `SaturatingAdd` and `saturating_add`)
- Incorporate corrections from GitHub
- Add type trait `Laplace Domain`, and the other ones from base Laplace: `D::Atom`, `SampleLaplace`, `Float`, and `InfCast(D::Atom)`.
- Add function `is_sign_negative()` from Laplace
- Add function `.recip()` from Laplace
- Add function `sample_laplace` (idealized, for now)
- Add `IntDistance`, although Mike said to keep using `u32` in the pseudocode type signing for now.
- Definition of the Rust iterator
- Do we need `fold`? Maybe always use loops

0.2 Versions of definitions documents

When looking for definitions for terms that appear in this document, the following versions of the definitions documents should be used.

- **Proof definitions document:** This file uses the version of the proof definitions document available as of September 6, 2021, which can be found at this link (archived here).

1 Types

Note 1.1 (A note on `()` vs. `[]`)**.** Parentheses, `()`, are used to create an instance of a domain. Square brackets, `[]`, are used to describe the type of a domain.

For example, `AllDomain(i8)` is the domain of all values of type `i8`. However, the type of `AllDomain(i8)` – the domain itself, not the elements of the domain – is `AllDomain[i8]`; note the square brackets. This typing style is inspired by the notation used in Python; see https://docs.python.org/3/library/typing.html.

Definition 1.2 (`bool`)**.** The type `bool` represents a value which can only be either `True` or `False`. If a `bool` is casted to an integer, `True` will be `1` and `False` will be `0`.

Definition 1.3 (`::Carrier`)**.** `SomeDomain::Carrier` is the type of a member in `SomeDomain`, where `SomeDomain` is a domain. See section 2 for more information on domains.

For example, `AllDomain(T)::Carrier` is `T`.

Definition 1.4 (`f32`)**.** `f32` is the Rust 32-bit floating point type.
See https://doc.rust-lang.org/std/primitive.f32.html.

Definition 1.5 (`f64`)**.** `f64` is the Rust 64-bit floating point type.
See https://doc.rust-lang.org/std/primitive.f64.html.

TODO (future – not enough info yet): Add / pointers to "binary64" type defined in IEEE 754-2008.

Definition 1.6 (`::Imputed`)**.** `SomeDomain::Imputed` is the type of `SomeDomain` after imputation, where `SomeDomain` is a domain.

Note however, that `SomeDomain::Imputed` may contain a null value. See the proof section of proof section of `impute_constant`.

Definition 1.7 (`IntDistance`)**.** `IntDistance` is equivalent to `u32`.

Definition 1.8 (`isize`)**.** `isize` is defined differently on 32-bit and 64-bit machines. This is because the size of this primitive is equal to the number of bytes it takes to reference any location in memory.

- **32-bit machines:** for `v` a value of type `usize`, `v` $\in \{-2^{31}, \ldots, -1, 0, 1, \ldots, 2^{32} - 1\}$
- **64-bit machines:** for `v` is a value of type `usize`, `v` $\in \{-2^{63}, \ldots, -1, 0, 1, \ldots, 2^{63} - 1\}$

See https://doc.rust-lang.org/std/primitive.isize.html.

Definition 1.9 (`Option`)**.** The type `Option[T]` is the type for all values of type `T`, as well as the type for the values `None` and `Some`, where `Some(val:T)` is equal to `val`.
See https://doc.rust-lang.org/std/option/enum.Option.html.

Definition 1.10 (`u32`)**.** `u32` is the Rust 32-bit unsigned integer type. If v is a value of type `u32`, then we know that $v \in \{0, 1, 2, \ldots, 2^{32} - 1\}$.
See https://doc.rust-lang.org/std/primitive.u32.html.

Definition 1.11 (`usize`)**.** `usize` is defined differently on 32-bit and 64-bit machines. This is because the size of this primitive is equal to the number of bytes it takes to reference any location in memory.

