Master WIP

Kjetil Midtgarden Golid

February 2, 2016

1 Introduction

1.1 Paradoxes

A theory in propositional logic is semantically *inconsistent* if it has no model, i.e., there exists no variable assignment making all formulae in the theory true. Consider the following example:

$$x \wedge \neg x \tag{1}$$

While a sentence like $\neg a \rightarrow b$ can be satisfied by, for instance, letting both a and b be true, no such assignment can be made for the statement above. The statement is therefore inconsistent.

A paradox is usually informally defined as something along the lines of *"a statement that can be neither true nor false"*. We can immediately note one thing from this intuitive definition: Since no paradoxes can be true, all paradoxes are, by definition, inconsistent. It does, however, seem like a stretch to say that all inconsistent theories are paradoxes. Just consider the inconsistent statement, $x \wedge \neg x$ again: this statement simply seem false, and thus not paradoxical.

A different view is that a paradox is a *dialetheia*, a sentence that is *both* true and false[1]. We will however not spend much time exploring these philosophical differences, being this is not a philosophical paper and it won't change much for our definitions.

The liar sentence is probably the most famous example of a paradox:

$$\textit{"This sentence is false".} \tag{2}$$

If the statement is true, then the statement is false, but if the statement is false, then the statement is true. It can thus neither be true nor false. Note how the liar sentence is a statement about other statements (in this case itself). In order to study these kinds of meta-statements, we need a way to reference other statements within a statement. In propositional logic, we can do this by giving statements "names" in the form of

adding fresh variables with equivalences to their correponding statements[1]. Consider the left statements below, together with their corresponding named statements on the right.

$$a \qquad\qquad x_1 \leftrightarrow a \qquad\qquad (3)$$

$$a \wedge \neg a \qquad\qquad x_2 \leftrightarrow a \wedge \neg a \qquad\qquad (4)$$

$$a \vee \neg a \qquad\qquad x_3 \leftrightarrow a \vee \neg a \qquad\qquad (5)$$

By performing this naming-operation on these statements, one is obviously changing their truth value. Even though we have one consistent, one inconsistent and one tautological statement on the left, all the statements become consistent after they have been named. This is because we in all the cases above can find a truth value for x_i that matches the one of the corresponding statement. This will not be the case for paradoxes, so our new named statements will be consistent if and only if they are not paradoxical.

Consider the liar sentence. It can be written as a named statement in the following way:

$$x \leftrightarrow \neg x \qquad\qquad (6)$$

This statement is obviously inconsistent, making it a paradox by our newly acquired definition. The study of *discourses* takes this formalization a step further.

1.2 Discourses

A propositional theory is in **graph normal form (GNF)** if all its formulae is on the following form:

$$x \leftrightarrow \bigwedge_{i \in I_x} \neg y_i \qquad\qquad (7)$$

such that every variable occurs exactly once on the left of $\leftrightarrow$ across all the formulae in the theory.

There is a simple translation from conjunctive normal from to graph normal form (shown in the end of this section), showing that any propositional theory can be written in GNF. By interpreting the variable on the left as the name of the statement on the right, like shown earlier, one can start using GNF to model meta-statements.

We now formally define a **discourse** to be a theory in GNF, and a **paradox** to simply be an inconsistent discourse.

We will later show a very handy correspondence between these discourse theories and certain graphs. The correspondence lets us not only decide the satisfiability of a discourse

---

[1]bad wording?

theory (i.e. whether or not it is paradoxical) by looking at certain properties of the corresponding graph. The properties in the graph also provides us with the satisfying models, if they exist. In order to express this logic/graph correspondence, we first need to establish some graph terminology.

1.3 Graphs, Kernels and Solutions

A directed graph is a pair $\mathbf{G} = \langle G, N \rangle$ where G is a set of vertices while $N \subseteq G \times G$ is a binary relation representing the edges in $\mathbf{G}$. We use the notation $N(x)$ to denote the set of all vertices that are targeted by edges originating in x (successors of x). Similarly, $N^-(x)$ denotes the set of all vertices with edges targeting x (predecessors of x. We define these two predicates formally as follows:

$$N(x) := \{y \mid (x,y) \in N\} \tag{8}$$

$$N^-(x) := \{y \mid (y,x) \in N\} \tag{9}$$

We extend these relations to sets of vertices in the following way:

$$N(X) = \bigcup_{x \in X} N(x) \tag{10}$$

$$N^-(X) = \bigcup_{x \in X} N(x) \tag{11}$$

A kernel is a set of vertices $K \subseteq G$ such that:

$$G \setminus K = N^-(K) \tag{12}$$

The above equivalence can ble split up into two inclusions to be more easily understood:

$G \setminus K \subseteq N^-(K)$, saying that each vertex outside the kernel, has to have an edge into the kernel (K is dominating/absorbing).

