

---

# ANALYTICAL SOLUTIONS FOR THE INVERSE PROBLEM WITHIN GRADUAL SEMANTICS

---

Nir Oren, Bruno Yun, Assaf Libman, Murilo S. Baptista

University of Aberdeen  
Scotland

{n.oren, bruno.yun, a.libman, murilo.baptista}@abdn.ac.uk

Srdjan Vesic

CRIL, Univ. Artois  
France

vesic@cril.fr

## ABSTRACT

Gradual semantics within abstract argumentation associate a numeric score with every argument in a system, which represents the level of acceptability of this argument, and from which a preference ordering over arguments can be derived. While some semantics operate over standard argumentation frameworks, many utilise a weighted framework, where a numeric initial weight is associated with each argument. Recent work has examined the inverse problem within gradual semantics. Rather than determining a preference ordering given an argumentation framework and a semantics, the inverse problem takes an argumentation framework, a gradual semantics, and a preference ordering as inputs, and identifies what weights are needed to over arguments in the framework to obtain the desired preference ordering. Existing work has attacked the inverse problem numerically, using a root finding algorithm (the bisection method) to identify appropriate initial weights. In this paper we demonstrate that for a class of gradual semantics, an analytical approach can be used to solve the inverse problem. Unlike the current state-of-the-art, such an analytic approach can rapidly find a solution, and is guaranteed to do so. In obtaining this result, we are able to prove several important properties which previous work had posed as conjectures.

## 1 Introduction

Standard approaches to abstract argumentation consider a set of atomic arguments and the interactions between them, encoding these as a graph. They then identify which sets of arguments are justified together by considering the inter-argument interactions [1]. Within the argumentation community, there has been increasing interest in so-called *ranking-based* semantics. These latter semantics aim to identify a ranking over the arguments, with higher ranked arguments considered more justified (i.e. “less attacked”) than arguments ranked lower (i.e. “more attacked”). Many ranking-based semantics associate numerical *acceptability degrees* to all arguments within the system, with the final ranking depending on the associated numerical ordering. Furthermore, some ranking semantics compute the final acceptability degree of an argument based not only on the topology of the argumentation graph, but also based on some *initial weight* assigned to the argument.

Our focus in this paper is on the *inverse problem* found in ranking-based semantics. That is, rather than describing how an ordering over arguments can be obtained from some set of arguments and their associated properties, we ask *what properties must be associated with arguments so as to derive some desired final acceptability ordering*. This problem was previously tackled by [2], who employed a numerical approach to solve it. In their work, the authors focused on three specific semantics, which also serve as the focus of this paper. Our core contribution involves describing an analytic approach to solving the inverse problem for these three semantics. The advantages of pursuing such an analytic approach are twofold. First, it guarantees our ability to find a (unique) solution; and second, it is efficient in finding this solution.

Apart from describing an analytic approach to the inverse problem, we make advance the state-of-the-art in several ways. First, building on ideas from [3, 4, 5] we describe a very general class of functions, which can be used to underpin gradual argumentation semantics, and demonstrate the existence of a unique fixed-point for such functions. We also demonstrate continuity for this class of functions. Focusing on the h-categorizer gradual semantics [6], we prove that

	$a_0$	$a_1$	$a_2$	$a_3$	Argument ranking
$\sigma_{MB}^{\mathcal{F}}$	0.43	0.30	0.58	0.30	$a_1 \simeq a_3 \triangleleft a_0 \triangleleft a_2$
$\sigma_{HC}^{\mathcal{F}}$	0.43	0.30	0.38	0.30	$a_1 \simeq a_3 \triangleleft a_2 \triangleleft a_0$
$\sigma_{CB}^{\mathcal{F}}$	0.43	0.18	0.17	0.30	$a_2 \triangleleft a_1 \triangleleft a_3 \triangleleft a_0$

Table 1: Acceptability degrees of the arguments from Figure 1

this semantics obeys monotonicity, a conjecture made in [2] which is required for their numerical approach to operate. With these results, we are able to prove several other important properties hitherto not discussed in the literature.

The remainder of this paper is structured as follows. In the Section 2 we introduce the semantics for which we address the inverse problem. In Section 3, we consider a generalisation of several gradual semantics, demonstrating that all semantics in this generalisation have a unique fixed point and are continuous. In Section 4 we expand on the approach advanced in [2], demonstrating how it can applied to the more generalised semantics. Following this, Section 5 details our analytic approach and proves some important properties. Section 6 concludes and discusses potential future work.

## 2 Background

We begin by providing an overview of the three gradual semantics around which our work revolves. These semantics all operate over *weighted argumentation frameworks* [6, 7].

**Definition 2.1** (WAF). A weighted argumentation framework (WAF) is a triple  $\mathcal{F} = \langle \mathcal{A}, \mathcal{D}, w \rangle$ , where  $\mathcal{A}$  is a finite set of arguments,  $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$  is a binary attack relation, and  $w : \mathcal{A} \rightarrow [0, 1]$  is a weighting function assigning an initial weight to each argument.

We denote the set of attackers of an argument  $a \in \mathcal{A}$  as  $\text{Att}(a) = \{b \in \mathcal{A} \mid (b, a) \in \mathcal{D}\}$ . While myriad ranking based-semantics have been described (see e.g., [8]), we concentrate on three of the semantics described in [4] which allow for an initial weight to be assigned to an argument. This is because the inverse problem we focus on considers what initial weight needs to be assigned to obtain some preference ordering.

**Definition 2.2** (Gradual Semantics). A gradual semantics  $\sigma$  is a function that associates to each weighted argumentation graph  $\mathcal{F} = \langle \mathcal{A}, \mathcal{D}, w \rangle$ , a scoring function  $\sigma^{\mathcal{F}} : \mathcal{A} \rightarrow [0, 1]$  that provides an acceptability degree to each argument. We consider the three semantics  $\sigma_x$ , for  $x \in \{MB, CB, HC\}$ , defined as follows.

- The *weighted max-based* semantics  $\sigma_{MB}$  [4] is defined such that the acceptability degree of an argument  $a \in \mathcal{A}$  is  $\sigma_{MB}^{\mathcal{F}}(a) = MB_{\infty}(a)$ , where  $MB_i(a) = \frac{w(a)}{1 + \max_{b \in \text{Att}(a)} MB_{i-1}(b)}$  and for all  $b \in \mathcal{A}$ ,  $MB_0(b) = w(b)$ .
- The *weighted card-based* semantics  $\sigma_{CB}$  [4] is defined such that the acceptability degree of an argument  $a \in \mathcal{A}$  is  $\sigma_{CB}^{\mathcal{F}}(a) = CB_{\infty}(a)$  where  $CB_i(a) = \frac{w(a)}{1 + |\text{Att}^*(a)| + \frac{\sum_{b \in \text{Att}^*(a)} CB_{i-1}(b)}{|\text{Att}^*(a)|}}$ , for all  $b \in \mathcal{A}$ ,  $CB_0(b) = w(b)$ , and  $\text{Att}^*(a) = \{b \in \text{Att}(a) \mid w(b) > 0\}$  if  $\text{Att}^*(a) \neq \emptyset$  and  $w(a)$  otherwise.
- The *weighted h-categorizer* semantics  $\sigma_{HC}$  [4] is defined such that the acceptability degree of an argument  $a \in \mathcal{A}$  is  $\sigma_{HC}^{\mathcal{F}}(a) = HC_{\infty}(a)$  where  $HC_i(a) = \frac{w(a)}{1 + \sum_{b \in \text{Att}(a)} HC_{i-1}(b)}$  and for all  $b \in \mathcal{A}$ ,  $HC_0(b) = w(b)$ .