- **32-bit machines:** if v is a value of type `usize`, then $\texttt{v} \in \{0, 1, 2, \ldots, 2^{32} - 1\}$
- **64-bit machines:** if v is a value of type `usize`, then $\texttt{v} \in \{0, 1, 2, \ldots, 2^{64} - 1\}$

See https://doc.rust-lang.org/std/primitive.usize.html.

Definition 1.12 (`Vec[T]`). The Rust type `Vec[T]` consists of ordered lists of type `T`. For example, if `T = bool`, then values of type `Vec[T]` include `[], [0], [1], [0, 0], ...`. A vector can contain no more than `get_max_value(usize)` elements.

See https://doc.rust-lang.org/std/vec/struct.Vec.html.

1.1 Notes, todos, questions

Question for reviewers: Should we have a general definition for "floats" (and "integers"?), or is it sufficiently understood what a float is?

2 Domains

A *data domain* is a representation of the set of values on which the function associated with a transformation or measurement can operate. Each metric (see section 7) is associated with certain data domains. Types used for implementing domains in OpenDP have trait `Domain` (defined in definition 3.5).

Definition 2.1 (`AllDomain`). `AllDomain(T)` is the domain of all values of type `T`. This domain has type `AllDomain[T]`.

For example, `AllDomain(u32)` is the domain of all values of type `u32`.

Definition 2.2 (`IntervalDomain`). For any type `T` with trait `TotalOrd` (see definition 3.19),[1] `IntervalDomain(L:T, U:T)` is the domain of all values `v` of type `T` such that `L <= v` and `v <= U`, for a type `T` that has a total ordering (`T` has trait `TotalOrd`) and for values `L <= U` of type `T`. This domain has type `IntervalDomain[T]`.

An important remark is that the Rust implementation of `IntervalDomain` checks that `L <= U`, and returns an error if `L > U`. Therefore, any transformation or measurement that uses `IntervalDomain` does not need to re-check this constraint and raise a possible exception for it.

Note that, because both `L` and `U` are of type `T`, there is no need to explicitly pass `T`; the type `T` can be inferred. `IntervalDomain` is defined on any type that implements the trait `TotalOrd`. For example, `IntervalDomain(1:u32, 17:u32)` corresponds to a domain that contains all the `u32` values `v` such that `1 <= v` and `v <= 17`; it has type `IntervalDomain[u32]`.

Definition 2.3 (`InherentNullDomain`). `InherentNullDomain(inner_domain:D)` is the domain of all values of data domain `inner_domain` unioned with a `null` value. Note that this means that a domain may have only *one* `null` value; as a result, applying `InherentNullDomain` to a domain already containing a `null` value will not affect the domain. This domain has type `InherentNullDomain[D]`.

[1] As of June 28, the OpenDP library requires the weaker condition of partial ordering (implements `PartialOrd`) instead.

Definition 2.4 (`SizedDomain`). `SizedDomain(inner_domain:D, n:usize)` is the domain of all elements from domain `D` restricted to length `n`. This domain has type `SizedDomain[D]`.

For example, `SizedDomain(VectorDomain(AllDomain(u32)), n)` is the domain of all vectors of length `n` with elements of type `u32`.

Definition 2.5 (`VectorDomain`). `VectorDomain(inner_domain:D)` is the domain of all vectors of elements drawn from domain `inner_domain`. This domain has type `VectorDomain[D]`.

2.1 Subdomains

2.2 Notes, todos, questions

TODO (future – not enough info yet): As of July 20, OpenDP plans to include subdomains (see the Architecture meeting notes for 20/7). We have to include them and prove that they are indeed subdomains. Then in the metric definition it is enough to list the most general domain, since the domain-metric compatibility in inherited. We will add the necessary information here after Mike and Andy have finished the implementation details.

Preliminary theorems:

Theorem 2.6 (Domain-metric compatibility inheritance.). *Given a domain* `D`*, for any subdomain* `S` $\subseteq$ `D`*, if* `D` *is compatible with metric* `M` *then* `S` *is compatible with metric* `M`*.*

3 Traits

Definition 3.1 (`Abs`). A type `T` has trait `Abs` if and only if the absolute value of a value of type `T` can be taken.

