Partial inconsistency and mathematics of software

Michael Bukatin

Nokia Corporation, Cambridge, MA

Joint work with Ralph Kopperman and Steve Matthews

- - -

CCNY, Joint Math-CS Colloquium, April 11, 2013

Electronic coordinates

These slides are linked from my new page on partial inconsistency and vector semantics of programming languages:

http://www.cs.brandeis.edu/∼bukatin/partial_inconsistency.html

E-mail:

bukatin@cs.brandeis.edu

Abstract

The mathematics of partial inconsistency starts with adjoining formally inconsistent elements, such as negative probabilities or formal intervals $[a, b]$ with the contradictory property that $b < a$.

Then one progresses to more complex subjects, such as the "bilattice pattern" which arises in different situations, bitopology, ordered Banach spaces of measures, Scott domains, and a new type of "possible worlds" models.

Actual and potential applications include handling inconsistency in databases, non-monotonic reasoning in AI, semantics of probabilistic programs, and, with some luck, better machine learning schemes over spaces of programs.

Puzzle connectors

This talk overviews how a variety of different studies done by different people seem to fall together like pieces of a single puzzle. Here is the list of the puzzle connectors:

- Bilattice pattern
- Partial inconsistency
- Non-monotonic inference
- Bitopology
- Negative probability and parametrization by signed measures
- Group and vector space semantics of programming languages compatible with Scott domain semantics

Outline

1. Introduction
   - Partially inconsistent interval numbers
   - Negative probability
2. Bilattice pattern
   - Examples
   - History and definition
   - Bilattice pattern
   - Bitopology and *d*-frames
3. Signed measures and vector semantics

Interval numbers

Segments $[a, b]$ on real line, $a \leq b$.

What $[a, b]$ means: $[a, b]$ stands for a partially defined number x, what is known about x is the constraint $a \leq x \leq b$.

Partial order on the interval numbers:
$[a, d] \sqsubseteq [b, c]$ iff $a \leq b$ $(\leq)$ $c \leq d$.

Here $[b, c]$ is better (more precisely) defined than $[a, d]$.

Addition and weak minus

Addition: $[a_1, b_1] + [a_2, b_2] = [a_1 + a_2, b_1 + b_2]$.

Weak minus: $-[a, b] = [-b, -a]$.

These are monotonic operations:
$x \sqsubseteq y \Rightarrow x + z \sqsubseteq y + z$ and $-x \sqsubseteq -y$.

However, the minus is weak, e.g. $-[2, 3] = [-3, -2]$, so
$-[2, 3] + [2, 3] = [-1, 1] \sqsubset [0, 0]$.

So one does not get a group here.
And it would be nice to have a group.

Partially inconsistent interval numbers

Add pseudosegments $[a, b]$, such that $b < a$.

This corresponds to contradictory constraints, $x \leq b \& a \leq x$.

The new set consists of segments and pseudosegments.

Addition: $[a_1, b_1] + [a_2, b_2] = [a_1 + a_2, b_1 + b_2]$.

True minus: $-[a, b] = [-a, -b]$.

$-[a, b] + [a, b] = [0, 0]$.

This gets us a group.

True minus is antimonotonic

$x \sqsubseteq y \Rightarrow -y \sqsubseteq -x$.

True minus maps precisely defined numbers, $[a, a]$, to precisely defined numbers, $[-a, -a]$.

Other than that, true minus maps segments to pseudosegments and maps pseudosegments to segments.

This suggests that we ought to look beyond Scott paradigm that "computable implies monotonic" and bring bitopology into play.

Negative self-distances

The standard partial metric on interval numbers is $p([a_1, b_1], [a_2, b_2]) = \max(b_1, b_2) - \min(a_1, a_2)$.

Hence for $x = [a, b]$ the self-distance is $p(x, x) = b - a$.

Hence if x is a pseudosegment, and if we expect the formula above to hold, the self-distance is negative.