**Example 2.3.** (taken from [2]). Let  $\mathcal{F} = \langle \mathcal{A}, \mathcal{D}, w \rangle$  be a WAF, where  $\mathcal{A} = \{a_0, a_1, a_2, a_3\}$ ,  $\mathcal{D} = \{(a_0, a_2), (a_1, a_1), (a_1, a_2), (a_2, a_2), (a_3, a_2)\}$ ,  $w(a_0) = 0.43$ ,  $w(a_1) = 0.39$ ,  $w(a_2) = 0.92$ , and  $w(a_3) = 0.3$ . The WAF is represented in Figure 1 whereas the acceptability degrees and the associated rankings on arguments for the semantics of Definition 2.2 are shown in Table 1.

Note that we use the following notation for the ordering on arguments, where  $a \triangleright b$  denotes that  $a$  is at least as preferred as  $b$ ,  $a \simeq b$  iff  $a \triangleright b \wedge b \triangleright a$ ,  $a \triangleright b$  iff  $a \triangleright b \wedge a \not\triangleright b$ , and  $a \triangleleft b$  iff  $a \not\triangleright b$ .

## 3 Kernel-based Gradual Semantics

Pu [3] demonstrated that a unique fixed-point exists for the h-categorizer semantics, i.e. in the case where all arguments have an initial weight of 1. This result was generalised to the three semantics described in Definition 2.2 by Amgoud et

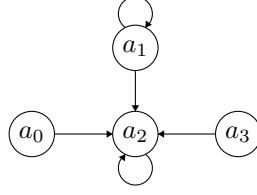


Figure 1: Graphical representation of a WAF

al. [4]. In this section, we describe a more general class of semantics for which such a unique fixed-point is guaranteed to exist. Our core result is as follows.

**Theorem 3.1.** *We say that  $\varphi_i(x) : [0, 1]^n \rightarrow [0, \infty)$ , where  $i = 1, \dots, n$ , is an argumentation kernel function if  $\varphi_i$ :*

- *is continuous*
- *is monotonic, i.e., for all  $x, y \in [0, 1]^n$  where  $x \preceq y^1$ ,  $\varphi_i(x) \leq \varphi_i(y)$*
- *is homogeneous, i.e.,  $\varphi_i(tx) = t\varphi(x)$  for any  $x \in p0, 1]^n$  and every  $0 \leq t \leq 1$*

*If  $\Phi_w(x)_i = \frac{w_i}{1+\varphi_i(x)}$ , then  $\Phi_w : [0, 1]^n \rightarrow [0, 1]^n$  has a unique fixed point.*

We can consider the class of all gradual semantics  $\sigma$  that takes as input a WAF  $\mathcal{F} = \langle \{a_1, a_2, \dots, a_n\}, \mathcal{D}, w \rangle$  and return a scoring function  $\sigma^{\mathcal{F}}$  such that for all  $x \in [0, 1]^n$ , for all  $a_i \in \mathcal{A}$ , there exists  $\Phi_w$  (as defined in Theorem 3.1) such that  $\Phi_w(x)_i = \sigma^{\mathcal{F}}(a_i)$ . The previous theorem states that all the gradual semantics in the aforementioned class are guaranteed to “converge”, including the previous three semantics of Definition 2.2. Indeed, they are all in this class with  $\varphi_i(x) = \max_{a_j \in \text{Att}(a_i)} x_j$ ,  $\varphi_i(x) = 1 + |\text{Att}^*(a_i)| + (\sum_{a_j \in \text{Att}^*(a_i)} x_j) / |\text{Att}^*(a_i)|$ , and  $\varphi_i(x) = \sum_{a_j \in \text{Att}(a_i)} x_j$  respectively for  $\sigma_{MB}$ ,  $\sigma_{CB}$ , and  $\sigma_{HC}$ . Of course, our result is more general and one could consider many other semantics in this class based on other argumentation kernel functions. Other examples of argumentation kernels include the geometric mean, i.e.  $\varphi_i(x) = \sqrt[p]{b_i x_1 \cdots x_n}$  for some  $b_i > 0$ , or the  $L^p$ -norm<sup>2</sup>, i.e.  $\varphi_i(x) = \sqrt[p]{\sum_j |x_j|^p}$  for some  $1 \leq p < \infty$ , among others. More generally, we can combine argumentation kernels together as the collection of argumentation kernel functions is:

1. *closed under linear combination with non-negative coefficients*, i.e. if  $\varphi_1, \dots, \varphi_k$  are argumentation kernels and  $\lambda_1, \dots, \lambda_k \geq 0$  then  $\lambda_1 \varphi_1 + \dots + \lambda_k \varphi_k$  is an argumentation kernel.
2. *closed under geometric*, i.e. if  $\varphi_1, \dots, \varphi_k$  are argumentation kernels then  $\sqrt[k]{\varphi_1(x) \cdots \varphi_k(x)}$  is an argumentation kernel.
3. *closed under limits*, i.e. if  $\varphi_k$  is a sequence of argumentation kernels such that the limit  $\varphi := \lim_k \varphi_k$  exists and continuous (e.g the convergence is uniform), then  $\varphi$  is an argumentation kernel.

The following theorem builds on the following result, adapted from [3].

**Theorem 3.2.** *Let  $X = [0, 1]^n$ , and let  $f : X \rightarrow X$  be a function which*

- *is order reversing, i.e.,  $x \preceq y$  implies  $f(y) \preceq f(x)$*
- *there is some  $0 < \alpha \leq 1$  such that for any  $x \in [0, 1]^n$  and any  $0 \leq t \leq 1$*

$$f(tx) \preceq \frac{1}{t + \alpha(1-t)} f(x)$$

*then  $f$  has a unique fixed point  $y \in [0, 1]^n$ . Furthermore,  $y = \lim_{k \rightarrow \infty} x^{(k)}$  where  $x^{(k)}$  is any sequence defined recursively by choosing  $x^{(0)} \in X$  arbitrarily and  $x^{(k+1)} = f(x^{(k)})$ .*

*Proof.* Define a sequence  $u^{(n)}$  in  $X$  by recursion,  $u^{(0)} = 0$  and  $u^{(n+1)} = f(u^{(n)})$  for every  $n \geq 0$ . Then  $u^{(0)} \preceq u^{(1)}$  since  $0 \in X$  is the minimum of  $X$ , and since  $f$  satisfies the second item of Theorem 3.2, we get  $u^{(0)} \preceq u^{(2)} \preceq u^{(1)}$ .

<sup>1</sup>We equip  $\mathbb{R}^n$  with the partial order  $x \preceq y$  if  $x_i \leq y_i$  for all  $i = 1, \dots, n$ . We will also write  $x \preceq 0$  to denote  $x_i \leq 0$  for all  $i$ .

<sup>2</sup>We will also use the notation  $\|x\|_p$  to denote the  $L^p$  norm of  $x$ . For  $p = \infty$  we obtain the max-norm, i.e.  $\|x\|_\infty = \max_i |x_i|$ .

By induction one easily shows

- (i)  $u^{(2k)} \preceq u^{(2k-1)}$  for all  $k \geq 1$ ,
- (ii)  $u^{(2k)} \preceq u^{(2k+2)}$  for all  $k \geq 0$ ,
- (iii)  $u^{(2k+1)} \preceq u^{(2k-1)}$  for all  $k \geq 1$ .

Thus, the sequence  $u^{(2k)}$  is increasing (in the sense that  $u_i^{(2k)}$  is an increasing sequence in  $\mathbb{R}$  for every  $i = 1, \dots, n$ ) and bounded above by  $1 \in X$ , and  $u^{(2k-1)}$  is decreasing and bounded below by  $0 \in X$ , so we define

$$\begin{aligned} u^{(\text{ev})} &= \sup_k u^{(2k)} = \lim_k u^{(2k)} \\ u^{(\text{odd})} &= \inf_k u^{(2k-1)} = \lim_k u^{(2k-1)}. \end{aligned}$$

It follows from (i) that  $u^{(\text{ev})} \preceq u^{(\text{odd})}$ . Our next goal is to show that  $u^{(\text{ev})} = u^{(\text{odd})}$ .