Definition 3.2 (`Bounded`). A type `T` has trait `Bounded` if and only if `T` has some upper bound and some lower bound (some smallest possible value and some largest possible value).

Definition 3.3 (`CheckedMul`). A type `T` has trait `CheckedMul` if it performs multiplication that returns `None` if overflowing.

Definition 3.4 (`DistanceConstant`). A type `TO` has trait `DistanceConstant(TI)` if:

- `TO` has trait `Mul(Output=TO)` (multiplication can be done with type `TO`)
- `TO` has trait `Div(Output=TO)` (some form of inverse mapping can be done with type `TO`)
- `TO` has trait `PartialOrd` (`TO` has a partial ordering)
- `TO` has trait `InfCast(TI)`

In OpenDP (Rust), this is called `DistanceConstant`. See `https://github.com/opendp/opendp/blob/main/rust/opendp/src/traits.rs`.

Definition 3.5 (`Domain`)**.** A type `T` has trait `Domain` if and only if it can represent a set of values that make up a domain. The `Domain` implementation prescribes a type for members of the domain, as well as a method to check if any instance of that type is a member of that domain.

Definition 3.6 (`ExactIntCast`)**.** A type `TO` has trait `ExactIntCast(TI)` if and only if Every value of type `TI` can be `exact_int_cast`ed exactly to a value of type `TO`, as long as the original value of type `TI` is no smaller than `get_min_consecutive_int(TO)` and no larger than `get_max_consecutive_int(TO)`.

A cast error is returned when the value being `exact_int_cast`ed is greater than `get_max_consecutive_int(TO)` or less than `get_min_consecutive_int(TO)`.

Definition 3.7 (`Float`)**.** Generic trait for floating point numbers. A type `T` with trait `Float` automatically implements trait `Div<Self, Output = Self>`.

Definition 3.8 (`ImputableDomain`)**.** Any domain for which the `ImputableDomain` trait is implemented for it, has

- an associated type `Imputed`. `VectorDomain::Imputed` is the data type after imputation.
- an imputation function `impute_constant` that replaces a null value with a constant or passes a non-null value through. The pseudocode is
  `def impute_constant(x: DA) -> DA::Imputed:`
  `return constant if x.is_null else x`
- a function `is_null` to check if a value is `null`
- a function `new` to construct an instance of the domain

Definition 3.9 (`InfCast`)**.** A type `TO` has trait `InfCast(TI)` if and only if for every value `val` of type `TI`, `inf_cast(val:TI,TO)` is defined (note in definition 4.16 for `inf_cast` that it is acceptable for `inf_cast(val:TI,TO)` to return an error).

Note: the name "InfCast" comes from the idea of rounding the result of the cast toward positive ***inf****inity if needed.*

Definition 3.10 (`InherentNull`)**.** A type `T` has trait `InherentNull` if and only if type `T` can hold some value `null`.

As of July 16, 2021, only `f32` and `f64` have the trait `InherentNull`.

Definition 3.11 (`MaxConsecutiveInt`)**.** A type `T` has trait `MaxConsecutiveInt` if and only if there is some maximum integer `i` such that all integers from 0 up to `i` (inclusive) can be expressed as a value of type `T`; but such that the next integer that can be expressed by `T` is not `i` + 1.

Definition 3.12 (`MinConsecutiveInt`)**.** A type `T` has trait `MinConsecutiveInt` if and only if there is some minimum integer `i` such that all integers from 0 up to `i` (inclusive) can be expressed as a value of type `T`; but such that the next integer that can be expressed by `T` is not `i` + 1.

Definition 3.13 (`Metric`)**.** A type `T` has trait `Metric` if and only if it can represent a metric for quantifying distances between values in a set. The `Metric` implementation additionally prescribes the type to use for representing distances.

Definition 3.14 (`One`)**.** A type `T` has trait `One` if and only if `T` has some multiplicative identity element. A type `T` with trait `One` automatically implements trait `Mul(Output = T)`.

