34. A Mathematical Theory of Randomized Computation. I

By Shinichi YAMADA
Waseda University and Nihon Unisys, Ltd.

(Communicated by Shokichi IYANAGA, M. J. A., April 12, 1988)

1. Introduction. Randomized algorithms, or probabilistic algorithms, are the extended algorithms which incorporate random input data and random choices. They have recently been recognized as an important area of information science and undergone intensive studies. Whereas a mathematical theory of randomized computation is indispensable to guarantee the reliability of stochastic software, there has been no theory comparable to Scott's theory [3, 4] for deterministic computation. The present series of communications outlines a new randomized domain theory [5] which, while naturally extending Scott's deterministic domain theory, readily provides a denotational semantics for a class of high level probabilistic programming languages and is also applicable to machine learning and algorithmic information theory.

2. The postulates for the probability theory. There have been pointed out many pathological phenomena that arise within the framework of the axiomatic theory of probability. In order to avoid these defects, we postulate the following Axioms 1 and 2 for the probability theory. The space Ω in a measurable space $(\Omega, \mathcal{A})$ is called a *basic space*.

Axiom 1. *Every basic space is compact, Hausdorff and second countable, if it is endowed with a topology. Every basic space has power* $\leq \aleph_1$, *if it is not endowed with any topology.*

Axiom 2. *Every probability space is separable and perfect.*

Given a measurable space $(\Omega, \mathcal{A})$, we shall call a measure μ on $\mathcal{A}$ a *subprobability measure* if it satisfies $\mu(\Omega) \leq 1$.

3. The Scott topology. The essence of Scott's theory is the idea of finite approximation coupled with the Kleene first recursion theorem. A computation is thought as a sequence or a directed set of increasingly refined approximations whose supremum is the desired result. To formulate the idea, we introduce "undefined value" $\perp$ and "approximation ordering" $\sqsubseteq$ in a domain of computation C and think of the domain as a poset $\boldsymbol{D}=(D, \sqsubseteq)$, where $D := C \cup \{\perp\}$. The intuitive meaning of $\sqsubseteq$ is as follows: Every x in C is a definite value "well defined" compared with the undefined value $\perp$. So $\perp$ is considered to "approximate x with respect to the degree of definition". In other words, "x is better defined than $\perp$", and written $\perp \sqsubseteq x$. For $x, y \in C$, if $x \neq y$ then "x and y do not approximate each other". Technically we define it as follows:

(1) A poset $\boldsymbol{D}=(D, \leq)$ is a *conditionally complete poset* (*ccp* for short)

if (i) $\exists \perp \in D$, $\forall x \in D$ $[\perp \leq x]$ ($\perp$ is called the *bottom*) and (ii) every non-empty upwards directed subset X of D which is bounded from above has a supremum in D.

To have a well-behaved "limit", we assume that every element in $\boldsymbol{D}$ is itself the supremum in the ordering $\sqsubseteq$ and then we say that an element y s.t. $y \sqsubseteq x$ is an approximation of x if, for every directed set Z s.t. $x \sqsubseteq \sup Z$, there exists z in Z s.t. $y \sqsubseteq z$:

(2) Let $\boldsymbol{D}=(D, \leq)$ be a ccp. For $\forall x, y \in D$, x is *way below* y (in symbols, $x \ll y$) if for every upwards directed subset Z of D possessing the supremum z (in symbols, $Z \uparrow z \subset D$), $[y \leq z \Rightarrow \exists z_0 \in Z\, [x \leq z_0]]$. $x \in D$ is *compact* iff $x \ll x$.

In a ccp $\boldsymbol{D}=(D, \leq)$, we shall write $\downarrow x := \{y \in D \mid y \leq x\}$.

The set of all compact elements in a ccp $\boldsymbol{D}$ is denoted by $K(\boldsymbol{D})$.