Negative self-distance and negative probability

Partial metrics are often expressed via probability measures of certain sets associated with pairs of points.

If partial metrics can take negative values, we should consider using signed measures and negative probabilities.

Negative probability in the literature

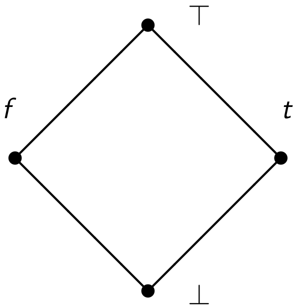
The idea of negative probability is not new, see e.g.

http://en.wikipedia.org/wiki/Negative_probability

and links and references therein for some of the history of this notion and its applications in physics and in mathematical finance.

See in particular Richard Feynman, "Negative Probability," in *Quantum Implications : Essays in Honour of David Bohm*, F. David Peat (Ed.), Basil Hiley (Ed.), Routledge & Kegan Paul Ltd, London and New York, 1987, pp. 235–248.

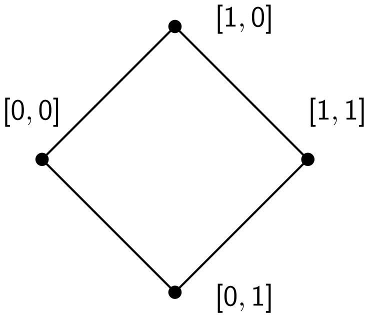
Example of a bilattice



$f < \bot < t, f < \top < t$

$\bot \sqsubset f \sqsubset \top, \bot \sqsubset t \sqsubset \top$

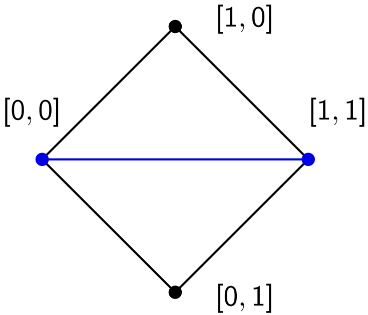
Partially inconsistent interval numbers within a segment



$[a, b] \leq [c, d]$ iff $a \leq c, b \leq d$

$[a, d] \sqsubseteq [b, c]$ iff $a \leq b, c \leq d$

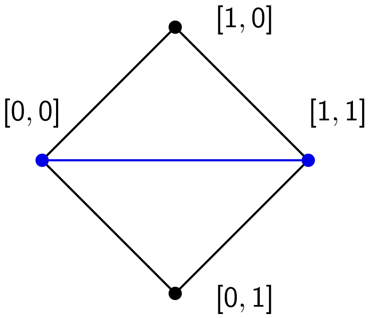
Partially inconsistent interval numbers within a segment



$[a, b] \leq [c, d]$ iff $a \leq c, b \leq d$

$[a, d] \sqsubseteq [b, c]$ iff $a \leq b, c \leq d$

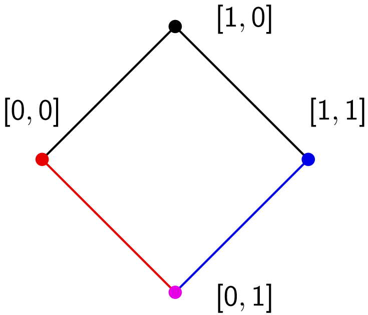
Partially inconsistent interval numbers within a segment