$N^-(K) \subseteq G \setminus K$, saying that each edge targeting a vertex within the kernel has to come from outside, thus no two vertices in the kernel are connected by an edge (K is independent).

Kernels heve been of great interest over several decades, mainly within the fields of game theory and economics[**neumann**]. In a graph representing some sort of a turn-based game, where vertices are states and edges are transitions, one can often work out winning strategies whenever one finds a kernel in the graph. Observe that whenever one is outside of the kernel, one always has the possibility of moving inside the kernel (since the kernel is absorbing), while inside the kernel one *has* to move out of the it (since the kernel is independent). If you are the player with the choice outside the kernel, you can control the game and choose to stabilize it by always moving inside the kernel, forcing the opponent to move out again.

Deciding the existence of kernels in a graph has been shown to be an NP-complete[2] problem. This should not be surprising, since we are in the middle of showing the equivalence between this problem and the problem of finding satisying models of PL theories (SAT), which we know is NP-complete.

We will get the correspondence between satisfying models of a discourse theory and kernels in a graph through an alternative, equivalent kernel definition called a *solution*. The equivalence of kernels and solutions was shown by Roy Cook in [3].

Given a directed graph $\mathbf{G} = \langle G, N \rangle$, an assignment $\alpha \in 2^G$ is a function mapping every vertex in the graph to either 0 or 1. A **solution** is an assignment α such that for all $x \in G$:

$$\alpha(x) = 1 \iff \alpha(N(x)) = \{0\} \tag{13}$$

In simple words, this means that for any node x, if x is assigned to 1, then all its successors has to be assigned to 0, and if x is assigned to 0, then there has to exist a node assigned to 1 among its successors. A consequence of this definition is that all sink nodes (nodes with no outgoing edges) in the graph have to be assigned to 1, since it vacuously does not point to any node assigned to 1. We use the notation $sol(\mathbf{G})$ to denote the set of all solutions of the graph $\mathbf{G}$.

## 1.4 Discourse Theories and Digraphs

As mentioned earlier, there is a close connection between (1) models of a discourse theory, (2) kernels of a graph and (3) solutions of a graph. While Roy Cook has given us the equality of (2) and (3), we will now look at two functions connecting (1) and (2). We get the following definitions from []

$\mathcal{T}$: translating a digraph $\mathbf{G}$ into a corresponding theory $\mathcal{T}(\mathbf{G})$ such that $sol(\mathbf{G}) = mod(\mathcal{T}(G))$.

$\mathcal{G}$: translating a theory T into a corresponding digraph $\mathcal{G}(T)$ such that $mod(T) = sol(\mathcal{G}(T))$.

Given any digraph $\mathbf{G}$ we get the theory $\mathcal{T}(\mathbf{G})$ by taking, for each $x \in G$, the formula $x \leftrightarrow \bigwedge_{y \in N(x)} \neg y$ where $\bigwedge \emptyset = 1$.

Example 1.
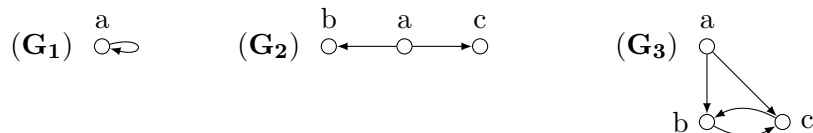
Using the graphs from above, we get the following theories using $\mathcal{T}$:

$$\mathcal{T}(\mathbf{G_1}) = \{a \leftrightarrow \neg a\} \tag{14}$$

$$\mathcal{T}(\mathbf{G_2}) = \{a \leftrightarrow (\neg b \wedge \neg c), b, c\} \tag{15}$$

$$\mathcal{T}(\mathbf{G_3}) = \{a \leftrightarrow (\neg b \wedge \neg c), b \leftrightarrow \neg c, c \leftrightarrow \neg b\} \tag{16}$$

We will not be proving that $sol(\mathbf{G}) = mod(\mathcal{T}(G))$, but observe that $\mathbf{G_1}$ has no solution, just like its corresponding theory $\mathcal{T}(\mathbf{G_1})$ has no satisfying models. $\mathbf{G_2}$ has one solution, where one assigns $a = 0, b = 1, c = 1$. This assignment also works as a satisfying model for $\mathcal{T}(\mathbf{G_2})$. In $\mathbf{G_3}$, we get two solutions, both with a assigned to 0, but with 0 and 1 distributed on b and c. These are also the only two models satisfying $\mathcal{T}(\mathbf{G_3})$. It is generally true that $\mathcal{T}$ gives us our requested correspondence[4].

