Report on ECMA MS 21631

Hodge allocation for cooperative rewards: a generalization of Shapley's cooperative value allocation theory via Hodge theory on graphs by Tongseok Lim

1. Summary

The Shapley value has been subject of intense study by economists ever since its introduction by Lloyd Shapley [21]. For instance, various characterizations were developed, in particular, [14] emphasizes the characterization based on the balanced contributions property by [24], and the characterization based on marginality by [15]. Moreover, many applications and extensions have been studied, as well as mutually illuminating connections between noncooperative and cooperative game theory ("Nash program"), e.g. by [8], [20], and [13]. More recently, the Shapley value became of high interest in the context of interpretations of black box machine learning models [12] as well as its application for the evaluation of advertiser actions [1]. These applications are novel in the sense that the Shapley value is not only computed exemplarily by various game theorists, but applied frequently by numerous users.[1] The Shapley value is therefore a very important concept, not only for game theorists but also for econometricians.

A player's Shapley value is computed as this player's average marginal contribution. [5] emphasizes that this "comes as a surprise at first glance: uniqueness is the consequence of four basic axioms, and nothing in those axioms hints at the marginality principle, of long tradition in economic theory. In the clarification of this puzzle, Young (1985) provided a key piece." Recently, [23] added further clarification to this relationship through the study of the combinatorial Hodge decomposition of individual marginal contributions. The present paper builds upon this work, finds new structural insights as well as interpretations and plausible generalizations of the Shapley value.

Let me start with the result in the present paper that gives a new and interesting interpretation of the Shapley value as the expected contribution of a player in a random coalition formation process. We consider $|N|$ players and the oriented $|N|$-dimensional hypercube graph $G=(\mathcal{V},\mathcal{E})$, where the set of vertices represents the set of coalitions and the directed edges represent the addition of a player, i.e.,

$$\mathcal{V}=2^N, \qquad \mathcal{E}=\{(S,S\cup\{i\}\in\mathcal{V}\times\mathcal{V}|\, S\subseteq N\setminus\{i\}, i\in N)\}$$

Note that a coalitional game is a function $v\colon\mathcal{V}\to\mathbb{R}$ such that $v(\varnothing)=0$. Moreover, it is natural to consider the gradient $\mathrm{d}v\colon\mathcal{E}\to\mathbb{R}$, which assigns to every edge the marginal con-

1. For its actual usage, we refer to https://github.com/slundberg/shap and https://developers.google.com/ads-data-hub/guides/shapley.

tribution $\mathrm{d}v(S, S\cup\{i\}) = v(S\cup\{i\}) - v(S)$. This further motivates the partial differential operator d_i, which filters the individual marginal contributions from the gradient,

$$\mathrm{d}_i v(S, S\cup\{i\}) = \begin{cases} \mathrm{d}v(S, S\cup\{i\}) & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Note that $\sum_{i\in N} \mathrm{d}_i = \mathrm{d}$. Now let us consider a random walk on the hypercube graph G, i.e., we start from the empty coalition and we go from coalition to coalition, where at each step a player is equally likely chosen to join (if the player is not yet member of the coalition) or to leave (if the player is already a member of the coalition). This is described by the Markov chain $(X_n)_{n\in\mathbb{N}_0}$ with state space $\mathcal{V}$, initial state $X_n = \varnothing$ and transition probabilities

$$p_{S,T} = \begin{cases} 1/N & \text{if } (S,T)\in\mathcal{E} \text{ or } (T,S)\in\mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that this is not a strategic coalition formation process in the sense that the players join randomly (the author develops a generalization that accommodates for unequal probabilities).

