Proposal for a postdoctoral position

Acronym : DAfPUP

Scientific axis of the corresponding GT : SciLex

Principal investigator : Jean Goubault-Larrecq
(`http://www.lsv.fr/~goubault/`),

Hosting Digicosme lab : LSV
(`http://www.lsv.fr/`)

Managing lab : Laboratoire Spécification et Vérification (CNRS UMR 8643)
École Normale Supérieure Paris-Saclay
61, avenue du Président Wilson
94235 Cachan Cedex
Scientific contact: Jean Goubault-Larrecq

Corresponding GT : Systèmes hybrides
(`http://labex-digicosme.fr/GT+SHY`)

Involved teams : Apart from LSV, team Cosynus at Ecole Polytechnique, under the direction of Eric Goubault
(`http://www.lix.polytechnique.fr/Labo/Eric.Goubault/`).

Duration and contract dates : one year, April 2018–March 2019.

Type of contract : postdoc.

Title : Domain Approximations for Programs with Uncertain Probabilities

Context and state of the art : The starting point of this proposal has its origins in the paper [ABG+14] by Adjé, Bouissou, Goubault-Larrecq, Goubault, and Putot, where a static analysis framework is proposed for numerical programs with inputs given as imprecise probabilities.

Such programs form an important family of critical mission programs, including controllers in planes and trains. The imprecise probabilistic inputs are typically given by sensors, whose value is probabilistic, but obeys partially known probability distributions. For example, it might be known that a temperature sensor returns a Gaussian distributed value whose average is the outside temperature T, with standard deviation ± 0.5°C, but the outside temperature T itself is only known to be between $-55°$ and $-45°$ when the plane reaches its cruise altitude.

A close piece of work is due to Sankaranarayanan [SCG13], with applications in the medical domain. One may find other related work by following references. Yet other related work, pertaining to the mathematical models to be used, will be cited below.

Objectives and avenues of work : The above paper by Adjé *et al.* rests on a formal concrete semantics that one may consult in [AG12]. It is based on a measure-theoretic approach, where values of sensors, hence also of program variables, are certain convex sets of probability distributions encoded as *measure-theoretic previsions*. Those are defined as (superlinear, resp. sublinear) ω-continuous functionals $F\colon \mathcal{M}(X) \to [0,1]$, where $\mathcal{M}(X)$ is the family of all measurable maps f from X to $[0,1]$, and can be thought of as generalized integrals of the input function f.

While the authors manage to derive a static analysis framework from such a concrete semantics, that semantics is somehow unsatisfactory: it completely ignores computability requirements. In other words, it interprets programs in so large a domain of definition that this domain contains many non-computable functions. Concretely, although time has taught us that computable functions must be continuous in some precise, domain-theoretic sense, the functions used in the above semantics are very far from being continuous, and are only measurable. Correspondingly, this concrete semantics has *no* theory of approximation whatsoever, and the static analysis found in the papers above is ad hoc.

The goal of this postdoctoral position is to improve on such a situation. The impact is not only theoretical, it is expected to lead to new static analysis domains, found in a principled way. The spirit of the research is in line with work done by Abbas Edalat and colleagues on approximation domains for integration with application to the theory of fractals [Eda95], or to Brownian motion [BE17], and approximation domains for the Clarke subgradient operator [EM17].

Let us deal with the the case at hand, approximations for uncertain probabilities. The approximations that are used by Adjé *et al.* are given in terms of *P-boxes*. For uncertain probabilities on the real line $\mathbb{R}$, those consist essentially in one upper bound and one lower bound on the set of probability distributions represented by F, and those bounds are represented as staircase-shaped cumulative distribution functions. This is a relatively coarse approximation. In practice, one knows that certain laws obey, or are close to some of the well-known probability laws, Gaussian, binomial, or other. This is what Sankaranarayanan *et al.* manage to compute on [SCG13], however for a limited number of known probability distributions, and still without a unifying theory of approximation.

On the contrary, *domain theory* is the preferred theory of approximation and computation in computer science. The principal investigator of this proposal is an internationally recognized expert in that field [Gou13]. He is responsible for laying out the corresponding, analogous theory of *domain-theoretic previsions* in [Gou07, Gou16], with application to the semantics of higher-order functional programs with uncertain (or certain) probabilities [Gou15].

However, and although it predated the paper by Adjé *et al.* [ABG$^+$14], that theory of domain-theoretic previsions did not find its way in the latter. That is unnatural, and should be investigated. The main theoretical question of this proposal is therefore:

> Can one find a measure-theoretic semantics of uncertain probabilities as a form of limit of domain-theoretic semantics of uncertain probabilities?

This should come with a theory of approximation, in the style of domain theory: new continuous domains, with yet to be understood way-below relations.

In the case of (not uncertain) probabilities, we already have so-called measure extension theorems, of which the most complete account is given by Keimel and Lawson [KL05]. In the case of uncertain probabilities, the question is open.

The most obvious way to think of the problem is the relation between spaces of *measurable* functions, as used in the definition of measure-theoretic previsions, and *lower semicontinuous* functions, as used in domain-theoretic previsions.

> Can one find measurable maps from X to $\mathbb{R}$ as some form of limit of lower semicontinuous maps?