Conversely, given any discourse theory T (in fact, this will work given any PL theory, since we can translate CNF to GNF), we can derive the corresponding graph $\mathcal{G}(T)$ in the following way: All variables in the theory are vertices, and for each formula $x \leftrightarrow \bigwedge_{i \in I_x} y_i$ make a directed edge $\langle x, y_i \rangle$ for each $i \in I_x$.

Example 2.

$$T_1 = \neg a \qquad \Longleftrightarrow \quad (a \leftrightarrow \neg a'), (a' \leftrightarrow \neg a), (y_1 \leftrightarrow (\neg a' \wedge \neg y_1)) \tag{17}$$

$$T_2 = a \vee b \quad \Longleftrightarrow \quad (a \leftrightarrow \neg a'), (a' \leftrightarrow \neg a), (b \leftrightarrow \neg b'), (b' \leftrightarrow \neg b), (y_1 \leftrightarrow (\neg a \wedge \neg b \wedge \neg y_1)) \tag{18}$$

Using $\mathcal{G}$ on the above theories – translated to GNF – gives us the following graphs:
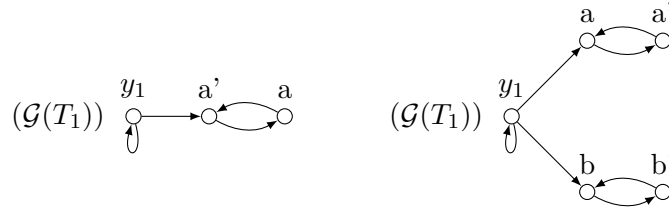
Again, will we not be proving the correspondence, but notice that T_1 has one model satisfying it, where $a = 0$. $\mathcal{G}(T_1)$ also has one solution, namely where $a = 0, a' = 1, y_1 = 0$. T_2 has three solutions, where either a, b or both are assigned to 1. This reflects onto the graph since y_1 has to be assigned to 0, thus forcing a or b to be assigned to 1. It is generally true that $\mathcal{G}$ gives us the correspondence we are looking for[4].

With the problem of solutions being equal to the problem of models, we get our final equivalence between kernels and models. This equivalence connects the fields of logic with the fields of graph theory: A graph has a kernel if and only if its corresponding discourse is consistent (non-paradoxical). A theory is paradoxical if and only if its corresponding graph has no kernels. Because of this tight link, we will often refer to graphs without kernels as paradoxical graphs.

The applicability of kernels should by now be obvious. In the next section we will explore some of the various findings within Kernel Theory, and especially the findings related to infinitary graphs.

## 2   Results in Kernel Theory

Ultimately, the end goal of Kernel Theory would be to have an easy way to answer the question "Does this digraph have a kernel?" no matter the graph, and no matter the answer. We are not quite there, but a lot of work has been put into trying to identify special circumstances under which one is guaranteed to have (or guaranteed to not have) a kernel in the given graph. One of the results is the Richardson's Theorem, worked out by Moses Richardson in 1953:

Theorem 3. [5] If D is a finitary[2] digraph without odd cycles, then D has a kernel.

This theorem gives us the confirmation that whenever dealing with finitary dags, for instance, one can be certain that its corresponding theory is consistent.

Intuitively, one is tempted to believe that *all* digraphs without odd cycles have kernels, but this is not the case. Until now, our paradoxes have always been statements that – directly or indirectly – have been referring back to themselves (giving cycles in the graph) and thus causing a logical conflict, and it is hard to imagine any other way to construct paradoxical statements. The following construction will however reveal our lack of imagination.

The **Yablo Graph** [6] is an example of an acyclic graph with no kernel. It is constructed with an infinite set of vertices $\{x_i | i \in \mathbb{N}\}$ and a set of edges N such that $\langle x_i, x_j \rangle \in N$ iff $i < j$.
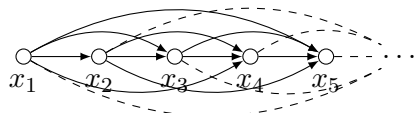
Since there exist no two numbers $x, y \in \mathbb{N}$ such that $x < y$ and $y < x$, we get that the Yablo graph indeed is acyclic[3]. Furthermore, since any natural number has infinitely many numbers strictly larger than it, we get that all the vertices are infinitely branching, making the Yablo graph infinitary (not finitary).

The corresponding discourse theory of the Yablo graph would – informally – be the situation with an infinite number of statements, all saying "Every statement after this statement is false".

