

Quantitative Equational Reasoning

Prakash Panangaden¹

¹School of Computer Science
McGill University

January 2018, IRIF, Paris

Outline

- 1 Introduction
- 2 Equational reasoning
- 3 Categorical algebra
- 4 Metrics
- 5 Quantitative equations
- 6 Examples

- Equations are at the heart of mathematical reasoning.
- Reasoning about programs is also based on program equivalences.
- A trinity of ideas: Equationally given algebras, Lawvere theories, Monads on **Set**
- The dawning of the age of quantitative reasoning.
- We want quantitative analogues of algebraic reasoning.
- (Pseudo)metrics instead of equivalence relations.
- Equality indexed by a real number $=_{\epsilon}$.
- Monads on **Met**.
- Enriched Lawvere theories?

Finitary equational theories

- Signature $\Omega = \{(Op_i, n_i) \mid i = 1 \dots k\}$
- Terms $t ::= x \mid Op(t_1, \dots, t_n)$
- Equations $s = t$
- Axioms, sets of equations Ax
- Deduction $Ax \vdash s = t$
- Usual rules for deduction: equivalence relation, congruence,...
- Theories: set of equations closed under deduction.

Equational deduction rules

- Axiom $Ax \vdash s = t$ if $s = t \in Ax$
- Equivalence

$$\frac{\overline{Ax \vdash t = t}}{Ax \vdash s = t, Ax \vdash t = u} \\ Ax \vdash s = u \\ \frac{Ax \vdash s = t}{Ax \vdash t = s}$$

- Congruence

$$\frac{Ax \vdash t_1 = s_1, \dots, Ax \vdash t_n = s_n}{Ax \vdash Op(t_1, \dots, t_n) = Op(s_1, \dots, s_n)}$$

- Substitution

$$\frac{Ax \vdash t = s}{Ax \vdash t[u/x] = s[u/x]}$$

- We assume that there is one set of “basic things” – one-sorted algebras.
- Fix a set Ω of *operations*, each with a fixed arity $n \in \mathbb{N}$. These include *constants* as arity zero “operations.” Such an Ω is called a signature.
- Everything has finite arity.
- As Ω -algebra \mathcal{A} is a set A to interpret the basic sort and, for each operation f of arity n a function $f_{\mathcal{A}} : A^n \rightarrow A$.

Algebras equationally II

- Can define homomorphisms and subalgebras easily.
- What about equations that are required to hold?
- Given a set X we define the *term algebra generated by X* , TX
- The elements of X are in TX .
- If t_1, \dots, t_n are in TX and f has arity n then $f(t_1, \dots, t_n)$ is in TX .

Algebras from equations I

- Want to write things like $\forall x, y, z; f(x, f(y, z)) = f(f(x, y), z)$.
- X , set of *variables*.
- Let s, t be terms in TX , we say the equation $s = t$ *holds* in an Ω -algebra \mathcal{A} if *for every* homomorphism $h : TX \rightarrow \mathcal{A}$ we have $h(s) = h(t)$ where, in the latter, $=$ means identity.
- Let S be a set of equations between pairs of terms in TX . We define a *congruence relation* \sim_S on TX in the evident way.

Algebras from equations II

- Easy to check that if $t_1 \sim_S s_1, \dots, t_n \sim_S s_n$ then $f(t_1, \dots, t_n) \sim_S f(s_1, \dots, s_n)$ we can define f_{\sim_S} on TX / \sim_S .
- Let $[t]$ be an equivalence class of \sim_S ; $f_{\sim_S}([t_1], \dots, [t_n])$ is well defined by $[f(t_1, \dots, t_n)]$.
- A class of Ω -algebras satisfying a set of equations is called a variety of algebras (not the same as an algebraic variety!).
- When are a set of equations bad? If we can derive $x = y$ from S then the only algebras have one element.

Examples

- Monoids, groups, rings, lattices, boolean algebras are all examples.
- Vector spaces have two sorts.
- Fields are annoying because we have to say $x \neq 0$ implies x^{-1} exists. Fields do not form an equational variety.
- Sometimes we need to state conditional equations; these are called *Horn clauses*. Example: cancellative monoids, $x \cdot y = x \cdot z \vdash y = z$.
- Stacks are equationally definable but queues are not.