Definition 3.15 (`OptionNull`)**.** A type `Option[T]` has trait `OptionNull` if and only if `null` can be represented as a value of type `Option[T]`.

Definition 3.16 (`PartialEq`)**.** A type `T` has trait `PartialEq` if and only if values of type `T` have a partial equivalence relation defined on them. A relation R is a partial equivalence relation if and only if for all `a,b,c` of type `T`, we have:

- **symmetry:** if `a`R`b`, then `b`R`a`.
- **transitivity:** if `a`R`b` and `b`R`c`, then `a`R`c`

See https://doc.rust-lang.org/std/cmp/trait.PartialEq.html.

Definition 3.17 (`PartialOrd`)**.** A type `T` has trait `PartialOrd` if for all elements `a, b, c` of type `T`, the following properties are satisfied:

1. **asymmetry**: if `a < b` then `not(a > b)`; likewise, if `a > b` then `not(a < b)`

   Question for reviewers: Is the meaning of "`not`" above clear? In math, the first statement above would be written as "if $a < b$ then $a \not> b$".

2. **transitivity**: if `a < b` and `b < c`, then `a < c`; the same holds for `==` and `>`.

See https://doc.rust-lang.org/std/cmp/trait.PartialOrd.html.

Definition 3.18 (`SaturatingAdd`)**.** A type `T` has trait `SaturatingAdd` if the function `saturating_add` (defined in definition 4.25) can be called on values of type `T`.

Definition 3.19 (`TotalOrd`)**.** A type `T` has trait `TotalOrd` if and only if `T` has trait `PartialOrd` and moreover all elements are comparable; that is, for all elements a, b of type `T`, either $a \leq b$ or $b \leq a$.

Definition 3.20 (`Zero`)**.** A type `T` has trait `Zero` if and only if `T` has some additive identity element. A type `T` with trait `Zero` automatically implements trait `Add(Output = T)`.

3.1 Math-related definitions

Definition 3.21 (`Add(Output=T)`)**.** A type `T` has trait `Add(Output=T)` if and only if addition can be performed between elements of type `T`, with the result of the addition also being of type `T`.

See https://doc.rust-lang.org/std/ops/trait.Add.html.

Definition 3.22 (`Div(Output=T)`)**.** A type `T` has trait `Div(Output=T)` if and only if division can be performed between elements of type `T`, with the result of the division also being of type `T`.

See https://doc.rust-lang.org/std/ops/trait.Div.html.

Definition 3.23 (`Mul(Output=T)`)**.** A type `T` has trait `Mul(Output=T)` if and only if multiplication can be performed between elements of type `T`, with the result of the multiplication also being of type `T`.

See https://doc.rust-lang.org/std/ops/trait.Mul.html.

Definition 3.24 (`Sub(Output=T)`)**.** A type `T` has trait `Sub(Output=T)` if and only if subtraction can be performed between elements of type `T`, with the result of the subtraction also being of type `T`.

See https://doc.rust-lang.org/std/ops/trait.Sub.html.

Definition 3.25 (`Sum(Output=T)`)**.** A type `T` has trait `Sum(Output=T)` if and only if such type can be created by summing up an iterator. This trait is used to allow the `sum` function on iterators to be used. Types which implement the trait can be generated by the `sum()` method.

See https://doc.rust-lang.org/std/iter/trait.Sum.html.

3.2 Traits that need not appear in the preconditions

- `'static`. Notes: `'static` is not a type; it is a lifetime name (this is a Rust definition).
- `Clone`
- `Copy`

3.3 Notes, todos, questions

4 Functions

4.1 Functions in the pseudocode language

Definition 4.1 (`abs`)**.** Given an element `var` of type `T`, where `T` must have trait `Abs`, the function `abs` reuturns the absolute value of `var`.

Definition 4.2 (`assert`)**.** The function `assert` is followed by an expression. If `some_expression` evaluates to `False`, then `assert some_expression` results in an error that prevents the code from proceeding further.