For every  $k \geq 1$  set

$$\pi_k = \sup \{t : t \cdot u^{(2k-1)} \preceq u^{(2k)}, 0 \leq t \leq 1\}.$$

The set on the right is not empty since  $0 \cdot u^{(2k-1)} = 0 \preceq u^{(2k)}$ , so

$$0 \leq \pi_k \leq 1$$

and by construction

$$\pi_k \cdot u^{(2k-1)} \preceq u^{(2k)}.$$

Since  $f$  is order reversing, (ii) implies that for every  $k \geq 1$

$$\begin{aligned} u^{(2k+1)} &= f(u^{(2k)}) \preceq \frac{1}{\pi_k + \alpha(1 - \pi_k)} \cdot f(u^{(2k-1)}) \\ &= \frac{1}{\pi_k + \alpha(1 - \pi_k)} \cdot u^{(2k)} \preceq \frac{1}{\pi_k + \alpha(1 - \pi_k)} \cdot u^{(2k+2)}. \end{aligned}$$

Hence  $(\pi_k + \alpha(1 - \pi_k)) \cdot u^{(2k+1)} \preceq u^{(2k+2)}$ , so by the definition of  $\pi_{k+1}$

$$\pi_k + \alpha(1 - \pi_k) \leq \pi_{k+1}.$$

It follows that  $1 - \pi_{k+1} \leq (1 - \alpha)(1 - \pi_k)$  for all  $k \geq 1$  and therefore

$$1 - \pi_{k+1} \leq (1 - \pi_1)(1 - \alpha)^k \xrightarrow{k \rightarrow \infty} 0.$$

But  $\pi_k \leq 1$  for all  $k$ , so  $\lim_k \pi_k = 1$ .

Consider some  $\epsilon > 0$ . Then  $\pi_k > 1 - \epsilon$  for all  $k \gg 0$ . Equation (i) implies

$$(1 - \epsilon) \cdot u^{(2k-1)} \preceq \pi_k \cdot u^{(2k-1)} \preceq u^{(2k)} \preceq u^{(2k-1)}.$$

Letting  $k \rightarrow \infty$  this implies  $(1 - \epsilon)u^{(\text{odd})} \preceq u^{(\text{ev})} \preceq u^{(\text{odd})}$ . Since  $\epsilon > 0$  was arbitrary,

$$u^{(\text{ev})} = u^{(\text{odd})}$$

as needed. We will denote  $u^* := u^{(\text{ev})} = u^{(\text{odd})}$ .

Write  $f^n : X \rightarrow X$  for the  $n$ -fold composition of  $f$  with itself:  $f^n = f \circ \dots \circ f$ . Since  $0 \in X$  is minimal,  $u^{(0)} \preceq x$  for any  $x \in X$ . Since  $f$  is order reversing,  $u^{(0)} \preceq f(x) \preceq u^{(1)}$ . It then follows by induction that for all  $k \geq 1$

$$\begin{aligned} f^{2k-1}(X) &\subseteq [u^{(2k-2)}, u^{(2k-1)}]_X \\ f^{2k}(X) &\subseteq [u^{(2k)}, u^{(2k-1)}]_X. \end{aligned}$$

We deduce that

$$\begin{aligned} \bigcap_{n \geq 1} f^n(X) &= \bigcap_{k \geq 1} f^{2k-1}(X) \cap f^{2k}(X) \\ &\subseteq \bigcap_{k \geq 1} [u^{(2k-2)}, u^{(2k-1)}]_X \cap [u^{(2k)}, u^{(2k-1)}]_X \\ &= [u^{(\text{ev})}, u^{(\text{odd})}]_X = \{u^*\}. \end{aligned}$$

Notice that  $u^* \in [u^{(2k)}, u^{(2k-1)}]_X$  (resp.  $u^* \in [u^{(2k-2)}, u^{(2k-1)}]_X$ ) for all  $k \geq 1$ , so since  $f$  is order reversing  $f(u^*) \in [u^{(2k)}, u^{(2k+1)}]_X$  (resp.  $u^* \in [u^{(2k)}, u^{(2k-1)}]_X$ ) for all  $k$ . Therefore  $f(u^*) \in \{u^*\}$ , so  $u^*$  is a fixed point of  $f$ .

If  $y \in X$  is a fixed point of  $f$  then by induction  $y \in f^n(X)$  for all  $n$ , so  $y \in \bigcap_n f^n(X) = \{u^*\}$  and it follows that  $y = u^*$ . Therefore  $u^*$  is the unique fixed point of  $f$ .

Finally, choose some  $x \in X$  and define a sequence by recursion  $x^{(0)} = x$  and  $x^{(n+1)} = f(x^{(n)})$ . Then  $u^{(0)} = 0 \preceq x$  and since  $f$  is order reversing,  $u^{(0)} \preceq x^{(1)} \preceq u^{(1)}$ . Then one proves by induction that  $u^{(2k-2)} \preceq x^{(2k-1)} \preceq u^{(2k-1)}$  and  $u^{(2k)} \preceq x^{(2k)} \preceq u^{(2k-1)}$  for all  $k \geq 1$ . By the sandwich rule  $\lim_n x^{(n)} = u^*$ . □

We can now prove the following.

**Theorem 3.3.** *A function  $\Phi_w : X \rightarrow X$ , based on some argumentation kernel functions, satisfies the conditions of Theorem 3.2. That is,  $\Phi_w$  is order reversing and there exists  $0 < \alpha \leq 1$  such that  $\Phi_w(tx) \preceq \frac{1}{t+\alpha(1-t)}\Phi_w(x)$ , for any  $x \in X$  and any  $0 \leq t \leq 1$ .*

*Proof.* Since  $\varphi_i$  are monotonic, we get that if  $x \preceq y$  then for any  $i$

$$\Phi_w(x)_i = \frac{w_i}{1 + \varphi_i(x)} \geq \frac{w_i}{1 + \varphi_i(y)} = \Phi_w(y).$$

So  $\Phi_w : X \rightarrow X$  is order reversing. For the second condition, observe that the functions  $\psi_i : X \rightarrow \mathbb{R}$  defined by  $\psi_i(x) = \frac{1}{1+\varphi_i(x)}$  are well defined, continuous and take value in the interval  $(0, 1]$ . Since  $X$  is compact, set

$$\alpha = \min_i \min_{x \in X} \frac{1}{1+\varphi_i(x)}.$$

Then  $0 < \alpha \leq 1$ . By construction, for any  $x \in X$  and any  $0 \leq t \leq 1$  and any  $i = 1, \dots, n$

$$\begin{aligned} \Phi_w(tx)_i &= \frac{w_i}{1 + \varphi_i(x)} \\ &= \frac{w_i}{(1-t) + t(1 + \varphi_i(x))} \\ &= \frac{\frac{w_i}{1+\varphi_i(x)}}{t + (1-t)\frac{1}{1+\varphi_i(x)}} \\ &\leq \frac{1}{t + \alpha(1-t)}\Phi_w(x)_i. \end{aligned}$$

Notice that the denominator is always positive. It follows that  $\Phi_w(tx) \preceq \frac{1}{t+\alpha(1-t)}\Phi_w(x)$ . □

This theorem, combined with Theorem 3.2 directly proves Theorem 3.1.

**Definition 3.4.** Let  $\varphi_i$  be some argumentation kernel functions, for  $i = 1, \dots, n$  and  $W = [0, 1]^n$ . We define a function  $h_\varphi : W \rightarrow X$  such that  $h_\varphi((w_1, \dots, w_n))$  is the unique fixed point of  $\Phi_w$ . The image of  $h_\varphi$  will be denoted by  $H_\varphi$  and will also be referred to as the *acceptability degree space* in later sections. We also define a function  $k_\varphi : X \rightarrow \mathbb{R}^n$  such that  $k_\varphi(x)_i = x_i(1 + \varphi_i(x))$ .

We can demonstrate that any semantics which builds on an argumentation kernel is continuous through the following theorem.