Example: barycentric algebras (Stone 1949)

- Signature:

$$\{+_{\epsilon} \mid \epsilon \in [0, 1]\}$$

- Axioms:

$$(B_1) \vdash t +_1 t' = t$$

$$(B_2) \vdash t +_{\epsilon} t = t$$

$$(SC) \vdash t +_{\epsilon} t' = t' +_{1-\epsilon} t$$

$$(SA) \vdash (t +_{\epsilon} t') +_{\epsilon'} t'' = t +_{\epsilon\epsilon'} (t' +_{\frac{\epsilon' - \epsilon\epsilon'}{1 - \epsilon\epsilon'}} t'')$$

Universal properties

- Let $\mathbb{K}(\Omega, S)$ be the collection of algebras satisfying the equations in S . $\mathbb{K}(\Omega, S)$ becomes a category if we take the morphisms to be Ω -homomorphisms.
- Let X be a set of generators. We write $T[X]$ for TX / \sim_S . There is a map $\eta_X : X \rightarrow T[X]$ given by $\eta_X(x) = [x]$.
- Universal property.

Set		$\mathbb{K}(\Omega, S)$
X	$\xrightarrow{\eta_X}$	$T[X]$
	$\searrow \alpha$	$\downarrow h$
		\mathcal{A}

Variety theorem

Birkhoff

A collection of algebras is a variety of algebras if and only if it is closed under homomorphic images, subalgebras and products.

There are analogous results for algebras defined by Horn clauses: quasivariety theorems.

Example

Consider $\mathbb{Z}_2 \times \mathbb{Z}_2$. It's not a field because, *e.g.* $(1, 0) \times (0, 1) = (0, 0)$. Hence fields cannot be described by equations!

- Capturing universal algebra categorically.
- Data: (i) Endofunctor $T : \mathcal{C} \rightarrow \mathcal{C}$, (ii) $\eta : I \rightarrow T$ natural, and (iii) $\mu : T^2 \rightarrow T$ also natural.
- Some diagrams are required to commute.

$$\begin{array}{ccc} T^3A & \xrightarrow{\mu_{TA}} & T^2A \\ T\mu_A \downarrow & & \downarrow \mu_A \\ T^2A & \xrightarrow{\mu_A} & TA \end{array}$$

$$\begin{array}{ccccc} & & & & \\ & & & & \\ TA & \xrightarrow{\eta_{TA}} & T^2A & \xleftarrow{T\eta_A} & TA \\ & \searrow & \downarrow \mu_A & \swarrow & \\ & & TA & & \end{array}$$

- Examples: powerset, “free” constructions e.g. monoid, group, the Giry monad.

The Kleisli construction

- From a monad $T : \mathcal{C} \rightarrow \mathcal{C}$ make a new category: the Kleisli category \mathcal{C}_T .
- Objects, the same as those of \mathcal{C} .
- Morphisms $f : A \rightarrow B$ in \mathcal{C}_T are $f : A \rightarrow TB$ in \mathcal{C} .
- Composition? $f : A \rightarrow TB$ and $g : B \rightarrow TC$ don't match.
- $f : A \rightarrow TB$ and $Tg : TB \rightarrow T^2C$ to match but we are in T^2C .
- Compose with $\mu_C : T^2C \rightarrow TC$ to get $A \rightarrow TC$.
- The Kleisli category of the powerset monad is the category of sets and relations.

- **Mes**: objects are sets equipped with a σ -algebra (X, Σ) , morphisms $f : (X, \Sigma) \rightarrow (Y, \Lambda)$ are functions $f : X \rightarrow Y$ such that $\forall B \in \Lambda, f^{-1}(B) \in \Sigma$.
- $\mathcal{G} : \mathbf{Mes} \rightarrow \mathbf{Mes}$, $\mathcal{G}(X, \Sigma) = \{p \mid p \text{ is a probability measure on } \Sigma\}$.
- For each $A \in \Sigma$, define $e_A : \mathcal{G}(X) \rightarrow [0, 1]$ by $e_A(p) = p(A)$. Equip $\mathcal{G}(X)$ with the smallest σ -algebra making all the e_A measurable.
- $f : X \rightarrow Y$, $\mathcal{G}(f) : \mathcal{G}(X) \rightarrow \mathcal{G}(Y)$ given by $\mathcal{G}(f)(p)(B \in \Lambda) = p(f^{-1}(B))$.