In Python, this is called `assert`. See https://docs.python.org/3/reference/simple_stmts.html#the-assert-statement.

Definition 4.3 (`can_cast`)**.** The function `can_cast(type1,type2)` returns `True` if and only if no data would be lost by casting from `type1` to `type2`. In other words, it returns `True` if and only if there is an injection from `type1` to `type2`.

See https://doc.rust-lang.org/std/convert/trait.TryFrom.html.

For example, `can_cast(u32,u64)` will return `True` because a `u32` can always be expressed as a `u64`; conversely, `can_cast(u64,u32)` will return `False` because a `u64` could be too big to be expressed as a `u32`, and then data would be lost.

Definition 4.4 (`cast`)**.** `cast(val:TI, TO)` converts `val` of type `TI` to the corresponding `val` of type `TO`, and returns `val` of type `TO`. Returns an error if the conversion is unsuccessful.

Definition 4.5 (`checked_mul`)**.** Given two elements `var1, var2` of type `T` with trait `Mul(Output=T)`, `checked.mul(var1, var2)` returns `var1*var2` if the result does not overflow, and else returns `None`.

Definition 4.6 (`exact_int_cast`)**.** This function only works for types `TO` that have trait `ExactIntCast(TI)`. For any given `val` such that `val` is between `get_min_consecutive_int(TO)` and `get_max_consecutive_int(TO)`, then `exact_int_cast(val:TI,TO)` returns an integer value of type `TO` equal to the integer value held by `val` (which was of type `TI`); otherwise, a cast error is returned.

Definition 4.7 (`get_input_domain`)**.** The function `get_input_domain(function)` returns the input domain of arguments passed to function `function`.

Definition 4.8 (`get_input_metric`)**.** The function `get_input_metric(some_relation)` returns the input metric used by the relation `some_relation`.

Definition 4.9 (`get_max_consecutive_int`)**.** This function is only defined on types `T` that have trait `ExactIntCast`. The function `get_max_consecutive_int(T)` returns the greatest integer $\texttt{i} \geq 0$ such that all integers from 0 up to `i` (inclusive) can be expressed exactly as a value of type `T`; but such that $\texttt{i} + 1$ cannot be expressed exactly as a value of type `T`. The return value is of type `T`.

Definition 4.10 (`get_max_value`)**.** This function is only defined on types `T` that have a total ordering. The function `get_max_value(T)` returns the maximum value that can be expressed by an object of type `T`. The return value is of type `T`.

Definition 4.11 (`get_min_consecutive_int`)**.** This function is only defined on types `T` that have trait `ExactIntCast`. The function `get_min_consecutive_int(T)` returns the smallest integer $\texttt{i} \leq 0$ such that all integers from 0 down to `i` (inclusive) can be expressed exactly as a value of type `T`; but such that $\texttt{i} - 1$ cannot be expressed exactly as a value of type `T`. The return value is of type `T`.

Definition 4.12 (`get_min_value`)**.** This function is only defined on types `T` that have a total ordering. The function `get_min_value(T)` returns the minimum value that can be expressed by an object of type `T`. The return value is of type `T`.

Definition 4.13 (`get_output_domain`)**.** The function `get_output_domain(function)` returns the output domain of values returned by function `function`.

Definition 4.14 (`get_output_metric`)**.** The function `get_output_metric(some_relation)` returns the output metric used by the relation `some_relation`.

Definition 4.15 (`has_trait`)**.** The function `has_trait(T,(trait1,trait2,...))` is a function that returns `True` if and only if the type `T` implements `trait1`, `trait2`, etc.

Definition 4.16 (`inf_cast`)**.** This function is only defined for casting to types `TO` that have trait `InfCast(TI)`. The function `inf_cast(val:TI, TO)` casts `val` to a value of type `TO` and returns that value. Specifically, `val` will be casted to the value of type `TO` that is closest to `val` and at least as large as `val`. If `inf_cast` is not able to cast `val` to a value of type `TO` at least as large as `val`, then an error is returned instead.