**Theorem 3.5.** *The restriction of  $k_\varphi$  to  $H_\varphi$  gives a homeomorphism  $k_\varphi : H_\varphi \rightarrow W$  whose inverse is  $h_\varphi$ . In particular  $h_\varphi : W \rightarrow X$  restricts to a homeomorphism onto  $H_\varphi$ .*

*Proof.* Consider some  $x \in H_\varphi$ . Then  $x = h(w)$  for some  $w = (w_1, \dots, w_n) \in W$ . By definition  $x$  is a fixed point of  $\Phi_w : X \rightarrow X$  so  $x_i = \frac{w_i}{1+\varphi_i(x)}$  for all  $i = 1, \dots, n$ , which is equivalent to  $w_i = x_i(1 + \varphi_i(x))$ . In other words,  $k_\varphi(x) = w$ . We see that  $k_\varphi$  restricts to a function  $k_\varphi : H_\varphi \rightarrow W$  and moreover  $k_\varphi(h_\varphi(w)) = w$  for all  $w \in W$ . Observe thus that  $k_\varphi : H_\varphi \rightarrow W$  is surjective.

Suppose that  $x, x' \in H_\varphi$  are such that  $k_\varphi(x) = k_\varphi(x')$ . That is,  $k_\varphi(x) = k_\varphi(x') = w \in W$ . But  $k_\varphi(x) = w$  means that  $w_i = x_i(1 + \varphi_i(x))$  so  $x$  is a fixed point of  $\Phi_w: X \rightarrow X$ . Similarly,  $k_\varphi(x') = w$  means that  $x'$  is a fixed point of  $\Phi_w$ . But  $\Phi_w$  has a unique fixed point in  $X$ , so  $x = x'$ . It follows that  $k_\varphi: H_\varphi \rightarrow W$  is injective. Thus,  $k_\varphi$  is bijective, and  $k_\varphi \circ h_\varphi$  yielding  $w$  implies that  $k_\varphi^{-1} = h_\varphi$ .

Next, we claim that  $H_\varphi$  is a closed subset of  $X$ , hence it is compact. To see this, let  $x^{(i)} \in H_\varphi$  be a convergent sequence in  $X$  with limit  $y$ . Set  $w^{(i)} = k_\varphi(x^{(i)})$ . Then  $w^{(i)}$  in  $W$  and since  $k_\varphi$  is continuous  $\lim_i w^{(i)} = \lim_i k_\varphi(x^{(i)}) = k_\varphi(y)$ . Since  $W$  is closed in  $\mathbb{R}^n$  we deduce that  $k_\varphi(y) \in W$ . But since  $h_\varphi$  is the inverse of  $k_\varphi: H_\varphi \rightarrow W$  this implies that  $y = h_\varphi(k_\varphi(y)) \in H_\varphi$ . This shows that  $H_\varphi$  is closed, as needed.

Thus,  $H_\varphi$  is a compact subset of  $\mathbb{R}^n$  and  $k_\varphi$  is a bijective continuous function between compact metric spaces. It is therefore a homeomorphism, and consequently so is  $h_\varphi: W_\varphi \rightarrow H$ .  $\square$

The previous theorem is crucial as it shows that for any gradual semantics based on argumentation kernel functions and any arbitrary WAF, one cannot obtain the same scoring function with two different weighting functions on arguments. The continuity is also an important result which will be used during the bisection method to solve the inverse problem, as per [2].

## 4 The Inverse Problem

Recall that our goal is to find a set of initial weights which, when applied to a specific argumentation framework under the chosen semantics will result in a desired preference ordering, derived from the numeric acceptance degree computed for each argument.

The approach described in [2] makes use of two phases to solve the inverse problem. In the first phase, a target acceptance degree is computed for each argument. In the second phase, a numerical method (the bisection method) is used to find the initial weights which lead to this acceptance degree. Since the bisection method is designed to find the zeros of a function with only one variable, and since changing the initial weight of one argument can affect the acceptance degree of other arguments, repeated applications of the bisection method are often necessary. [2] identify several strategies for selecting the argument to which the bisection method should be applied, and demonstrate that selecting the argument whose current acceptance degree is furthest away from its target acceptance degree works well in practice.

In this work we will provide an alternative approach to calculate the initial weights to achieve a desired acceptance degree. We therefore recall how [2] identifies target acceptance degrees for solving the inverse problem. We begin by noting that the desired preference orderings can be represented as a sequence of non-empty sets  $[Ar_0, \dots, Ar_n]$  which partition the set of arguments  $\mathcal{A}$  such that for any  $a, b \in Ar_i$ ,  $0 \leq i \leq n$ ,  $a \simeq b$ , and for any  $a \in Ar_i, b \in Ar_j$  where  $0 \leq i < j \leq n$ ,  $a \triangleright b$ .

Algorithm 1, taken from [2], associates a *minimal upper bound* with every argument in the system. This minimal upper bound — in effect — identifies an achievable final degree for an argument by assuming that all of an argument's attackers have some large value.

If we consider argumentation kernel functions  $\varphi_i(x)$  instead, what Algorithm 1 does is identify a maximum value for  $\varphi_i(x)$  in terms of other arguments in the system, and set the minimal upper bound based on this value. The main part of the algorithm can therefore be more generally rewritten for our larger class of semantics as shown in Algorithm 2.

We now depart from [2]; whereas they used the bisection method to identify initial weights for arguments which achieve the desired acceptability degrees, we consider an analytic approach for doing so.

## 5 Computing Initial Weights from Acceptability Degrees

In this section we identify a vector representation for each semantics which allows us to compute appropriate initial weights. Namely, given an arbitrary WAF  $\mathcal{F} = \langle \mathcal{A}, \mathcal{D}, w \rangle$  such that  $\mathcal{A} = \{a_1, a_2, \dots, a_n\}$ ,  $\vec{w}$  is the column vector  $(w(a_1), w(a_2), \dots, w(a_n))^T$ , and  $\mathbb{A}$  is the adjacency matrix where  $\mathbb{A}_{ij} = 1$  iff  $(a_j, a_i) \in \mathcal{D}$  and 0 otherwise. Note that  $\vec{1}$  is a column vector containing 1s of length equal to  $|\mathcal{A}|$ .

For each semantics  $\sigma_X$ , for  $X \in \{MB, CB, HC\}$ , we denote by  $\vec{X}_\infty$ , the column vector  $(\sigma_X^{\mathcal{F}}(a_1), \sigma_X^{\mathcal{F}}(a_2), \dots, \sigma_X^{\mathcal{F}}(a_n))^T$  containing the acceptability degrees of all arguments. We now consider each semantics in turn.

---

**Algorithm 1** Computing arguments' minimal upper bounds.
 

---

```

function COMPUTEBOUNDS( $\emptyset, \_, \_$ )
    return  $\{\}$ 
end function

function COMPUTEBOUNDS( $[Ar_0, \dots, Ar_n], max, \sigma$ )
    switch  $\sigma$  do
        case  $\sigma_{MB}$ :  $min \leftarrow max / (1 + max + \zeta)$ 
        case  $\sigma_{HC}$ :  $min \leftarrow max / (1 + \max_{a \in Ar_0} |\text{Att}(a)| + \zeta)$ 
        case  $\sigma_{CB}$ :  $min \leftarrow max / (2 + \max_{a \in Ar_0} |\text{Att}(a)| + \zeta)$ 

    return  $\{(Ar_0, min)\} \cup \text{COMPUTEBOUNDS}([Ar_1, \dots, Ar_n], min, \sigma)$ 
end function

function COMPUTEBOUNDS( $[Ar_0, \dots, Ar_n], \sigma$ )
    return COMPUTEBOUNDS( $[Ar_0, \dots, Ar_n], 1, \sigma$ )
end function
    
```