blue – precisely defined numbers

pseudosegments are above the blue

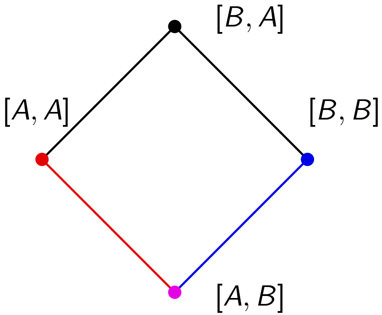
Negative and positive subspaces



Negative – space of upper bounds $[0, x]$

Positive – space of lower bounds $[x, 1]$

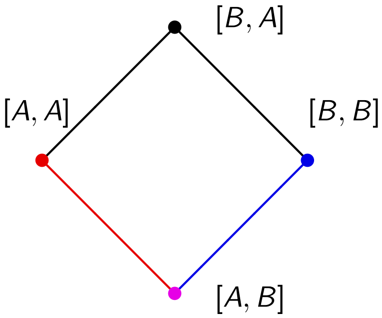
Negative and positive subspaces



Negative – space of upper bounds $[A, x]$

Positive – space of lower bounds $[x, B]$

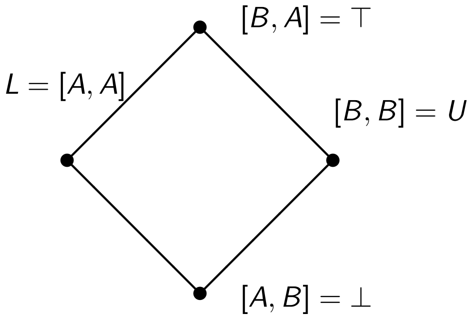
Negative and positive subspaces



We require $A < B$.

We can even allow $A = -\infty, B = \infty$.

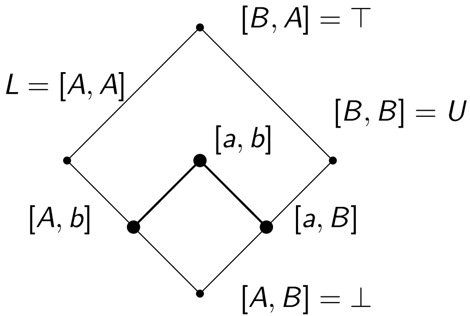
Negative and positive subspaces



We require $A < B$.

We can even allow $A = -\infty, B = \infty$.

Decomposition into negative and positive subspaces



$$[A, b] = [a, b] \wedge \bot = [a, b] \sqcap L$$
$$[a, B] = [a, b] \vee \bot = [a, b] \sqcap U$$
$$[a, b] = [A, b] \sqcup [a, B]$$

The idea of bilattice

Matthew L. Ginsberg. Multivalued logics: a uniform approach to inference in artificial intelligence. *Computational Intelligence*, 4(3):256–316, 1992.

Free versions of this paper and all papers referenced in subsequent slides can be found online.

Original context

Matthew Ginsberg created a number of successful systems in applied AI.

Bilattices were introduced to provide a unified framework for a variety of practical inference schemes used in AI, such as **non-monotonic inference**, inference with uncertainty, etc.

Besides purely theoretical interest, they seemed to increase modularity and efficiency of inference implementations.

Non-monotonic inference

Plenty of examples of commonsense non-monotonic inference.

E.g. if one knows that Larry is a bird, then one infers that Larry can fly. If one then learns that Larry is actually a pinguin, then one takes this inference back and instead infers that Larry cannot fly. If one then learns that Larry is a magical flying pinguin, then...

Other examples include “negation as failure” in Prolog, etc.

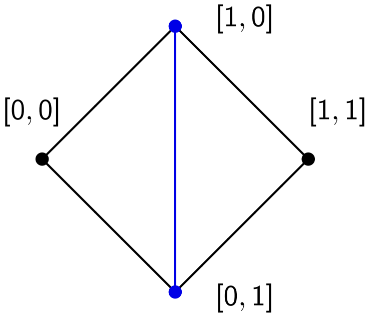
Non-monotonic motives in our work

"True minus" is anti-monotonic.

Consider a partial metric p, and its associated quasi-metric, $q(x, y) = p(x, y) - p(x, x)$.

q is monotonic with respect to one of its arguments, and anti-monotonic with respect to another. This is so annoying in a number of respects that Ralph Kopperman even introduces $q' : A \times A' \to R$ instead of the original $q : A \times A \to R$ in some contexts to mitigate the situation. Here A' is A with reversed order, making q' monotonic with respect to both variables.