Property: `inf_cast`ed distances are never less than input distances.

Note: the name "InfCast" comes from the idea of rounding the result of the cast toward positive ***inf****inity if needed.*

Definition 4.17 (`is_instance`)**.** The function `is_instance(var,T)` returns `True` if and only if the variable `var` is of type `T`.

Definition 4.18 (`is_none`)**.** The function `var.is_none` returns `True` if and only if `var` which is of `float` is equal to `None`.

Definition 4.19 (`is_null`)**.** The function `var.is_null` returns `True` if and only if `var` which is of type `Option[T]` is not equal to `null`.

Remark 4.20 (`ImputableDomain`)**.** To check for nullity, we use `v.is_null()` if v is a float, or `v.is_none()` if v is an `Option[T]`. To abstract these functions, we define the `ImputableDomain` trait to capture both notions of nullity.

Definition 4.21 (`len`)**.** The function `len(vector_name)` returns the number of elements (with multiplicities) in `vector vector_name`. Output is of type `usize`, so the return value v on 32-bit machines is $v \in \{0, 1, 2, \ldots, 2^{32} - 1\}$; likewise, the return value on 64-bit machines is $v \in \{0, 1, 2, \ldots, 2^{64} - 1\}$.

See https://doc.rust-lang.org/std/vec/struct.Vec.html#method.len.

Note: we do not call it `length` *to avoid notational clashes with, for example, the Bounded Sum code.*

Definition 4.22 (`map`)**.** A `map` applies a given function to all the items in an iterable without using an explicit for loop. Hence, `map(f, iter)` is an iterator that maps the values of `iter` with `f`.

See https://doc.rust-lang.org/std/iter/struct.Map.html.

Definition 4.23 (`max`)**.** The function `max(var1, var2)` compares `var1` and `var2`, and returns the greater of the two values. When `var1` and `var2` are equivalent, it returns `var2`. The return type of `map` is also an iterator. The function `max` requires that `var1` and `var2` have trait `TotalOrd`.

See https://doc.rust-lang.org/std/cmp/fn.max.html.

Definition 4.24 (`min`)**.** The function `min(var1:T, var2:T)` compares `var1` and `var2`, and returns the lesser of the two values. When `var1` and `var2` are equivalent, it returns `var1`. The function `min` requires that `var1` and `var2` have trait `TotalOrd`.

See https://doc.rust-lang.org/std/cmp/fn.min.html.

Definition 4.25 (`saturating_add`)**.** Given two elements `var1, var2` of type `T` with trait `Add(Output=T)`, the function `saturating_add(var1, var2)` returns `var1+var2` whenever `var1+var2` $\in$ `[get_min_value(T), get_max_value(T)]`. If `var1+var2` $<$ `get_min_value(T)`, then `saturating_add(var1, var2)` $:=$ `get_min_value(T)`. Similarly, in the case where `var1+var2` $>$ `get_max_value(T)`, then `saturating_add(var1, var2)` $:=$ `get_max_value(T)`.

See https://doc.rust-lang.org/std/intrinsics/fn.saturating_add.html.

Definition 4.26 (`sum`)**.** The `sum` function adds all terms in an iterable incrementally and returns their sum.

This is done by first summing the first two terms and calculating an intermediate result with the same type as the input type (with rounding if applicable). Then the third term is summed with this result and a new intermediate result is calculated. This process continues until the most recent intermediate result is summed with the final term in the iterable, and the result, with the same type as the input type, is returned.

4.2 Notes, todos, questions

5 Data structures

Definition 5.1 (`list`)**.** A `list` is a data structure which is a changeable ordered sequence of elements.

Importantly, we remark that in some occasions in our Python-like pseudocodes we will write `list` as an equivalent for the Rust `Vec` in order to maintain a Python-like notation. For this reason, such a `list` will have type `Vec(T)` and be considered an element of `VectorDomain`. We will allow the use of the Rust-like term `Vec` when type signing the functions in the pseudocode and proving the corresponding domain properties in the proof.