---

**Algorithm 2** More general computation of arguments' minimal upper bound using an argumentation kernel.
 

---

```

function COMPUTEBOUNDS( $[Ar_0, \dots, Ar_n], max, \sigma$ )
    Let  $\varphi_i$  be the argumentation kernel of  $\sigma$ 
    Let  $m$  be the maximal value of  $\varphi_i$  when evaluated for  $i$  over all arguments in  $Ar_0$ .
     $min \leftarrow max / (1 + m + \zeta)$ 
    return  $\{(Ar_0, min)\} \cup \text{COMPUTEBOUNDS}([Ar_1, \dots, Ar_n], min, \sigma)$ 
end function
    
```

---

## 5.1 H-categorizer semantics

As we will do for all other semantics, we examine the equation for the weighted h-categorizer semantics once convergence has taken place. It holds that:

$$\overrightarrow{HC}_\infty = \frac{\overrightarrow{w}}{\mathbb{1} + \mathbb{A}\overrightarrow{HC}_\infty}$$

Here, the division occurs in an element-wise manner. We can rewrite this equation as:

$$\overrightarrow{HC}_\infty + \mathbb{M}\mathbb{A}\overrightarrow{HC}_\infty = \overrightarrow{w}$$

Where  $\mathbb{M}$  is the diagonal matrix such that  $\mathbb{M}_{ii} = \sigma_{HC}^F(a_i)$ , for  $1 \leq i \leq n$ , and 0 otherwise. Since we know the values of  $\overrightarrow{HC}_\infty$  and  $\mathbb{M}$  (the minimum upper bounds), as well as  $\mathbb{A}$  (the adjacency matrix obtained from the structure of our argumentation graph), we can therefore compute  $\overrightarrow{w}$  (the initial weights) directly from the above equation.

## 5.2 Card-Based Semantics

Following the same process as above, at convergence, the weighted card-based semantics can be written as:

$$\overrightarrow{CB}_\infty + \mathbb{D}\overrightarrow{CB}_\infty + \mathbb{D}^{-1}\mathbb{M}\mathbb{A}\overrightarrow{CB}_\infty = \overrightarrow{w}$$

Here  $\mathbb{D}$  is the diagonal matrix such that  $\mathbb{D}_{ii} = |\text{Att}^*(a_i)|$ , for  $1 \leq i \leq n$ , and 0 otherwise. Note that this matrix will contain only 0s in row  $i$  if argument  $a_i$  is unattacked or all its attackers have an initial weight of 0. However from the definition of the semantics, the acceptability degree of such an argument should be equal to its initial weight. In cases where the diagonal element is non-zero, we can simply solve the equation by performing an index-wise calculation. Notice that  $\mathbb{D}^{-1}$  is also a diagonal matrix where the elements are the reciprocal of the diagonal elements of  $\mathbb{D}$ .

	$a_0$	$a_1$	$a_2$	$a_3$
$\sigma_{HC}^{\mathcal{F}}$	0.5	0.167	0.0278	0.167
$\sigma_{MB}^{\mathcal{F}}$	0.333	0.111	0.037	0.111
$\sigma_{CB}^{\mathcal{F}}$	0.333	0.0833	0.012	0.083

 Table 2: Acceptability degrees computed by Algorithm 1 ( $\zeta = 1$ )

	$w(a_0)$	$w(a_1)$	$w(a_2)$	$w(a_3)$
$\sigma_{HC}$	0.5	0.194	0.052	0.167
$\sigma_{MB}$	0.333	0.123	0.0494	0.111
$\sigma_{CB}$	0.333	0.174	0.061	0.083

Table 3: Initial weights computed for each semantics

### 5.3 Weighted Max-Based Semantics

The max operation present in these semantics can be written in a vector format as:

$$\vec{w} = \overrightarrow{MB}_{\infty} + \mathbb{M} \max\{\mathbb{A}\odot\},$$

Where  $\odot$  is a square matrix whose columns are all equal to the vector  $\overrightarrow{MB}_{\infty}$ , and  $\max\{\mathbb{A}\odot\}$  takes the largest element from each row of  $\mathbb{A}\odot$  to form a column vector.

### 5.4 Generalisation

We observe that in general, the function  $k_{\varphi}$ , as per Definition 3.4, can be used to compute the initial weights by passing the desired final degree as input (i.e., given an acceptability degree vector  $h \in H_{\varphi}$ , we can compute  $k_{\varphi}(h)$  to obtain our initial weights analytically).

**Example 5.1.** Consider the argumentation system shown in Figure 1, and assume we wish to obtain the preference ordering  $a_0 \triangleright a_1 \simeq a_3 \triangleright a_2$ . Table 2 show, for each semantics, the acceptability degrees as computed by Algorithm 1 (assuming  $\zeta = 1$ ). Table 3 then shows the initial weights necessary to achieve these acceptability degrees.

### 5.5 Properties

Our results from the previous sections address the Weighting Validity conjecture described in [2], which states that initial weights can be found such that the acceptability degrees of each argument is equal to the corresponding the minimum upper bound generated by Algorithm 1.

We can also easily see that decreasing (resp. increasing) the acceptability degree of some arguments without increasing (resp. decreasing) the degree of any other arguments can only come about by only decreasing (resp. increasing) some initial weights while not increasing (resp. decreasing) any initial weights.

**Proposition 5.2.** *Let us consider two WAFs  $\mathcal{F} = \langle \{a_1, a_2, \dots, a_n\}, \mathcal{D}, w \rangle$  and  $\mathcal{F}' = \langle \{a_1, a_2, \dots, a_n\}, \mathcal{D}, w' \rangle$  and the two corresponding acceptability degree vectors  $\vec{X}_{\infty}, \vec{X}'_{\infty}$ , for  $X \in \{HC, MB, CB\}$  such that for any  $1 \leq i, j \leq n$ ,  $(\vec{X}_{\infty})_i \leq (\vec{X}'_{\infty})_j$  (resp.  $(\vec{X}_{\infty})_i \geq (\vec{X}'_{\infty})_j$ ), it is the case that for any  $1 \leq i, j \leq n$ ,  $\vec{w}_i \leq \vec{w}'_j$  (resp.  $\vec{w}_i \geq \vec{w}'_j$ ).*

It trivially follows that for  $X \in \{HC, MB, CB\}$ , if  $\vec{X}_{\infty}$  is a valid acceptability degree vector, so is any  $\vec{X}'_{\infty}$  such that for any  $1 \leq i \leq n$ ,  $0 \leq (\vec{X}'_{\infty})_i \leq (\vec{X}_{\infty})_i$ . Similarly, increasing the acceptability degree of one argument will increase (or not affect) all initial weights, but we must then ensure that no initial weights exceeds 1. For example, if  $\mathcal{F} = \langle \{a\}, \{(a, a)\}, w \rangle$  with  $w(a) = 1$  then  $\sigma_{HC}^{\mathcal{F}}(a) \approx 0.62$ . In this case, it is not possible to increase the acceptability degree of  $a$  without the initial weight of  $a$  going above 1.

The result of Theorem 3.5 demonstrates that for any given unweighted argumentation graph, the acceptability degree space (i.e., the set of all valid acceptability degree vectors  $H_{\varphi}$ ) for  $\sigma \in \{\sigma_{MB}, \sigma_{CB}, \sigma_{HC}\}$ , will be continuous. Namely, we can always find a sequence of acceptability degree vectors that “links” two acceptability degree vectors. This



builds on the result of [2], which demonstrated continuity, but only in the interval  $[0, \infty)$ . However, this result does not mean that the acceptability degree space will be convex, i.e. if  $\vec{X}_\infty^1$  and  $\vec{X}_\infty^2$  are two valid acceptability degree vectors, then  $\alpha\vec{X}_\infty^1 + (1 - \alpha)\vec{X}_\infty^2$ , for  $\alpha \in [0, 1]$ , will not always be a valid acceptability degree vector. For example, if  $\langle \{a_1, a_2, a_3\}, \{(a_1, a_2), (a_2, a_1), (a_1, a_3), (a_2, a_3)\} \rangle$  is an unweighted argumentation graph, we can find weighting functions such that both  $(0, 0, 1)$  and  $(1, 0, 0.5)$  are valid acceptability degree vectors w.r.t.  $\sigma_{HC}$  but there is no  $\alpha \in ]0, 1[$  such that  $\alpha(0, 0, 1) + (1 - \alpha)(1, 0, 0.5)$  is a valid acceptability degree vector.