Denote the probability space by $(\Omega, \mathcal{F}, \mathcal{P})$. For a sample path $\omega\in\Omega$, let $\tau_N(\omega)\in\mathbb{N}$ denote the first time the process visits $N\in\mathcal{V}$. The total contribution of player i along the path ω is

$$\sum_{n=1}^{\tau_N(\omega)} (v(S\cup\{i\}) - v(S)),$$

and taking the expectation over all paths defines the value

$$V_i(N) := \int_\Omega \sum_{n=1}^{\tau(\omega)} \mathrm{d}_i v(S, S\cup\{i\}) d\mathcal{P}(\omega).$$

The author now provides the following insight: A player i's Shapley value $\phi_i(v)$ equals this value, i.e.,

$$\phi_i(v) = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(|N|! - |S|! - 1)}{|N|!} \mathrm{d}_i v(S, S\cup\{i\}) = V_i(N).$$

The author further generalizes this value V_i (and hence the Shapley value) in various directions. First, let us consider the contributions a player makes not on paths all the way up to N, but toward some coalition $T\subseteq N$. Let $\tau_T(\omega)\in\mathbb{N}$ denote the first time the process starting in $\varnothing$ visits $T\in\mathcal{V}$. Then, the value given by

$$V_i(T) := \int_\Omega \sum_{n=1}^{\tau_T(\omega)} \mathrm{d}_i v(S, S\cup\{i\}) d\mathcal{P}(\omega) \tag{1}$$

represents a player's total contribution toward T. The author now establishes a relationship between $V_i(\cdot)$ and component game that lie at the heart of the recent work by [23]. There, a coalitional game v is decomposed into component games $v_i, i \in N$, such that $v = \sum_{i \in N} v_i$ and $\mathrm{d}v_i \in \mathcal{R}(\mathrm{d})$, where $\mathrm{d}v_i$ is the gradient of v_i and $\mathcal{R}(\mathrm{d})$ denotes the range of the transpose of the oriented incidence matrix of G.[2] In fact, the author establishes the surprising $V_i = v_i$.

The first main Theorem is now a characterization of these functions, based on the four Shapley axioms and a new axiom "reflection", which reflects the idea that the $v_i(T)$ represents a stochastic path integral as in (1).

Next, the author generalizes the Shapley value by considering more general notions of marginal contributions. In fact, the author considers arbitrary edge flows $f_i: \mathcal{E} \to \mathbb{R}$ in place of $\mathrm{d}_i v(S, S \cup \{i\})$ in (1) and argues that this generalizes the Shapley value to the f-Shapley value, that the above value function $V_i(\cdot)$ motivated by the random walk interpretation generalizes to $V_f(\cdot)$, and he demonstrates that the equality $V_i = v_i$ generalizes to $V_f = v_f$, where v_f originates from the decomposition of f_i instead of d_i.

Thereafter, the author argues that this generalized version of the Shapley value can be utilized in the context of the value introduced by [11], with the generalized marginal contribution in place of the usual marginal contribution.

Motivated by [23], the author generalizes the Shapley value further via the construction in (1) by allowing for weights on the edges (reflecting the idea that an edge, i.e. a player joining some coalition, can be more or less likely), allowing for arbitrary connected subgraph of the hypercube (reflecting the idea that not every coalition is feasible), and allowing for a starting point that is not necessarily the empty coalition. The author calls this the Hodge allocation and shows how it is related to a combinatorial Hodge decomposition.

Finally, the author studies a generalization to multigraphs that allow for multiple edges between nodes. Beyond revealing structural relationships, the author emphasizes how the results help to effectively calculate various concepts.

2. [23] study coalitional games $v: \mathcal{V} \to \mathbb{R}$ and functions on the edges $f: \mathcal{E} \to \mathbb{R}$, in their relationship to the oriented incidence matrix of the graph $(\mathcal{V}, \mathcal{E})$. Specifically, denote the transpose of the oriented incidence matrix of the graph $(\mathcal{V}, \mathcal{E})$ by B^T. If we adhere to the standard bases of the vector spaces $\{v: \mathcal{V} \to \mathbb{R}\}$ and $\{f: \mathcal{E} \to \mathbb{R}\}$, then we can think of this matrix as a linear operator d that converts a coalitional game v into a function on the edges $\mathrm{d}v$. That is, we can think of this matrix applied to v as the gradient of v, which assigns to a game v its marginal contributions, $\mathrm{d}v(S, S \cup \{j\}) = v(S \cup \{j\}) - v(S)$. Moreover, we introduce the partial differential operator $\mathrm{d}_i: \mathcal{E} \to \mathbb{R}$,

$$\mathrm{d}_i v(S, S \cup \{i\}) = \begin{cases} \mathrm{d}v(S, S \cup \{i\}) & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Note that $\sum_{i \in N} \mathrm{d}_i = \mathrm{d}$.