The Lebesgue-Hausdorff theorem [Sri98, Theorem 3.1.36 and Proposition 3.1.32] states that, provided X is a metrizable space, if you close the space of continuous real-valued maps from X to $\mathbb{R}$ under pointwise limits, you obtain exactly the usual measurable maps. That result is mostly used when X is more than just metrizable, and is Polish. Then, work by Lawson and by Martin (see [Gou13], Section 7.7.2) has shown that the regular spaces that have a good, domain-theoretic notion of approximation in the form of a *computational model*, are *exactly* the Polish spaces. It should therefore be a reasonable starting point to investigate whether the spaces of measurable functions, from a space X with a computational model to some well-behaved space such as $\mathbb{R}$, themselves have good computational models.

There are many related questions, which may be explored by the candidate. There are relations to the type 2 theory of effectivity, which heavily depends on the Polish space $\mathbb{N}^{\mathbb{N}}$ (Baire space). Questions of quasi-metrics between previsions [Gou08, GL17], with the purpose of finding ways to evaluate distances between approximations to the modeled system, or between two approximations, would also be natural questions. In general, there are many related questions, and those are only a sample.

Difficulties : The main questions are voluntarily open. The general question has eluded the principal investigator for years. I have listed some ideas which one may use, in the previous discussion.

The difficulties are clearly of a mathematical nature, hence what is required is theoretical work. If we ever manage to understand approximation so well in this context that we could hope for actual algorithms, that would be ideal. I do not think this is realistic, though, and we should concentrate on the mathematics for now.

Criteria of success : Publications in the best international venues. Since this is a wide open area, any positive result will be be valuable. Every negative result is welcome as well, as this is also the way science progresses.

References

[ABG$^+$14] Assalé Adjé, Olivier Bouissou, Jean Goubault-Larrecq, Éric Goubault, and Sylvie Putot. Static analysis of programs with imprecise probabilistic inputs. In Ernie Cohen and Andrey Rybalchenko, editors, *Revised Selected Papers of the 5th IFIP TC2/WG2.3 Conference Verified Software—Theories, Tools, and Experiments (VSTTE'13)*, volume 8164 of *Lecture Notes in Computer Science*, pages 22–47, Atherton, California, USA, 2014. Springer.

[AG12] Assalé Adjé and Jean Goubault-Larrecq. Concrete semantics of programs with non-deterministic and random inputs. Research Report cs.LO/1210.2605, Computing Research Repository, October 2012. 19 pages.

[BE17] Paul Bilokon and Abbas Edalat. A domain-theoretic approach to brownian motion and general continuous stochastic processes. *Theor. Comput. Sci.*, 691:10–26, 2017.

[Eda95] Abbas Edalat. Domain theory and integration. *Theoretical Computer Science*, 151:163–193, 1995.

[EM17] Abbas Edalat and Mehrdad Maleki. Differentiation in logical form. In *32nd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017, Reykjavik, Iceland, June 20-23, 2017*, pages 1–12. IEEE Computer Society, 2017.

[GL17] Jean Goubault-Larrecq. Complete quasi-metrics for hyperspaces, continuous valuations, and previsions. Technical report, LSV, ENS Paris-Saclay, July 2017. arXiv:1707.03784.

[Gou07] Jean Goubault-Larrecq. Continuous previsions. In Jacques Duparc and Thomas A. Henzinger, editors, *Proceedings of the 16th Annual EACSL Conference on Computer Science Logic (CSL'07)*, pages 542–557, Lausanne, Switzerland, September 2007. Springer-Verlag LNCS 4646.

[Gou08] Jean Goubault-Larrecq. Simulation hemi-metrics between infinite-state stochastic games. In Roberto Amadio, editor, *Proceedings of the 11th International Conference on Foundations of Software Science and Computation Structures (FoSSaCS'08)*, pages 50–65, Budapest, Hungary, March-April 2008. Springer-Verlag LNCS 4962.

[Gou13] Jean Goubault-Larrecq. *Non-Hausdorff Topology and Domain Theory—Selected Topics in Point-Set Topology*, volume 22 of *New Mathematical Monographs*. Cambridge University Press, 2013. ISBN 978-1-107-03413-6.

[Gou15] Jean Goubault-Larrecq. Full abstraction for non-deterministic and probabilistic extensions of PCF I: the angelic cases. *Journal of Logic and Algebraic Methods in Programming*, 84(1):155–184, January 2015.

[Gou16] Jean Goubault-Larrecq. Isomorphism theorems between models of mixed choice. *Mathematical Structures in Computer Science*, FirstView:1–36, 6 2016. To appear.

[KL05] Klaus Keimel and Jimmie Lawson. Measure extension theorems for T_0-spaces. *Topology and its Applications*, 149(1–3):57–83, 2005.

[SCG13] Sriram Sankaranarayanan, Aleksandar Chakarov, and Sumit Gulwani. Static analysis for probabilistic programs: Inferring whole program properties from finitely many paths. *SIGPLAN Not.*, 48(6):447–458, June 2013.

[Sri98] Sashi Mohan Srivastava. *A Course on Borel Sets*, volume 180 of *Graduate Texts in Mathematics*. Springer Verlag, 1998.