In the next proposition, we show that having self-attacking arguments prevent us from achieving some acceptability degree vectors.

**Proposition 5.3.** *Given an unweighted argumentation graph  $\langle \{a_1, a_2, \dots, a_n\}, \mathcal{D} \rangle$ , for all for  $X \in \{HC, MB, CB\}$ ,  $i \in \{1, \dots, n\}$ , there exists  $\vec{X}_\infty$  in the acceptability degree space such that  $[\vec{X}_\infty]_i = 1$  iff there is no  $a \in \{a_1, \dots, a_n\}$  such that  $(a, a) \in \mathcal{D}$ .*

*Proof.* This proof is split into two parts. First, assume that there is a self-attacking argument  $a_i \in \mathcal{A}$ , it is easy to show that the maximum acceptability degree of  $a_i$  is  $(-1 + \sqrt{5})/2 \simeq 0.618$ , for  $\sigma_{HC}$  and  $\sigma_{MB}$ , and  $-1 + \sqrt{2} \simeq 0.41$  for  $\sigma_{CB}$ , thus it is not possible to find an acceptability degree vector  $\vec{X}_\infty$  such that  $[\vec{X}_\infty]_i = 1$ , for  $X \in \{HC, MB, CB\}$ . Now, assume that there are no self-attacking arguments in  $\mathcal{A}$ , then for all  $X \in \{HC, MB, CB\}$ ,  $i \in \{1, \dots, n\}$ , we can create an acceptability degree vector  $\vec{X}_\infty$  such that  $[\vec{X}_\infty]_i = 1$  by putting the weight of  $a_i$  to 1 and the initial weights of all other arguments to 0.  $\square$

**Definition 5.4.** Given two unweighted argumentation graphs  $\langle \mathcal{A}, \mathcal{D} \rangle$  and  $\langle \mathcal{A}', \mathcal{D}' \rangle$ , we say that  $f$  is an isomorphism from  $\mathcal{A}$  to  $\mathcal{A}'$  if  $(a, b) \in \mathcal{D}$  iff  $(f(a), f(b)) \in \mathcal{D}'$ .

The next proposition shows that isomorphic arguments will induce a symmetry effect in the acceptability degree space.

**Proposition 5.5.** *Given an unweighted argumentation graph  $\langle \mathcal{A}, \mathcal{D} \rangle$ ,  $\mathcal{A} = \{a_1, \dots, a_n\}$  and  $a_i, a_j \in \mathcal{A}$  such that there is an isomorphism  $f$  from  $\mathcal{A}$  to  $\mathcal{A}$  and  $f(a_i) = a_j$ , it holds that for all  $X \in \{HC, MB, CB\}$ ,  $\vec{X}_\infty$  is in the acceptability degree space iff  $([\vec{X}_\infty]_1, \dots, [\vec{X}_\infty]_{i-1}, [\vec{X}_\infty]_j, [\vec{X}_\infty]_{i+1}, \dots, [\vec{X}_\infty]_{j-1}, [\vec{X}_\infty]_i, [\vec{X}_\infty]_{j+1}, \dots, [\vec{X}_\infty]_n)^T$  is in the acceptability degree space.*

The previous proposition follows directly from the anonymity principle which is satisfied by all three semantics studied in the paper [4]. One important observation is that, apart from the structure of the graph, the semantics chosen also plays an important role in the shape of the acceptability degree space. For instance, if we consider a complete graph of size  $n$  with self-attacking arguments, the acceptability degree space for  $\sigma_{MB}$  is  $\{(v_1, v_2, \dots, v_n)^T \mid \forall i \in \{1, \dots, n\}, 0 \leq v_i \leq (-1 + \sqrt{5})/2\}$  while for  $\sigma_{CB}$  it is  $\{(v_1, v_2, \dots, v_n)^T \mid \forall i \in \{1, \dots, n\}, 0 \leq v_i \leq (-1 + \sqrt{5})/2 \text{ and } v_i(1 + \sum_{j=1}^n v_j) \in [0, 1]\}$ . We provide a graphical representation of the acceptability degree space for  $\sigma_{HC}$  on an example in Figure 2. It is clear here that the acceptability degree space for  $\sigma_{CB}$  is a subset of the acceptability degree space for  $\sigma_{MB}$  for all  $n > 0$ . In the next proposition, we show that this inclusion holds in the general case.

**Proposition 5.6.** *Given an arbitrary unweighted argumentation graph  $\langle \mathcal{A}, \mathcal{D} \rangle$ ,  $\mathcal{A} = \{a_1, \dots, a_n\}$ , the acceptability degree space for  $\sigma_{HC}$  is a subset of the acceptability degree space for  $\sigma_{MB}$ .*

*Proof.* Let us consider an arbitrary acceptability degree vector  $\vec{HC}_\infty = (v_1, \dots, v_n)$  such that  $v_i \in [0, 1]$  for all  $1 \leq i \leq n$ . We show that this vector is in the acceptability degree space for  $\sigma_{MB}$ . By definition, we know that for all  $i \in \{1, \dots, n\}$ ,  $0 \leq v_i(1 + \sum_{a_j \in \text{Att}(a_i)} v_j) \leq 1$ . It follows that for all  $i \in \{1, \dots, n\}$ ,  $0 \leq v_i(1 + \max_{a_j \in \text{Att}(a_i)} v_j) \leq 1$  and thus,  $\vec{HC}_\infty$  is in the acceptability degree space for  $\sigma_{MB}$ .  $\square$

The final property we consider is monotonicity, which — together with continuity — is required for the bisection method of [2] to operate. Due to space constraints, we only prove this result for the weighted h-categoriser semantics. This means that we fix our argumentation kernel function of Section 3 to be  $\varphi_i(x) = \|(\mathbb{A}.x)_i\|_1 = \sum a_{i,j}x_j$ .

We will equip  $\mathbb{R}^n$  with the  $L^\infty$  norm. Recall that there is an induced operator norm [9] on the set of all linear transformations of  $\mathbb{R}^n$  which we denote as  $\text{End}(\mathbb{R}^n) \cong \text{Mat}_{n \times n}(\mathbb{R})$  where for any matrix  $A$  we set this operator norm

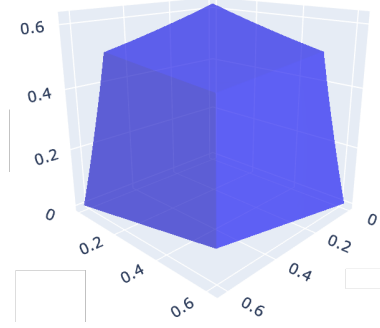


Figure 2: Representation of the acceptability degree space for  $\sigma_{HC}$  on a complete argumentation graph with 3 arguments (in blue). Each axis represents the acceptability degree of one argument.

as

$$\begin{aligned} \|A\| &= \sup\left\{\frac{\|Ax\|_\infty}{\|x\|_\infty} : 0 \neq x \in \mathbb{R}^n\right\} \\ &= \max\{\|Ax\|_\infty : x \in \mathbb{R}^n, \|x\|_\infty = 1\} \\ &= \max\left\{\frac{\|Ax\|_\infty}{\|x\|_\infty} : 0 \neq x \in \mathbb{R}^n\right\}. \end{aligned}$$

It is well known (and easy to check) that this defines a norm on  $\text{End}(\mathbb{R}^n)$ . Since all norms on  $\mathbb{R}^k$  are equivalent, the operator norm makes  $\text{End}(\mathbb{R}^n)$  a complete normed space. By definition for any  $x \in \mathbb{R}^n$

$$\|Ax\|_\infty \leq \|A\| \cdot \|x\|_\infty$$

Moreover, the operator norm is multiplicative, namely

$$\|AB\| \leq \|A\| \cdot \|B\|$$

This is a standard construction, see for example [9, Section 5.2 and Theorem 5.3 and Remark 5.4(1)].