On the other hand, we can consider the adjoint of d, denoted by d^* and given by the oriented incidence matrix B. This is the operator which applied to some $f: \mathcal{E} \to \mathbb{R}$ gives us the coalitional game $\mathrm{d}^* f$, which is defined by $(\mathrm{d}^* f)(S) = \sum_{T: (S,T) \in \mathcal{E} \text{ or } (T,S) \in \mathcal{E}} f(T, S)$.

By the fundamental theorem of linear algebra, we can decompose any function $f: \mathcal{E} \to \mathbb{R}$ into a component that lies in the range of d and a component that lies in the nullspace of d^*, i.e.,

$$f = \mathrm{d}u + p,$$

with $\mathrm{d}u \in \mathcal{R}(\mathrm{d})$ and $p \in \mathcal{N}(\mathrm{d}^*)$. The question of interest in [23] now is the following: Given a coalitional game v, consider for each player i the case $f = \mathrm{d}_i$ and find a coalitional game v_i such that $\mathrm{d}_i v = \mathrm{d}v_i + p$, where $\mathrm{d}v_i \in \mathcal{R}(\mathrm{d})$ and $p \in \mathcal{N}(\mathrm{d}^*)$. Remarkably, this recovers the Shapley value in a novel way, namely as $\phi_i(v) = v_i(N)$.

2. Contribution/Novelty

The present paper demonstrates that the connection made by [23] was the bridge to a very fruitful discovery. Indeed, the present paper gives a new interpretation of the Shapley value (being the expected contribution of a player on a random walk of coalition formation) and shows how this can be naturally extended in various directions. Moreover, the author provides a characterization of the decomposing games introduced by [23]. To the best of my knowledge, these are new and mesmerizing contributions. I am tempted to say that the present paper stands out amongst the many publications on the Shapley value in recent years and paves the way for future theoretical and applied research.

This is supported by inspiring examples and ideas that motivate further studies of the topic. This includes ideas that help understanding coalition formation processes and that go beyond research invoking the efficiency axiom.

A critique seems to be appropriate with regards to section 6, which deals with the generalization of the value introduced by [23]. Even though the proposed generalization is an interesting path for further research, the current contribution is perhaps not as illuminating and the economic intuition for this generalization remains rather vague.

3. Essential Changes

In its present form, the paper might not target the broad readership of the journal it was submitted to. It seems a little narrowed toward the many mathematicians among the readers. Even though this might not diminish the importance of the paper, it reduces the value it has for its targeted audience. Therefore, I suggest a few modifications regarding the interpretation of the findings and the connections to the literature. Some ideas regarding the simplification of the exposition (make no mistake: the paper is written very clearly; but it requires quite some math early on) are offered in the next section.

3.1. Interpretation

My understanding is that economic theory depends both on mathematical models and on interpretations. In the present paper, the mathematical insight is very compelling. Now I would encourage the author to work a little more on the interpretation and the "story" and present the math rather as an equally important ingredient that facilitates the economic interpretation. In fact, I am not sure if the characterization and the connection to Hodge decomposition is as interesting to the general readership of a journal for economists as is suggested by its prominence in the paper. Instead, I suggest to focus the opening on the interpretation of the Shapley value as the outcome of a coalition formation process.

Indeed, the idea of the random walk is very intuitive and the equality $V_i(N) = \phi_i(v)$ is a very nice insight to start with (an elementary proof would be appreciated and would make it easier to digest the more general theorems). Beginning with this result would offer the reader an easier entry into the paper. Emphasizing this finding and motivating further insights from there would ease the flow. Beyond the more general definitions, it would be appreciated if the author motivates further results. It appears that the current main purpose is about the mathematical structure and the computation of the generalizations of the Shapley value. I suggest to better motivate and explain how these generalizations make it possible for us to study question about coalition formation processes (that were perhaps raised by others before).