6 Classes

Definition 6.1 (`Transformation`)**.** We define a `Transformation` in the following way.

Question for reviewers: Which pseudocode style is preferred for this definition? With preconditions (section 6.1) or without preconditions (section 6.2)?

6.1 Pseudocode with preconditions

- `input_domain` must have trait `Domain`
- `output_domain` must have trait `Domain`
- `function` must operate on inputs from `input_domain`, and it must produce outputs in `output_domain`
- `input_metric` must have trait `Metric`
- `output_metric` must have trait `Metric`
- `stability_relation` must operate on input metrics equal to `input_metric`, and it must operate on output metrics equal to `output_metric`

```
class Transformation:
    def __init__(self, input_domain, output_domain, function, input_metric,
     output_metric, stability_relation):

        self.input_domain = input_domain
        self.output_domain = output_domain

        self.function = function

        self.input_metric = input_metric
        self.output_metric = output_metric

        self.stability_relation = stability_relation
```

6.2 Pseudocode without preconditions

```python
class Transformation:
    def __init__(self, input_domain, output_domain, function, input_metric,
     output_metric, stability_relation):

        assert has_trait(input_domain, Domain)
        self.input_domain = input_domain
        assert has_trait(output_domain, Domain)
        self.output_domain = output_domain

        assert get_input_domain(function) == input_domain
        assert get_output_domain(function) == output_domain
        self.function = function

        assert has_trait(input_metric, Metric)
        self.input_metric = input_metric
        assert has_trait(output_metric, Metric)
        self.output_metric = output_metric

        assert get_input_metric(stability_relation) == input_metric
        assert get_output_metric(stability_relation) == output_metric
        self.stability_relation = stability_relation
```

In OpenDP (Rust), this is called `Transformation`. See https://github.com/opendp/opendp/blob/35dbdc73d7d74e049f5101a704d4e036bed365e8/rust/opendp/src/core.rs#L369-L376. When we refer to a *valid transformation* in the proofs, this is the precise definition.

Therefore, there is no need to include the following code snippet in all of the pseudocodes:

```python
class Transformation:
    input_domain
    output_domain
    function
    input_metric
    output_metric
    stability_relation
```

TODO (future – not enough info yet): Add the equivalent pseudocode for Measurements. (Should be very similar; just change names.)

7 Metrics

7.1 Dataset metrics

Metrics are used to measure the distances between data. Metrics have a *domain* on which the function associated with the metric can measure distance, and an *associated type* that determines the type used to represent the distance between datasets.

Example: `SymmetricDistance` has a domain of `VectorDomain(AllDomain(T))`, which means that `SymmetricDistance` can be used to measure the distance between any objects that are vectors of elements of type `T`. `SymmetricDistance` has an associated type of `u32`, which means that a `u32` value is used to report the distance.

Please refer to the proof definitions document (see section 0.2) for definitions of the functions referred to by the metrics below.

Note 7.1. Once the subdomains have been implemented in OpenDP, it will no longer be necessary to list all of the subdomains in the compatible domains section.

Definition 7.2 (`AbsoluteDistance(T)`)**.** The definition of *absolute distance* in the proof definitions document (see section 0.2) tells how the distance between data is calculated.

- **Domain:** `AllDomain(T)`, where `T` has the traits `Sub(Output=T)` (defined in definition 3.24) and `TotalOrd` (defined in definition 3.19).
- **Associated type:** `Q`.
- ***d*-close:** For any two elements n, m in `AllDomain(T)`, where `T` denotes an arbitrary type with trait `Sub(Output=T)`, and d of generic type `Q`, we say that n, m are d-close under the absolute distance metric (abbreviated as d_{Abs}) whenever

$$d_{Abs}(n, m) = |n - m| \leq d.$$

Definition 7.3 (`SymmetricDistance`)**.** The definition of *symmetric distance* in the proof definitions document (see section 0.2) tells how the distance between data is calculated.