(3) A ccp $\boldsymbol{D}=(D, \leq)$ is *algebraic* if (i) $\downarrow x \cap K(\boldsymbol{D})$ is upwards directed and (ii) $x=\sup\{y \mid y \in \downarrow x \cap K(\boldsymbol{D})\}$, for $\forall x \in D$.

We call the topology induced by $\sqsubseteq$ the Scott topology:

(4) Let $\boldsymbol{D}=(D, \leq)$ be a ccp. The *Scott topology* on D is defined as follows: $\mathcal{O} \subset D$ is *open* if (i) $x \in \mathcal{O}$ and $x \leq y \Rightarrow y \in \mathcal{O}$ and (ii) $\forall X \uparrow x \subset D$ and $x \in \mathcal{O} \Rightarrow X \cap \mathcal{O} \neq \emptyset$.

(5) A set E in a ccp $\boldsymbol{D}=(D, \leq)$ is a *subbasis* of $\boldsymbol{D}$ if for $\forall x \in D$, $x=\sup\{e \mid e \in E \text{ and } e \leq x\}$. A subbasis $E \subset D$ is a *basis* of D if $\sup F \in E$ for $\forall$ finite $F \subset E$.

Now we consider function domains. Let f be the function recursively defined by $f(x) := F(f, x)$. We think of f as a fixed point of the "Scott continuous" functional F, then the computation of f is given as the sequence $\{g_n\}$ of increasingly refined approximate functions, whose supremum $\sup_n\{g_n\}$ is f: $g_0 \sqsubseteq g_1 \sqsubseteq \cdots \sqsubseteq g_n \sqsubseteq g_{n+1} \sqsubseteq \cdots$, where g_0 is the totally undefined function $\lambda x.\perp$ and $g_{n+1}(x)=F(g_n, x)=F^n(g_0, x)$ for $\forall n \geq 1$. So we define the products and exponents of ccp's as follows:

(6) Given ccp's $\boldsymbol{D}=(D, \leq)$ and $\boldsymbol{D}'=(D', \leq')$. (i) The *product* $\boldsymbol{D} \times \boldsymbol{D}'$ is the cartesian product of D and D' partially ordered by $\langle x, x' \rangle \leq \langle y, y' \rangle$ iff $x \leq y$ and $x' \leq y'$. (ii) The *function space* $[\boldsymbol{D} \to \boldsymbol{D}']$ is the set of all Scott continuous functions $T: \boldsymbol{D} \to \boldsymbol{D}'$ partially ordered $T \leq S$ in $[\boldsymbol{D} \to \boldsymbol{D}']$ iff $\forall x \in D$ $[Tx \leq Sx]$.

Then the following relations are easily deduced from (1)–(6):

(7) (i) *If* $\boldsymbol{D}$ *and* $\boldsymbol{D}'$ *are algebraic ccp's, then* $\boldsymbol{D} \times \boldsymbol{D}'$ *and* $[\boldsymbol{D} \to \boldsymbol{D}']$ *are algebraic ccp's.* (ii) *If ccp's* $\boldsymbol{D}$ *and* $\boldsymbol{D}'$ *have effectively given countable bases, then* $\boldsymbol{D} \times \boldsymbol{D}'$ *and* $[\boldsymbol{D} \to \boldsymbol{D}']$ *have effectively given countable bases.*

4. Randomized algorithms. We illustrate the semantics of randomized programs with the following *randomized McCarthy formalism*:

(8) Let X be the domain of computation and B a class of base functions on X. The class $C(B)$ of recursively defined functions on X is the smallest class of functions containing B and closed under the formation of function compositions, conditional expressions, recursive definitions,

λ-abstractions and label notations.