Specifically, the f-Shapley value deserves some better explanation. Indeed, it seems that investigating $f = \mathrm{d}_i$ was the brilliant idea of [23] to recover the Shapley value. Why should we be interested deviating from their setup? My understanding is that you could communicate the following reasons: (1) Getting beyond the efficiency assumption as discussed later in Remark 7.1 seems very compelling. To me it seems that there is ample interest in concepts that do without efficiency, which can be useful to motivate the f-Shapley value. Can you exemplarily clarify how f-Shapley values can be used to investigate relationship between efficiency and stability? (2) Functions that represent expected benefit from joining coalitions and this way encapsulate strategic thinking could be employed. (3) Redistribution as exemplified by the α-Shapley value. And of course the combination of all three. In order to motivate each of these reasons, a connection with existing literature is essential.

The other generalizations come natural. What happens if we eliminate implausible edges (e.g., edges that promises a player to loose value) or make them less likely... Can we still calculate the value for such seemingly intractable value function...

3.2. Connection to the literature

I think that the following literature also could be related to.

1. [9] introduce the notation of a discrete potential function and highlight the connection with the Shapley value. Even though their setup is different, it would be helpful to contrast it with your setup.

2. Note the relationship of [9] and [15], where the latter encapsulates a very compelling fairness property known as balanced contributions. Does Φ inherit a similar property from the Shapley value and can it be used to characterize Φ? For a survey of properties that are equivalent to balanced contributions, it is referred to [3].

3. There is a large literature on coalition formation, see for instance [19] and [18]. Within the context of coalitional games, [22] studies population monotonic allocation schemes, which investigate what payoffs support the formation of the growing coalition. Those works usually acknowledge strategic behavior of the players, whereas the present paper assumes players to join a coalition equally likely, or with a priori defined weights. I encourage the author to demonstrate how the new findings apply if weights are derived from the game in a plausible way, e.g., when giving weights proportional to the marginal contribution or to $f(T, S)$, reflecting a myopic willingness to join coalitions.

4. Your discussion of the f-Shapley value resembles the allocation rules studied by [16] and draws inspiration from [17], which in turn studies an allocation rule introduced by [22].
5. The presented α-Shapley value resembles the egalitarian Shapley values introduced by [10], and characterized by [2] based on additivity. Please clarify the relationship between the α-Shapley value and the egalitarian Shapley values. Also note the experimental evidence given by [4].
6. [7] and ensuing literature deals with characterizations of the Shapley value without efficiency.
7. Please provide a better survey of the related mathematical literature and other works that apply Hodge decompositions.

4. Other Comments and Suggestions

1. p1. "Recently, researchers have begun to use the Shapley value in machine learning [27]." Some further citations would be helpful, e.g., [11] and the ensuing literature. Moreover, there is a rich literature on the computation of the Shapley value which could benefit from being connected with the present paper.
2. p1. "Because it satisfies these axioms, the Shapley value is regarded as a unique and compelling solution to the problem of allocating the value of a cooperative game." I'm not sure about this. The characterization due to [24] seems more compelling.
3. p5, last paragraph. Please avoid name dropping. What are those papers about and how is it related to the current work?
4. p10. The notion of the null player axiom A3 resembles the null player out property by [6]. Moreover, "So $\Phi_i[v] \equiv 0$ is a consequence rather than a part of the axioms."—Then why is it in the axiom?
5. p10. Are the axioms in the theorem logical independent of each other? Please provide respective counterexamples.
6. p10. What is the advantage of A5 over A5'? After all, A5 is not used, neither for the proof nor for the interpretation. So why not replacing it with A5' throughout?
7. p12. Using S or $\varnothing$ for the starting coalition and T for the terminal/target coalition throughout your interpretation would help the reading.
8. p13. "its refection ω' from T to S" => You mean "its refection ω' from $S \cup \{i\}$ to $T \cup \{i\}$"?
9. p14. "is arguably the most important of the Shapley axioms" That is a strange statement. I'd say they are all equally important. This sentence needs some refinement.

10. p17. "The author believes that finding generalized Shapley axioms that can characterize the extended allocation scheme corresponding to the marginal value (5.5) and many other possible choices is an interesting question for future research." Can you provide such a characterization? I think [2] could be useful for the α-Shapley value.

11. p19, eq. (6.3): The last step, i.e., the 3rd equality is not obvious and deserves some explanation.

12. p20. Cancel "and the author's ability"

13. p30. "Notice then A3 (and ...)" => "Notice then A3 applied to player i (and ...)"