---

[2]In a finitary graph, every vertex has a finite number of out-neighbors; the graph has finite branching.
[3]This is a bit over-simplified

We will later show that this graph is indeed without a kernel, but for now we will take Yablo's word [6] for it.

One thing should be mentioned at this point; neither odd cycles nor infinitely branching vertices *entails* that their respective graphs are paradoxical. The two following graphs illustrate this point:

$$x_0$$

(19)

$$x_1 \qquad x_2$$

The above graph contains an odd cycle, but the singleton set $\{x_2\}$ is a kernel.

$$x_0$$

(20)

$$\ldots \qquad x_3 \qquad x_2 \qquad x_1$$

The above graph is has an infinitely branching vertex x_0, but the infinite set $\{x_i \mid x > 0\}$ is a kernel.

It is shown that every digraph (with at least one edge) can be transformed into a infinitary dag[4] such that α is a solution to the created dag if and only if it is a solution to the original digraph [4]. This means that for any finitary graph that is paradoxical by the virtue of having an odd cycle, there is an infinitary, *acyclic* digraph that is also paradoxical. Furthermore, if one is trying to find ways to identify paradoxical graphs, one does only need to look at dags. The reason for this is simple; say you have a method of identifying paradoxical dags. Then given any graph, you will be able to decide whether or not it is paradoxical. If it is a dag, you use your method. If it is cyclic, translate it, then use your method. Since the translation preserves the solutions, one can be sure that the cyclic graph is paradoxical if and only if the resulting dag was paradoxical.

This result will be of great importance to us, enabling us to look only at dags when searching for paradoxes.

---

[4]Directed acyclic graph

2.1 Proofs

2.1.1 Translating CNF to GNF

Since any PL theory can be expressed in CNF, showing that any theory P in CNF can be translated to a theory R in GNF such that $P \leftrightarrow R$ gives us that any PL theory can be expressed in GNF.

Given any CNF theory P, start with an empty theory R and for each formula in P, follow the steps below to acquire its corresponding GNF formulae.

Step 1: For each literal $\neg x_i$ in the formula, introduce a fresh variable x_i' and add the following two GNF formulae to R: $x_i' \leftrightarrow \neg x_i, x_i \leftrightarrow \neg x_i'$, (unless this has already been done while translating an earlier formula in the theory).

Step 2: In each clause, replace every negative literal $\neg x_i$ with its corresponding x_i' fom step 1. Every clause does now contain all positive literals. For every clause $(x_1 \lor x_2 \lor \cdots \lor x_n)$, create a fresh variable y and add the following GNF formula to R:

$$y \leftrightarrow (\neg x_1 \land \neg x_2 \land \cdots \land \neg x_n \land \neg y)$$

The combined set of all these acquired formulae will make up the the corresponding theory R. We have only added proper GNF formulae and all variables appear to the left in exactly one clause, so R will indeed be a GNF theory.

Example 4.

$$\text{CNF:} \quad (a \lor b) \tag{21}$$
$$\text{GNF:} \quad (a \leftrightarrow \neg a'), (a' \leftrightarrow \neg a), (b \leftrightarrow \neg b'), (b' \leftrightarrow \neg b), (y_1 \leftrightarrow (\neg a \land \neg b \land \neg y_1)) \tag{22}$$
$$\tag{23}$$
$$\text{CNF:} \quad (\neg a) \tag{24}$$
$$\text{GNF:} \quad (a \leftrightarrow \neg a'), (a' \leftrightarrow \neg a), (y_1 \leftrightarrow (\neg a' \land \neg y_1)) \tag{25}$$
$$\tag{26}$$

References

[1] G. Priest and F. Berto, "Dialetheism," in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., Summer 2013, 2013. [Online]. Available: `http://plato.stanford.edu/archives/sum2013/entries/dialetheism/`.

[2] V. Chvátal, "On the computational complexity of finding a kernel. technical report crm-300," Centre de Recherches Mathématiques, Univeristé de Montréal, Tech. Rep., 1973.

[3] R. T. Cook, "Patterns of paradox," *The Journal of Symbolic Logic*, vol. 69(3), pp. 767–774, 2004.

[4] M. Bezem, C. Grabmeyer, and M. Walicki, "Expressive power of digraph solvability," *Annals of Pure and Applied Logic*, vol. 163, pp. 200–213, 2012.

[5] M. Richardson, "Solutions of irreflexive relations," *The Annals of Mathematics, Second Series*, vol. 58(3), pp. 573–590, 1953.

[6] S. Yablo, "Paradox without self-reference," *Analysis*, vol. 53(4), pp. 251–252, 1993.