In a randomized algorithm (or program) $F \in \mathcal{C}(B)$ the state vector composed of the program variables and random oracles is thought as a random (vector) variable $x: (\Omega, \mathcal{A}, \mu) \to (X, \mathcal{B})$, where $(\Omega, \mathcal{A}, \mu)$ and $(X, \mathcal{B})$ are an a priori given sample space and a value space, respectively. Let $L(X)$ and $M(X)$ be the spaces of partial measurable functions from $(X, \mathcal{B})$ to itself and subprobability measures on $(X, \mathcal{B})$, respectively. Then a sample *execution of the program* F consists of picking up a sample point $\omega \in \Omega$, determining the initial value of the state vector x, F executes *deterministically*, changing the contents of the state vector. Upon exit, the *effect* of the execution may be described by a partial measurable function F mapping x to the final state vector $F \circ x$. So, as the first *meaning* $m_1: \mathcal{C}(B) \to L(X)$, we have $m_1[\![F]\!] := F$. F also induces the operator T_F mapping an input probability measure $\mu^x := \mu \circ x^{-1}$ to the output subprobability measure $\mu^x \circ F^{-1}$ by $T_F(\nu) := \nu \circ F^{-1}$ $(\forall \nu \in M(X))$. So we have $m_2[\![F]\!] := T_F$ as the second *meaning* $m_2: \mathcal{C}(B) \to \mathrm{Hom}(M(X), M(X))$. Thus the *semantics* m_1 and m_2 are defined inductively as follows:

(i) If $F \in B$, then $m_1[\![F]\!] := F$ and $m_2[\![F]\!] := T_F$.

(ii) If $F, G \in \mathcal{C}(B)$ and $H := G \circ F$, then $m_1[\![H]\!] := m_1[\![G]\!] \circ m_1[\![F]\!]$ and $m_2[\![H]\!] := m_2[\![G]\!] \circ m_2[\![F]\!]$.

(iii) Let $P, F, G \in \mathcal{C}(B)$ and $H(x) :=$ if $P(x)$ then $F(x)$ else $G(x)$, where P is a propositional function symbol. Then

$$m_1[\![H]\!] := \lambda x.\text{if } m_1[\![P]\!](x) \text{ then } m_1[\![F]\!](x) \text{ else } m_1[\![G]\!](x).$$

Define $\mu_p \in M(X)$ for P by $\mu_P(E) := \mu(\{x \mid m_1[\![P]\!](x)\} \cap E)$ for $\forall E \in \mathcal{B}$ and let the projection e_P be $e_P(\mu) := \mu_P$ for $\forall \mu \in M(X)$. Then

$$m_2[\![H]\!] := m_2[\![H]\!] \circ e_P + m_2[\![G]\!] \circ e_{\sim P},$$

where $\sim P$ denotes the negation of P.

(iv) If F is defined by the recursive definition $F := \tau(F)$. Then F denotes the least fixed point $m_1[\![\tau(F)]\!] := \mathrm{Fix}(\tau)$ of $F = \tau(F)$ and $m_2[\![\tau(F)]\!] := T_F$, where $T_F(\mu) := \mu \circ \mathrm{Fix}(\tau)^{-1}$ for $\forall \mu \in M(X)$. (The fixed point theorem is discussed later.) Thus we may think of a randomized program as a positive linear operator mapping an input probability measure to the output subprobability measure.

Now it is well known ([1] and [2]) that every space $V = V(X)$ of all bounded measures on a measurable space $(X, \mathcal{B})$ with a compact Hausdorff space X, endowed with the following natural vector lattice structure and total variation norm, is isometric and order isomorphic to an AL-space and conversely:

For $\forall \mu, \nu \in V$, $(\mu+\nu)(A) := \mu(A)+\nu(A)$ and $(\lambda\mu)(A) := \lambda\mu(A)$ for $\forall \lambda \in \boldsymbol{R}$, $\forall A \in \mathcal{B}$, $\mu \geq \nu$ iff $\mu(A) \geq \nu(A)$ for $\forall A \in \mathcal{B}$,

$(\mu \vee \nu)(A) := \sup\{\mu(B)+\nu(A-B),\ B \subset A\}$, and

$(\mu \wedge \nu)(A) := \inf\{\mu(B)+\nu(A-B),\ B \subset A\}$ for $\forall A \in \mathcal{B}$, and

$\|\mu\| := |\mu|(X)$, the total variation norm.