Definition of bilattice



The standard definition of bilattice: 1) $\leq$ and $\sqsubseteq$ form complete lattices;

2) an involutive "weak negation" monotonic with respect to $\sqsubseteq$ and antimonotonic with respect to $\leq$ preserving the appropriate lattice structures (in our case, a reflection with respect to the blue line).

Maximal elements are incompatible with group properties

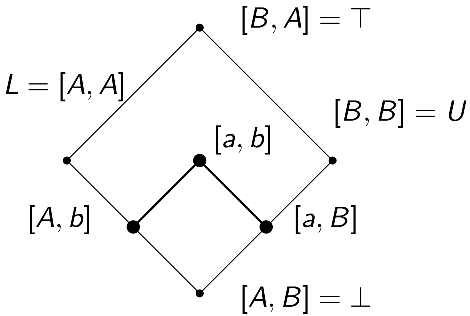
This is not a new situation. For example, a lot of applied math is based on linear algebra, but is implemented with computer representations of real numbers, and these computer representations do not form a group, but one is able to mostly ignore this.

In our case, if we take $A = -\infty, B = \infty$, but omit the segments with infinities, we get a group.

Here we have dropped the requirement that $\leq$ and $\sqsubseteq$ form complete lattices.

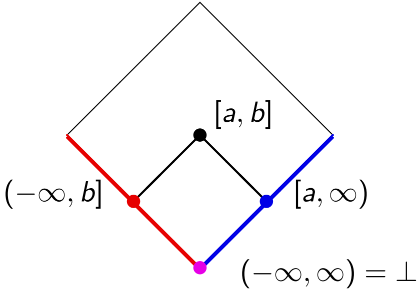
However, the positive and negative subspaces will disappear, and we need them. So let's adjoin their elements, $(-\infty, x]$ and $[x, \infty)$, externally for finite x, and also $\perp = (-\infty, \infty)$.

Decomposition into negative and positive subspaces



$[A, b] = [a, b] \wedge \bot = [a, b] \sqcap L$
$[a, B] = [a, b] \vee \bot = [a, b] \sqcap U$
$[a, b] = [A, b] \sqcup [a, B]$

Bilattice pattern



$(-\infty, b] = [a, b] \wedge \bot$
$[a, \infty) = [a, b] \vee \bot$
$[a, b] = (-\infty, b] \sqcup [a, \infty)$

Bitopology and *d*-frames

Achim Jung, M. Andrew Moshier. On the bitopological nature of Stone duality. Technical Report CSR-06-13. School of Computer Science, University of Birmingham, December 2006, 110 pages.

This text has a lot of very interesting material. I am only touching a bit of it here.

d-frames

Take two frames L_+ and L_- (the informal intent is for their elements correspond to open sets where the predicates are true and where they are false).

$L = L_+ \times L_-$ is a bilattice.

Introduce $Con, Tot \subseteq L$ with the informal intent that for pairs of open sets $U = \langle U_+, U_- \rangle$, $U \in Con$ when $U_+ \cap U_- = \emptyset$, and $U \in Tot$ when $U_+ \cup U_-$ covers the whole space.

This allows to handle partial inconsistency and the bilattice pattern does appear. (L_+, L_-, Con, Tot) is called a d-frame.

Bitopological Stone duality

This paper studies Stone duality modified to apply to bitopological spaces and d-frames.

It also demonstrates that a number of classical dualities, namely dualities of Stone, Ehresmann-Bénabou, and Jung-Sünderhauf, actually have bitopological nature, namely they are special cases of the Stone duality between bitopological spaces and d-frames.

Bitopological Stone duality

Jung and Moshier explain why specialization orders point in the opposite directions in the bitopological situations we tend to encounter. Namely, the property that the intersection of the two specialization orders in question is the equality relation is a corollary of bitopological sobriety when both topologies are T_0.