- $\eta_X : X \rightarrow \mathcal{G}(X)$ given by $\eta_X(x) = \delta_x$, where $\delta_x(A) = 1$ if $x \in A$ and 0 if $x \notin A$.
- $\mu_X(Q \in \mathcal{G}^2(X))(A) = \int e_A dQ$. Averaging over \mathcal{G} using Q .
- Probabilistic analogue of the powerset.

The Kleisli category of \mathcal{G}

- Objects: Same as **Mes**, morphisms from X to Y are measurable functions from X to $\mathcal{G}(Y)$.
- Compose: $h : X \rightarrow \mathcal{G}(Y)$, $k : Y \rightarrow \mathcal{G}(Z)$ by the formula:
 $(k\tilde{\circ}h) = (\mu_Z) \circ (\mathcal{G}(k)) \circ h$ where $\tilde{\circ}$ is the Kleisli composition and \circ is composition in **Mes**.
- Curry the definition of morphism: $h : X \times \Sigma_Y \rightarrow [0, 1]$. Markov kernels. We call this category **Ker**. Probabilistic relations.
- Composition in terms of kernels:
 $(k\tilde{\circ}h)(x, C \subset Z) = \int k(y, C)h(x, \cdot)$. Relational composition, matrix multiplication.

The Eilenberg-Moore category

- From T we can construct a category of algebras: objects $a : TA \rightarrow A$
- and morphisms $f : A \rightarrow B$ such that

$$\begin{array}{ccc} TA & \xrightarrow{a} & A \\ Tf \downarrow & & \downarrow f \\ TB & \xrightarrow{b} & B \end{array}$$

commute.

- Many categories of algebras (monoids, groups, rings, lattices) can be reconstructed this way.
- The Kleisli category = the category of “free” algebras.
- We get a monad on \mathbf{Set} from $X \mapsto T[X]$. The Eilenberg-Moore category for this monad is isomorphic to $\mathbb{K}(\Omega, S)$.
- Algebras for a monad \Leftrightarrow Algebras given by equations and operations.

- Quantitative analogue of an equivalence relation.
- Space M , (pseudo)metric $d : M \times M \rightarrow \mathbb{R}^{\geq 0}$
- $d(x, x) = 0$, $d(x, y) = d(y, x)$ and $d(x, z) \leq d(x, y) + d(y, z)$.
- If $d(x, y) = 0$ implies $x = y$ we say d is a **metric**.
- We can define usual notions of convergence, completeness, topology, continuity etc.
- Maps: $f(X, d) \rightarrow (Y, d')$ are *nonexpansive* $d'(f(x), f(y)) \leq d(x, y)$; automatically continuous
- We define \mathcal{M} : objects metric spaces, morphisms are nonexpansive functions.
- Quantitative equations give monads on \mathcal{M} .

Metrics between probability distributions

Let p, q be probability distributions on (X, d, Σ) .

- Total variation $tv(p, q) = \sup_{E \in \Sigma} |p(E) - q(E)|$.
- Kantorovich: $\kappa(p, q) = \sup_f \left| \int f dp - \int f dq \right|$ where f is nonexpansive.
- A *coupling* π between p, q is a distribution on $X \times X$ such that the marginals of π are p, q . Write $\mathcal{C}(p, q)$ for the space of couplings.
- Kantorovich: $\kappa(p, q) = \inf_{\mathcal{C}(p, q)} \int_{X \times X} d(x, y) d\pi(x, y)$.
Kantorovich-Rubinshtein duality.
- Wasserstein: $W^{(l)}(p, q) = \inf_{\mathcal{C}(p, q)} \left[\int_{X \times X} d(x, y)^l d\pi(x, y) \right]^{1/l}$. $l = 1$ gives Kantorovich.
- $W^{(l)}(\delta_x, \delta_y) = d(x, y)$.

Quantitative equations

- Signature Ω , variables X we get terms $\mathbb{T}X$.
- Quantitative equations: $\mathcal{V}(\mathbb{T}X)$:

$$s =_{\varepsilon} t, \quad s, t \in \mathbb{T}X, \quad \varepsilon \in \mathbb{Q} \cap [0, 1]$$

- A substitution σ is a map $X \rightarrow \mathbb{T}X$; we write $\Sigma(X)$ for the set of substitutions.
- Any σ extends to a map $\mathbb{T}X \rightarrow \mathbb{T}X$.
- Quantitative inferences: $\mathcal{E}(\mathbb{T}X) = \mathcal{P}_{\text{fin}}(\mathcal{V}(\mathbb{T}X)) \times \mathcal{V}(\mathbb{T}X)$

$$\{s_1 =_{\varepsilon_1} t_1, \dots, s_n =_{\varepsilon_n} t_n\} \vdash s =_{\varepsilon} t$$