Clearly the space of subprobability measures, or a basic *randomized*

domain, is the *positive unit hemisphere* $\mathcal{H}(V) := U \cap V^+$, i.e., the intersection of the *unit ball* $U := \{x \in V \mid \|x\| \leq 1\}$ and the *positive cone* $V^+ := \{x \in V \mid x \geq 0\}$ of the *AL*-space V of bounded measures.

5. **Banach lattices.** We recall the relevant definitions. Let $\boldsymbol{R}$ denote the real number field. (i) An *ordered vector space* (*OVS*) is a vector space V over $\boldsymbol{R}$ endowed with a partial order relation $\leq$ satisfying (a) $x \leq y \Rightarrow x+z \leq y+z$ for $\forall x, y, z \in V$, and (b) $x \leq y \Rightarrow \lambda x \leq \lambda y$ for $\forall x, y \in V$ and $\forall \lambda \in \boldsymbol{R}$ s.t. $\lambda \geq 0$. (ii) A *vector lattice* (*VL*) is an *OVS* V in which $x \vee y := \sup\{x, y\}$ and $x \wedge y := \inf\{x, y\}$ exist in V, for $\forall x, y \in V$. (iii) Let V be a *VL*. A function $\| \ \| : V \to \boldsymbol{R}$ is called a *norm* of V if (a) $\|x\| \geq 0$ and $\|x\| = 0 \Leftrightarrow x = 0$ for $\forall x \in V$, (b) $\|x+y\| \leq \|x\| + \|y\|$ for $\forall x, y \in V$, and (c) $\|\lambda x\| = |\lambda| \cdot \|x\|$ for $\forall \lambda \in \boldsymbol{R}$ and $\forall x \in V$. A norm $\| \ \|$ on V is called a *lattice-norm* (or, *norm*) on V if (d) $\|x\| = \| |x| \|$ $(\forall x \in V)$ and (e) $x \leq y \Rightarrow \|x\| \leq \|y\|$ $(\forall x, y \in V^+)$. The pair $(V, \| \ \|)$ is called a *normed VL* if $\| \ \|$ is a lattice-norm on V. We shall also denote a normed *VL* $(V, \| \ \|)$ by V. (iv) A normed *VL* $(V, \| \ \|)$ is called a *Banach lattice* (a *BL*, for short) if $(V, \| \ \|)$ is complete under the metric topology generated by the metric $\rho(x, y) := \|x-y\|$. (v) An *AL-space* is a *BL* V s.t. $\|x+y\| = \|x\| + \|y\|$ for $\forall x, y \in V^+$. (vi) An *AM-space* is a *BL* V s.t. $\|x \vee y\| = \max(\|x\|, \|y\|)$ for $\forall x, y \in V^+$. (vii) An *AM-space with unit* is an *AM*-space V in which the closed unit ball $U := \{x \in V \mid \|x\| \leq 1\}$ contains a largest element e (in the order topology) s.t. $\|e\| = 1$. e is called the *unit* of V.

We note that the order topology and the norm topology are independent on *BL*'s. For the order topology of a Banach lattice:

(9) Let V be a *BL*. Then (i) V is *σ-order complete* if every non-empty countable subset of V which is bounded from above has a supremum. (ii) V is *order complete* if every non-empty subset of V which is bounded from above has a supremum. (iii) V is *order separable* if every non-empty subset A of V possessing a supremum $\sup A$ contains a countable subset A_0 satisfying $\sup A = \sup A_0$.

(to be continued.)

References

[1] N. Bourbaki: Intégration. Hermann, Paris (1965).
[2] S. Kakutani: Ann. of Math., **42**, 994–1024 (1941).
[3] D. S. Scott: Lecture Notes in Math., vol. 274, 97–136 (1972).
[4] ——: SIAM J. Comput., **5**, 522–587 (1976).
[5] S. Yamada: D. Sc. Thesis, University of Tokyo (1987).