**Lemma 5.7.** Consider a matrix  $A \in \text{Mat}_{n \times n}(\mathbb{R})$  such that for every  $i = 1, \dots, n$

- (a)  $a_{i,i} = 0$
- (b)  $\sum_j |a_{i,j}| < 1$ .

Then  $\det(I + A) > 0$ .

*Proof.* First we show that any matrix  $A$  with these properties is invertible. For any  $x \in \mathbb{R}^n$  with  $\|x\|_\infty = 1$  we get

$$\begin{aligned} \|Ax\|_\infty &= \max_i |Ax_i| = \max_i \left| \sum_j a_{i,j} x_j \right| \\ &\leq \max_i \sum_j |a_{i,j}| \cdot |x_j| \leq \max_i \sum_j |a_{i,j}| \end{aligned}$$

It follows that  $\|A\| \leq \max_i \sum_j |a_{i,j}| < 1$ . Since  $\text{End}(\mathbb{R}^n)$  with the operators norm is complete,  $B = \sum_{k=0}^{\infty} (-1)^k A^k$  converges and  $B(I + A) = I$ . It follows that  $I + A$  is invertible.

Let  $U \subseteq \text{Mat}_{n \times n}(\mathbb{R})$  be the (open) subset of all matrices  $A$  satisfying the conditions in the lemma. Clearly  $0 \in U$ . Moreover  $U$  is path connected since for any  $A \in U$  the function  $\lambda: [0, 1] \xrightarrow{t \rightarrow tA} \text{Mat}_{n \times n}(\mathbb{R})$  gives a path in  $U$  from 0 to  $A$ . The function  $f: U \rightarrow \mathbb{R}$  defined by  $f(A) = \det(I + A)$  is clearly continuous. Since  $I + A$  is invertible,  $f(A) \neq 0$  for all  $A \in U$ . Since  $f(0) = 1 > 0$ , the intermediate value theorem implies that  $f(A) > 0$  for all  $A \in U$ .  $\square$

**Corollary 5.8.** Let  $M \in \text{Mat}_{n \times n}(\mathbb{R})$  be a matrix which satisfies the conditions

- 1.  $m_{i,j} \geq 0$  for all  $i, j$

2.  $m_{i,i} > \sum_{j \neq i} m_{i,j}$  for every  $i = 1, \dots, n$ .

Then  $\det(M) > 0$ .

*Proof.* The conditions imply  $m_{i,i} > 0$  for all  $i = 1, \dots, n$ . By inspection

$$\text{diag}(m_{1,1}^{-1}, \dots, m_{n,n}^{-1}) \cdot M = I + A$$

where  $A$  is of the form in Lemma 5.7. Then  $\det(M) = \det(I + A) \cdot \prod_i m_{i,i} > 0$ .  $\square$

$$M = \begin{bmatrix} 1 + a_{1,1}x_1 + \sum_j a_{1,j}x_j & a_{1,2}x_1 & a_{1,3}x_1 & \cdots & a_{1,n}x_1 \\ a_{2,1}x_2 & 1 + a_{2,2}x_2 + \sum_j a_{2,j}x_j & a_{2,3}x_2 & \cdots & a_{2,n}x_2 \\ \vdots & & \ddots & \ddots & \\ a_{n,1}x_n & a_{n,2}x_n & a_{n,3}x_n & \cdots & 1 + a_{n,n}x_n + \sum_j a_{n,j}x_j \end{bmatrix}$$

Figure 3: The matrix of the  $M$  with entries  $\frac{\partial h_i}{\partial x_\ell}$

**Corollary 5.9.** Let  $A$  be an  $n \times n$  matrix with  $0 \leq a_{i,j} \leq 1$ . Let  $x \in \mathbb{R}^n$  be a column vector with  $0 \leq x_i \leq 1$ . Let  $\delta_1, \dots, \delta_n > 0$ . Let  $M$  be the matrix given by

$$m_{i,j} = \begin{cases} \delta_i + \sum_k a_{i,k}x_k & \text{if } i = j \\ a_{i,j}x_j & \text{if } i \neq j \end{cases}$$

Then  $\det(M) > 0$  and moreover the diagonal entries of  $M^{-1}$  are positive, i.e.  $(M^{-1})_{i,i} > 0$  for all  $i$ .

*Proof.* First, we show that all matrices  $M$  of this form have  $\det(M) > 0$ . Let  $M'$  be the matrix with  $m'_{i,i} = m_{i,i}$  and  $m'_{i,j} = a_{i,j}x_j$ . Observe that  $m'_{i,j} \geq 0$  and that for any  $i$

$$m'_{i,i} = \delta_i + \sum_j a_{i,j}x_j > \sum_{j \neq i} a_{i,j}x_j = \sum_{j \neq i} m'_{i,j}.$$

By Corollary 5.8  $\det(M') > 0$ .

Let  $S_n$  denote a permutation of  $n$  symbols. For any  $\sigma \in S_n$ , let  $\text{fix}(\sigma)$  denote those symbols which the permutation does not shift (i.e., the fixed points of  $\sigma$ ) and  $\text{supp}(\sigma)$  those symbols which are permuted (i.e., its support). Clearly,  $\sigma$  induces a permutation of  $\text{supp}(\sigma)$ . Therefore

$$\begin{aligned} \prod_{i \in \text{supp}(\sigma)} m_{i,\sigma(i)} &= \prod_{i \in \text{supp}(\sigma)} a_{i,\sigma(i)}x_i \\ &= \prod_{i \in \text{supp}(\sigma)} a_{i,\sigma(i)} \cdot \prod_{i \in \text{supp}(\sigma)} x_i \\ &= \prod_{i \in \text{supp}(\sigma)} a_{i,\sigma(i)} \cdot \prod_{i \in \text{supp}(\sigma)} x_{\sigma(i)} \\ &= \prod_{i \in \text{supp}(\sigma)} a_{i,\sigma(i)}x_{\sigma(i)} = \prod_{i \in \text{supp}(\sigma)} m'_{i,\sigma(i)}. \end{aligned}$$

We can now compute

$$\begin{aligned} \det(M) &= \sum_{\sigma \in S_n} (-1)^\sigma \prod_i m_{i,\sigma(i)} \\ &= \sum_{\sigma \in S_n} (-1)^\sigma \prod_{i \in \text{fix}(\sigma)} m_{i,i} \cdot \prod_{i \in \text{supp}(\sigma)} m_{i,\sigma(i)} \\ &= \sum_{\sigma \in S_n} (-1)^\sigma \prod_{i \in \text{fix}(\sigma)} m'_{i,i} \cdot \prod_{i \in \text{supp}(\sigma)} m'_{i,\sigma(i)} \\ &= \sum_{\sigma \in S_n} (-1)^\sigma \prod_i m'_{i,\sigma(i)} \\ &= \det(M') > 0. \end{aligned}$$

We have shown that all matrices  $M$  of the form in the statement (for any  $n$ ) have positive determinant and in particular they are invertible. Let  $M[i, j]$  denote the  $(i, j)$ -minor of  $M$ , namely the  $(n - 1) \times (n - 1)$  matrix formed by deleting the  $i$ th row and  $j$ th column of  $M$ . A consequence of Cramer's rule

$$(M^{-1})_{i,i} = \frac{\det(M[i, i])}{\det(M)}.$$

Notice that  $M[i, i]$  is a matrix of the form in the statement (with  $A$  replaced with  $A' = A[i, i]$  and  $x$  replaced with  $x' = (x_1, \dots, \hat{x}_i, \dots, x_n)$  and  $\delta_i$  replaced with some  $\delta'_i > \delta_i$ ). So  $\det(M[i, i]) > 0$  and as a consequence  $(M^{-1})_{i,i} > 0$ .  $\square$

Let  $\mathbb{A}$  denote the adjacency matrix of a directed WAF  $\mathcal{F}$ . We can now apply these results to the argumentation kernel functions for  $\sigma_{HC}$ .