Deducibility relations

- (Refl) $\emptyset \vdash t =_0 t$
- (Symm) $\{t =_\varepsilon s\} \vdash s =_\varepsilon t.$
- (Triang) $\{t =_\varepsilon s, s =_{\varepsilon'} u\} \vdash t =_{\varepsilon + \varepsilon'} u.$
- (Max) For $\varepsilon' > 0$, $\{t =_\varepsilon s\} \vdash t =_{\varepsilon + \varepsilon'} s.$
- (Arch) For all $\varepsilon \geq 0$, $\{t =_{\varepsilon'} s \mid \varepsilon' > \varepsilon\} \vdash t =_\varepsilon s.$ **Infinitary!**
- (NExp) For $f : n \in \Omega$,
 $\{t_1 =_\varepsilon s_1, \dots, t_n =_\varepsilon s_n\} \vdash f(t_1, ..t_i, ..t_n) =_\varepsilon f(s_1, ..s_i, ..s_n)$
- (Subst) If $\sigma \in \Sigma(X)$, $\Gamma \vdash t =_\varepsilon s$ implies $\sigma(\Gamma) \vdash \sigma(t) =_\varepsilon \sigma(s).$
- (Cut) If $\Gamma \vdash \phi$ for all $\phi \in \Gamma'$ and $\Gamma' \vdash \psi$, then $\Gamma \vdash \psi.$
- (Assumpt) If $\phi \in \Gamma$, then $\Gamma \vdash \phi.$

- Given $S \subset \mathcal{E}(\mathbb{T}X)$, \vdash_S : smallest deducibility relation containing S .
- Equational theory: $\mathcal{U} = \vdash_S \cap \mathcal{E}(\mathbb{T}X)$.

- Ω : signature; $\mathcal{A} = (A, d)$,
 A an Ω -algebra and (A, d) a metric space.
- All functions in Ω are nonexpansive.
- Morphisms are Ω -algebra homomorphisms that are nonexpansive.
- $\mathbb{T}X$ is an Ω -algebra. $\sigma : \mathbb{T}X \rightarrow A$, Ω -homomorphism.

- (A, d) **satisfies** $\{s_i =_{\varepsilon_i} t_i / i = 1, \dots, n\} \vdash s =_{\varepsilon} t$ if

$$\begin{aligned} \forall \sigma, d(\sigma(s_i), \sigma(t_i)) \leq \varepsilon_i, i = 1, \dots, n \\ \text{implies} \\ d(\sigma(s), \sigma(t)) \leq \varepsilon. \end{aligned}$$

- We write $\{s_i =_{\varepsilon_i} t_i / i = 1, \dots, n\} \models_{\mathcal{A}} s =_{\varepsilon} t$.
- We write $\mathbb{K}(\mathcal{U}, \Omega)$ for the algebras satisfying \mathcal{U} .

$$d^{\mathcal{U}}(s, t) = \inf\{\varepsilon \mid \emptyset \vdash s =_{\varepsilon} t \in \mathcal{U}\}$$

- Why not use the following?

$$d^{\mathcal{U}}(s, t) = \inf\{\varepsilon \mid \forall V \in \mathcal{P}_f(\mathcal{V}(X)), V \vdash s =_{\varepsilon} t \in \mathcal{U}\}$$

- They are the same!
- The (pseudo)metric can take on infinite values.
- The kernel is a congruence for Ω .
- If we take the quotient we get an (extended) metric space.
- The resulting algebra is in $\mathbb{K}(\Omega, \mathcal{U})$.
- We can do this for any set M of generators and produce a “free” quantitative algebra.

$\forall \mathcal{A} \in \mathbb{K}(\mathcal{U}, \Omega), \Gamma \models_{\mathcal{A}} \phi$ if and only if $[\Gamma \vdash \phi] \in \mathcal{U}$.

- Analogue of the usual completeness theorem for equational logic.
- Right to left is by definition.
- Left to right is by a model construction.
- The proof needs to deal with quantitative aspects and uses the archimedean property.