A bitopology with two specialization orders pointing in the opposite directions is what seems to be required to handle antimonotonic functions well.

Current events

"Marie Curie IEF Project: Bilattices meet d-frames" by Umberto Rivieccio and Achim Jung:

http://www.cs.bham.ac.uk/∼rivieccu/bmdf.html

25th European Summer School in Logic, Language and Information (ESSLLI 2013), Heinrich Heine University, Düsseldorf, Germany, August 5-16, 2013. Course "From bilattices to d-frames" by Umberto Rivieccio and Achim Jung (Logic and Computation, Advanced). See Week 1 Slot 4 at:

http://esslli2013.de/schedule/

Semantics of probabilistic programs

Dexter Kozen. Semantics of probabilistic programs. *Journal of Computer and System Sciences*. **22** (3), 1981, pp. 328–350.

Programs are understood as transformers of probability distributions over spaces of input data.

Kozen finds it convenient to generalize to transformers of signed measures over spaces of input data, which introduces vector spaces and makes it applicable to our situation.

Hahn-Jordan decomposition

Every signed measure μ has unique decomposition into the difference of positive measures μ^+ and μ^-, such that there is no set A with both $\mu^+(A)$ and $\mu^-(A)$ being non-zero: $\mu = \mu^+ - \mu^-$.

The total variation norm, $\|\mu\| = \sup_A \mu^+(A) + \sup_B \mu^-(B)$, makes the space of measures with bounded $\|\mu\|$ a Banach space.

Partial order on the space of signed measures

One takes positive measures as the positive cone in this space.

Our partial order: $\nu < \mu$ iff $\mu - \nu$ is a positive measure.

This is a vector lattice (a Riesz space) and a Banach space, so people call this a Banach lattice.

Bilattice pattern on the space of signed measures

$\mu^+ = \mu \vee 0$ (0 is the zero measure)
$-\mu^- = \mu \wedge 0$

There are two ways to think about $\mu = \mu^+ - \mu^-$.

We can just say that $\mu = \mu^+ + (-\mu^-)$ looks sufficiently similar to our earlier formulas to constitute a bilattice pattern.

Or we can define $\nu \sqsubseteq \mu$ if $\nu^+ \leq \mu^+$ and $\nu^- \leq \mu^-$, and then $\mu = \mu^+ \sqcup (-\mu^-)$, and then it is obviously a bilattice pattern.

(Note: Matthew Ginsberg denotes $\sqcup$ as $+$ (and $\top$ as $\perp$, while denoting $\perp$ as *u*) which might lead to extra confusion here).

Probabilistic semantics via linear operators

Denotations of programs are taken to be continuous linear operators on a Banach space of signed measures with finite norms.

Probabilistic power domain is embedded into the positive cone, and $\leq$ plays the role of Scott's $\sqsubseteq$.

In general, constructions from denotational semantics can be transferred from Scott domains into this setting, and one can iterate them for higher types with what seem to be relatively mild technical complications, although I am not aware of the limits (meaning reflexive domains) having been studied in this setting.

Distances between programs

On one hand, Anthony Seda and Máire Lane note that there is a natural norm in this situation, which allows to define conventional metric.

Anthony Seda, Máire Lane. On continuous models of computation: towards computing the distance between (logic) programs. *IWFM'03 Proceedings*, 2003.

On the other hand, in this context, where everything is a function of a measure, the popular constructions of partial metrics over Scott domains which tend to be parametrized by measures look quite natural (we tend to view their dependency from a measure as an obstacle which needs to be overcome, but perhaps it is actually a desirable feature).

Possible worlds indexed by measures

A paper by William Wadge on intensional logic and data flow programming explores various ways to index possible worlds:

William Wadge, Intensional Logic in Context, in *Intensional Programming II: based on the papers at Islip 99*, Manolis Gergatsoulis and Panos Rondogiannis, Eds., pp. 1–13, World Scientific, 2000.