- **Domains:** `VectorDomain(inner_domain)` and `SizedDomain(VectorDomain(inner_domain))`, where `inner_domain` is any domain.
- **Associated type:** `IntDistance`.
- ***d*-close:** For any two vectors $u, v \in$ `VectorDomain(D)` and any d of type `u32`, we say that u, v are d-close under the symmetric distance metric (abbreviated as d_{Sym}) whenever

$$d_{Sym}(u, v) = |\text{MultiSets}(u)\Delta\text{MultiSets}(v)| \leq d.$$

Note: the associated type of `SymmetricDistance` is hard-coded as `IntDistance`, so when declaring that the metric being used is `SymmetricDistance`, we only need to write `metric = SymmetricDistance()`; by contrast, we need to write `AbsoluteDistance(T)` where `T` is the type on which we are taking the absolute distance since the associated type for `AbsoluteDistance` is not hard-coded.

7.2 Sensitivity metrics

Note 7.4. Sensitivity metrics are used for measurements (rather than transformations), so our most finalized proofs – which all deal with transformations – do not require correctness of the definitions in this section.

Definition 7.5 (`L1Distance`)**.** The definition of *L1 distance* in the proof definitions document (see section 0.2) tells how the distance between data is calculated.

Domain: `VectorDomain(inner_domain))`, where `inner_domain` is any domain.

Associated type: `Q`.

***d*-close:** For any two vectors $u, v \in$ `VectorDomain(D)` and d of generic type `Q`, we say that u, v are d-close under the L1 distance metric (abbreviated as d_{L1}) whenever

$$d_{L1}(u, v) = \sum_{i=0}^{n} |u_i - v_i| \leq d.$$

Definition 7.6 (`L2Distance`).

Definition 7.7 (`L2Distance`). The definition of *L2 distance* in the proof definitions document (see section 0.2) tells how the distance between data is calculated.

Domain: `VectorDomain(inner_domain)`, where `inner_domain` is any domain.

Associated type: `Q`.

***d*-close:** For any two vectors $u, v \in$ `VectorDomain(D)` and d of generic type `Q`, we say that u, v are d-close under the L2 distance metric (abbreviated as d_{L2}) whenever

$$d_{L2}(u, v) = \sqrt{\sum_{i=0}^{n} |u_i - v_i|^2} \leq d.$$

Question for reviewers: Are the domain and associated type correct? Are any further traits needed?

Maybe: include p-norms? (See the Wikipedia article at `https://en.wikipedia.org/wiki/Norm_(mathematics)#p-norm` for a definition of p-norms.)

7.3 Measures

Definition 7.8 (`MaxDivergence`). The definition of *max divergence* in the proof definitions document (see section 0.2) tells how the distance between data is calculated.

- **Domain:** `VectorDomain(AllDomain(T))`.

  Question for reviewers: Is there any trait required of `T`? Should it be `VectorDomain(D)` instead?

- **Associated type:** `Q`.

- ***d*-close:** For any two vectors u, v in `VectorDomain(AllDomain(T))` and any d of generic type `Q`, we say that u, v are d-close under the max divergence measure (abbreviated as D_∞) whenever

$$D_\infty(u, v) = \max_{S \subseteq \mathrm{Supp}(Y)} \left[\ln \frac{\Pr[f(u) \in S]}{\Pr[f(v) \in S]} \right] \leq d.$$

  Question for reviewers: Thoughts on this definition? Should D_∞ operate on u, v (so $D_\infty(u, v)$), or on $f(u), f(v)$ (so $D_\infty(f(u), f(v))$)?

7.4 Notes, todos, questions

TODO (future – not enough info yet): Need to learn how to cross-reference TeX files. After Prof. Vadhan's comment on 19/7, we should think of a systematic way to do dependency tracking.

Question for reviewers: With regard to the todo above, any recommendations for how to cross-reference between TeX files?

8 Code-Definitions update loop

List of items that have recently changed in the code and have not yet been implemented, or that we know will be implemented soon:

- Subdomains for domain-metric compatibility check.
- Renaming: `https://github.com/opendp/opendp/issues/181`.