Free construction from a metric space

- Starting from a **metric space** (M, d) we can define $\mathbb{T}M$ by adding constants for each $m \in M$
- and axioms $\emptyset \vdash m =_e n$ for every rational e such that $d(m, n) \leq e$.
- Call this extended signature Ω_M and the extended theory \mathcal{U}_M .
- Any algebra in $\mathbb{K}(\mathcal{U}_M, \mathcal{U}_M)$ can be viewed as an algebra in $\mathbb{K}(\Omega, \mathcal{U})$ by forgetting about the interpretation of the constants from M .
- Given any $\alpha : M \rightarrow A$ non-expansive we can turn $\mathcal{A} = (A, d)$ into an algebra in $\mathbb{K}(\Omega_M, \mathcal{U}_M)$ by interpreting each $m \in M$ as $\alpha(m) \in A$.

Universal property

Met

$\mathbb{K}(\Omega, \mathcal{U})$

$$\begin{array}{ccc} (M, d^M) & \xrightarrow{\eta_M} & T[M] \\ & \searrow \alpha & \downarrow h \\ & & (A, d^A) \end{array} \qquad \begin{array}{c} T[M] \\ \downarrow h \\ \mathcal{A} \end{array}$$

\mathcal{U}_M is consistent if and only if the map η_M is an isometry.

We have a monad on \mathcal{M} .

Barycentric algebras again

- $\Omega = \{+_e : 2|e \in [0, 1]\}$; uncountably many operations!
- **(B1)** $\emptyset \vdash x +_1 y =_0 x$
- **(B2)** $\emptyset \vdash x +_e x =_0 x$
- **(SC)** $\emptyset \vdash x +_e y =_0 y +_{1-e} x$
- **(SA)** $(x +_{e_1} y) +_{e_2} z =_0 x +_{e_1 e_2} (y +_{\frac{e_2 - e_1 e_2}{1 - e_1 e_2}} z)$ where $e_1, e_2 \in (0, 1)$
- **(LI)** $x +_e z =_\varepsilon y +_e z$ where $e \leq \varepsilon \in \mathbb{Q} \cap [0, 1]$
- The last equation uses one of the new indexed equations in a nontrivial way.
- We call it the *left-invariant* axiom scheme; LIB algebras for short.
- What does this axiomatize?
- The total variation metric on probability distributions.

$$TV(p, q) = \sup_{E \in \Sigma} |p(E) - q(E)|.$$

- It measures the size of the set on which p, q disagree the most.
- There is a duality theorem that gives it as a minimum rather than a maximum.

- Let $\mathcal{B}(M, \Sigma)$ be the Borel measures on a metric space M with Borel algebra Σ .
- We have a product space $M \times M$ with product σ -algebra $\Sigma \otimes \Sigma$ and Borel measures $\mathcal{B}(M \times M, \Sigma \otimes \Sigma)$.
- Given probability measures p, q a *coupling* is a probability measure ω on $(M \times M, \Sigma \otimes \Sigma)$ such that for all $E \in \Sigma$:

$$\omega(E \times M) = p(E) \quad \text{and} \quad \omega(M \times E) = q(E).$$

- $\mathcal{C}(p, q)$ is the set of couplings for (p, q) .

- Write Δ for the diagonal in $M \times M$.
- TV duality: $TV(p, q) = \min\{\omega(\Delta^c) \mid \omega \in \mathcal{C}(p, q)\}$; min is attained.
- Convex combinations of couplings are couplings.
- Splitting lemma: If p, q are Borel probability measures on M and $e = T(p, q)$. There are p', q', r such that

$$p = ep' + (1 - e)r \text{ and } q = eq' + (1 - e)r.$$

- We know there is a freely generated LIB algebra from a metric space M . What is it concretely?
- Let $\Pi[M]$ be the LIB algebra obtained by taking the *finitely-supported* probability measures on M and interpreting $+_e$ as convex combination.
- We endow it with the total-variation metric to make it a quantitative algebra.

- Theorem: $\Pi[M] \in \mathbb{K}(\mathcal{B}, \mathcal{U}^{LI})$.
- Use convexity and splitting lemma to show LI and Nexp.
- Theorem: $\Pi[M]$ is the free algebra generated by M .
- Use the embedding of convex spaces into vector spaces (Stone 49).
- The axioms give rise to the total-variation metric.