In our case, possible worlds would be indexed by measures, which is quite attractive and feels natural (a world is distinguished by how often one sees various things, and we do sampling to figure out what kind of world we currently inhabit).

More sophisticated applications?

The resulting setup does look promising for our chances to extend various methods of applied math based on linear algebra and functional analysis to spaces of programs.

Without claiming any real progress here, I'd like to sketch a tiny bit which I seem to understand somewhat.

Addition of programs

If $0 < \alpha < 1$, and we have programs **P** and **Q** denoting transformers of positive measures *p* and *q*, and a generator **random** of random real numbers uniformly distributed in $[0, 1]$, we know how to write a program denoting $\alpha p + (1 - \alpha) q$.

Namely, the program is: **if random** $<$ α **then P else Q**

Computations with signed measures

We should compute separately for positively and negatively valued components.

If we do that, then taking the minus is done by simply swapping components marked as positive with components marked as negative.

In particular, if we consider sampling-based probabilistic programming (e.g. MCMC-based engines, one can think about them as an unusual form of data flow programming languages), these two-component computations seem natural.

Retina computes with positive and negative components

To illustrate the last point I conclude with an example from:
David Marr, *Vision*, 1982.

One of the more important functions of retina is computing the convolution of various filters of the form $\nabla^2 G$ with the intensity of light. Here ∇^2 is the Laplacian, $\partial^2/\partial x^2 + \partial^2/\partial y^2$, and G is the 2D Gaussian depending on σ, $e^{-\frac{x^2+y^2}{2\pi\sigma^2}}$. Ignoring the differences between individual photons one can intepret the light intensity distribution as the probability distribution for where the next photon would hit.

Experimental data from Marr

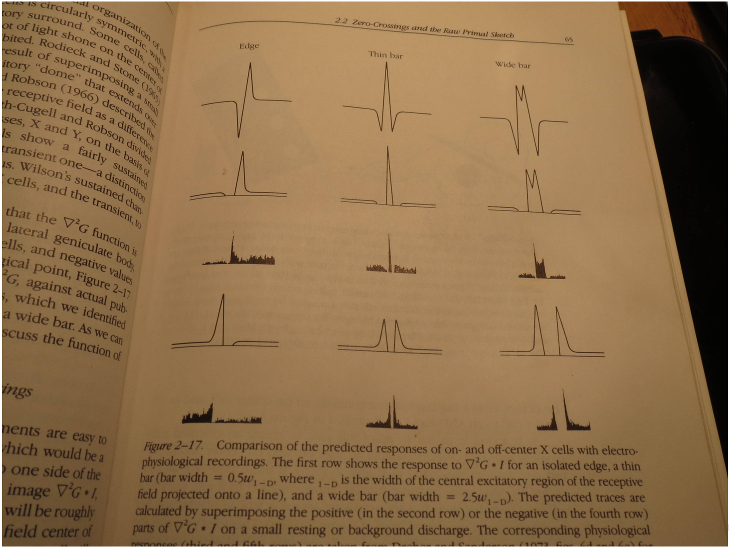


Figure 2–17. Comparison of the predicted responses of on- and off-center X cells with electrophysiological recordings. The first row shows the response to $\nabla^2 G * I$ for an isolated edge, a thin bar (bar width = $0.5w_{1-D}$, where $_{1-D}$ is the width of the central excitatory region of the receptive field projected onto a line), and a wide bar (bar width = $2.5w_{1-D}$). The predicted traces are calculated by superimposing the positive (in the second row) or the negative (in the fourth row) parts of $\nabla^2 G * I$ on a small resting or background discharge. The corresponding physiological

Electronic coordinates

These slides are linked from my new page on partial inconsistency and vector semantics of programming languages:

http://www.cs.brandeis.edu/∼bukatin/partial_inconsistency.html

E-mail:

bukatin@cs.brandeis.edu