**Theorem 5.10.** *Let  $h_\varphi: W \rightarrow H_\varphi$  and  $k_\varphi: H_\varphi \rightarrow W$  be associated with the argumentation kernel functions  $\varphi_i(x) = \|(\mathbb{A}x)_i\|_1 = \sum_j a_{i,j}x_j$ . Then  $h_\varphi$  and  $k_\varphi$  are  $C^\infty$  (differentiable infinitely many often) and moreover, for every  $w \in W$  and every  $i$*

$$\frac{\partial h_i}{\partial w_i}(w) > 0.$$

*In particular  $h_\varphi$  is increasing in each fibre, i.e if  $w, w' \in W$  are such that  $w' - w = (0, \dots, 0, \epsilon_i, 0, \dots, 0)$  for some  $\epsilon_i > 0$  then  $h_\varphi(w) < h_\varphi(w')$ .*

*Proof.* The function  $k_\varphi$  is the restriction to  $H_\varphi$  of the  $C^\infty$  function  $k_\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^n$

$$k_\varphi(x)_i = x_i(1 + \sum_j a_{i,j}x_j).$$

For any  $\ell = 1, \dots, n$ , we get

$$\frac{\partial k_i}{\partial x_\ell} = \delta_{i,\ell}(1 + \sum_j a_{i,j}x_j) + x_i a_{i,\ell}.$$

The derivative of  $k_\varphi$  at  $x \in \mathbb{R}^n$  is the matrix  $M$  with entries  $\frac{\partial h_i}{\partial x_\ell}$  so

$$M_{i,\ell} = \begin{cases} 1 + \sum_j a_{i,j}x_j + a_{i,\ell}x_i & \text{if } i = \ell \\ a_{i,\ell}x_i & \text{if } i \neq \ell. \end{cases}$$

In matrix form this can be written as  $M = I + \text{diag}(\mathbb{A}x) + \text{diag}(x)\mathbb{A}$ , which — by inspection — takes the form of the matrix shown in Figure 3.

Thus,  $M$  has the form in Corollary 5.9 and we deduce that  $\det(M) > 0$  and the diagonal of  $M^{-1}$  has positive entries. In particular  $M$  is invertible. By the inverse function theorem  $k$  is invertible in a neighbourhood of  $x$  with an inverse  $k_\varphi^{-1}$  defined in a neighbourhood  $U$  of  $y = k_\varphi(x)$  and the derivative of  $k_\varphi^{-1}$  at  $y$  is equal to  $M^{-1}$ .

Consider some  $w \in W$  and set  $x = h_\varphi(w)$ . We have seen that the matrix  $M$  of the derivative of  $k_\varphi$  at  $x$  is invertible and  $M^{-1}$  has positive diagonal entries. Also  $k_\varphi$  is invertible at a neighbourhood of  $x$  with inverse  $k_\varphi^{-1}$  defined in a neighbourhood  $U$  of  $k_\varphi(x) = w$  and with derivative  $M^{-1}$  at  $w$ . But  $h_\varphi$  is the inverse of  $k_\varphi|_{H_\varphi}$  so  $h_\varphi$  must coincide with  $k_\varphi^{-1}$  on  $U \cap W$  and in particular  $h_\varphi$  is  $C^\infty$  at  $w$  and its derivative is given by  $M^{-1}$  whose diagonal entries are positive.  $\square$

## 6 Conclusions and Future Work

Like [2], this paper primarily examines the inverse problem in gradual argumentation semantics. It proves several conjectures made in that work, namely demonstrating that the  $HC$ ,  $MB$  and  $CB$  semantics are continuous, and that the  $HC$  semantics are strongly monotonic. In addition, we prove that a more general class of gradual semantics is continuous, and has a unique fixed point, than done in work such as [3, 4]. Using these properties we demonstrate that — for any semantics which can be represented using an argumentation kernel — we can compute the initial weights analytically, obviating the need for the numerical method described in [2].

We are pursuing several avenues of future work. First, we want to investigate whether our monotonicity results can be extended to more general argumentation kernels, as well as investigating which existing weighted and ranking semantics (e.g., [10, 11, 12]) can be represented using such argumentation kernels, and whether there is an overlap between the

argumentation kernel approach and the properties described in [4, 8]. Second, we would like to further investigate the utility of the class of semantics described by argumentation kernels, and identify whether useful semantics exist which cannot be described by such kernels but which still respect monotonicity and continuity. If such semantics are found, then we wish to determine whether an analytical solution to the inverse problem for such semantics can be found, or whether a numerical method for solving the inverse problem in such cases is applicable. In addition, characterising the shape of the applicability degree in terms of a semantics and initial weights may yield interesting insights. We want to examine how our work can be applied to dynamic argumentation and sensitivity analysis of arguments. In effect, our work is an initial step to asking by how much an argument must change in order to change some conclusion. However, it does not (as yet) provide a lower bound to this change, but rather only a sufficient bound. Finally, we observe that the use of argumentation kernel functions have the potential to allow different arguments to compute a final acceptability degree in different ways. This could allow for the modelling of — for example — different acceptability degrees arising from different types of arguments, c.f., argumentation schemes, and serves as an exciting avenue of future research.

## References

- [1] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. An introduction to argumentation semantics. *Knowledge Eng. Review*, 26(4):365–410, 2011.
- [2] Nir Oren, Bruno Yun, Srdjan Vesic, and Murilo S. Baptista. The inverse problem for argumentation gradual semantics. *CoRR*, abs/2202.00294, 2022.
- [3] Fuan Pu, Jian Luo, Yulai Zhang, and Guiming Luo. Argument ranking with categoriser function. In Robert Buchmann, Claudiu Vasile Kifor, and Jian Yu, editors, *Knowledge Science, Engineering and Management - 7th International Conference, KSEM 2014, Sibiu, Romania, October 16-18, 2014. Proceedings*, volume 8793 of *Lecture Notes in Computer Science*, pages 290–301. Springer, 2014.
- [4] Leila Amgoud, Dragan Doder, and Srdjan Vesic. Evaluation of argument strength in attack graphs: Foundations and semantics. *Artificial Intelligence*, 302:103607, 2022.
- [5] Bruno Yun, Srdjan Vesic, and Madalina Croitoru. Ranking-based semantics for sets of attacking arguments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3033–3040, June 2020.
- [6] Leila Amgoud, Jonathan Ben-Naim, Dragan Doder, and Srdjan Vesic. Acceptability Semantics for Weighted Argumentation Frameworks. In *IJCAI*, pages 56–62, 2017.
- [7] Sylvie Coste-Marquis, Sébastien Konieczny, Pierre Marquis, and Mohand Akli Ouali. Selecting Extensions in Weighted Argumentation Frameworks. In *Computational Models of Argument - Proceedings of COMMA 2012, Vienna, Austria, September 10-12, 2012*, pages 342–349, 2012.
- [8] Elise Bonzon, Jérôme Delobelle, Sébastien Konieczny, and Nicolas Maudet. A Comparative Study of Ranking-Based Semantics for Abstract Argumentation. In *Proc. AAAI*, pages 914–920, 2016.
- [9] H. W. Alt. *Linear Functional Analysis*. Springer, 2016.
- [10] Paul E. Dunne, Anthony Hunter, Peter McBurney, Simon Parsons, and Michael Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artif. Intell.*, 175(2):457–486, 2011.
- [11] Antonio Rago, Francesca Toni, Marco Aurisicchio, and Pietro Baroni. Discontinuity-Free Decision Support with Quantitative Argumentation Debates. In *KR 2016*, pages 63–73, 2016.
- [12] Céilia da Costa Pereira, Andrea Tettamanzi, and Serena Villata. Changing One’s Mind: Erase or Rewind? In *Proc., IJCAI 2011*, pages 164–171, 2011.