Interpolative barycentric algebras

- Same signature as barycentric algebras.
- Axioms (B1), (B2), (SC), (SA); drop (LI).
- (\mathbf{IB}_p)

$$\{x =_{\varepsilon_1} y, x' =_{\varepsilon_2} y'\} \vdash x +_e x' =_{\delta} y +_e y',$$

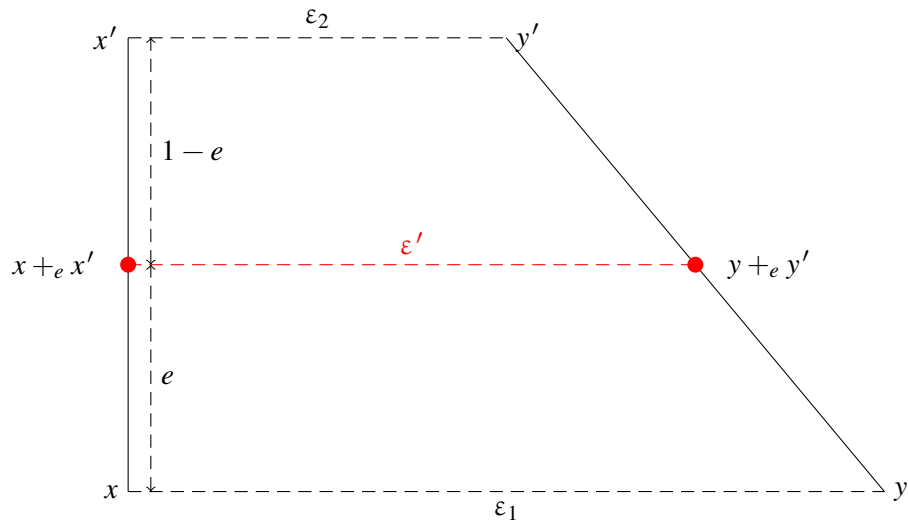
where $(e\varepsilon_1^p + (1 - e)\varepsilon_2^p)^{1/p} \leq \delta$.

- Now we need assumptions in the equation.
- If $p = 1$ we get

$$\{x =_{\varepsilon_1} y, x' =_{\varepsilon_2} y'\} \vdash x +_e x' =_{\delta} y +_e y',$$

where $e\varepsilon_1 + (1 - e)\varepsilon_2 \leq \delta$.

Picture of IB_1



Kantorovich-Wasserstein metric

Let (M, d) be a complete separable metric space and $p \geq 1$.

Wasserstein- p metric

$$W_d^p(\mu, \nu) = \inf \left\{ \left[\int_{M \times M} d^p(x, y) d\omega \right]^{1/p} \mid \omega \in \mathcal{C}(\mu, \nu) \right\}$$

Kantorovich

$$K_d(\mu, \nu) = \sup \left\{ \left| \int f d\mu - \int f d\nu \right| \right\}$$

Duality

$$K_d(\mu, \nu) = \min \left\{ \left[\int_{M \times M} d(x, y) d\omega \right] \mid \omega \in \mathcal{C}(\mu, \nu) \right\}$$

- We take the finitely supported measures on M and interpret it as a barycentric algebra as before.
- We give it the Wasserstein metric and show that we get an IB algebra.
- This uses the definition of the W_d^p metrics as an inf and convexity of couplings.
- We prove a splitting lemma for this case and show that we get the free algebra by similar, but more involved arguments.
- How do we lift it to the continuous case?

Weak convergence

- Suppose we have a sequence of measures $\{\mu_i | i \in I\}$. What does it mean to converge?
- For a “suitable” class of functions:

$$\int f d\mu_i \rightarrow \int f d\mu.$$

- For Kantorovich use contractive functions; for Wasserstein use a class of functions whose growth is controlled by d and p .
- The Wasserstein metrics give the topology of weak convergence on measures of finite p -moment.
- The finitely supported probability measures are *dense* in the space of all probability measures with weak topology.

Complete separable metric spaces

- A separable metric space has a countable dense subset.
- Define $\Delta[M]$ to be the space of all Borel probability measures on a complete separable metric space. We give it the W_d^p metric and interpret $+_e$ as convex combination.
- This gives an IB algebra.
- If we construct the term algebra $\mathbb{T}[M]$ as before and *complete it* we get an algebra isomorphic to $\Delta[M]$.
- In this case we get a monad on **CSMet**₁: complete separable 1-bounded metric spaces.

- Quantitative equations give a handle on otherwise arcane things like the Wasserstein metrics.
- Other examples: Hausdorff metric, pointed barycentric algebras.
- To do; many more examples:
 - Markov processes
 - Choquet capacities and games
 - quantitative theory of effects
 - quantitative equational axioms for probabilistic programming languages.