14. Is Φ invariant to dualization of the game? The dual v^D of a coalitional game v is given by $v^D(S) = v(N) - v(N \setminus S)$ for all $S \subseteq N$.

15. Section 6 on Generalized Nash-Kohlberg-Neyman's value for strategic games is poorly motivated. Either you can explain why this modification makes sense in the context of the motivation of the work of [11], or it might be better to drop the section and leave it for future work.

16. My intuition might not be as good as yours but maybe it makes sense to write ∂_i instead of d_i? This would at least ease the reading of the equation $\partial_i v = \mathrm{d}v_i + p$. I might however confuse notation and your choice of d_i could be well-motivated?

17. The subscript of $(v_i)_{i \in N}$ is sometimes missing => search for (v_i)

18. I suggest some modifications to simplify the exposition.

    a) Refer to the oriented incidence matrix B like [23]. This somewhat more concrete object eases the access to the setup. Later, when dealing with the f-Shapley value and Hodge allocations, please clarify how this affects the role and the content of B.

    b) It is often easier for the reader if the definition of an allocation rule and of the axioms are not part of the Theorem, but presented and interpreted beforehand. This would clarify the purpose of the theorem.

    c) In your description of the process, integrated contribution, and value function, it seems sufficient for the first part of the paper to omit superscript S and assume that we always start from the empty set $\varnothing$. The additional superscript can than be added in section 7 when it is needed. This might add some text but eases the access early on. Similar stepwise complification of the model and notation could help the uninitiated reader to gain access to your paper.

    d) I'd say that the definition of ϕ_f is a little surprising in that it does not contain any reference to player i. I reckon that you'd like to keep the notation concise but it is a little difficult since it is a new concept and a new framework at the same time. Perhaps, using ϕ_{f_i} would ease the understanding.

Bibliography

[1] R. Berman. Beyond the last touch: attribution in online advertising. *Marketing Science*, 37(5):771–792.

[2] A. Casajus and F. Huettner. Null players, solidarity, and the egalitarian Shapley values. *Journal of Mathematical Economics*, 49:58–61, 2013.

[3] A. Casajus and F. Huettner. Decomposition of solutions and the Shapley value. *Games and Economic Behavior*, 13(3):1–23, 2018.

[4] G. de Clippel and K. Rozen. Fairness through the lens of cooperative game theory: an experimental approach. 14(3):810–836.

[5] G. de Clippel and R. Serrano. Marginal contributions and externalities in the value. *Econometrica*, 76(6):1413–1436, 2008.

[6] J. J. M. Derks and H. H. Haller. Null players out? Linear values for games with variable supports. 1(3-4):301–314.

[7] E. Einy and O. Haimanko. Characterization of the Shapley-Shubik power index without the efficiency axiom. *Games and Economic Behavior*, 73(2):615–621, 2011.

[8] F. Gul. Bargaining foundations of shapely value. *Econometrica*, 57(1):81, 1989.

[9] S. Hart and A. Mas-Colell. Potential, value, and consistency. *Econometrica*, 57(3):589–614, 1989.

[10] R. Joosten. *Dynamics, equilibria and values*. PhD thesis, 1996.

[11] E. Kohlberg and A. Neyman. Cooperative strategic games. *Theoretical Economics*, 16(3):825–851, 2021.

[12] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. 30.

[13] B. McQuillin and R. Sugden. Backward induction foundations of the shapley value. *Econometrica*, 84(6):2265–2280, 2016.

[14] H. Moulin. *Fair Division and Collective Welfare*. The MIT Press.

[15] R. B. Myerson. Conference structures and fair allocation rules. *International Journal of Game Theory*, 9:169–182, 1980.

[16] C. C. Nembua. Linear efficient and symmetric values for TU-games: sharing the joint gain of cooperation. *Games and Economic Behavior*, 74(1):431–433, 2012-01.

[17] A. S. Nowak and T. Radzik. A solidarity value for n-person transferable utility games. 23:43–48.

[18] D. Ray and R. Vohra. Maximality in the farsighted stable set. *Econometrica*, 87(5):1763–1779, 2019.

[19] D. Ray and R. Vohra. Coalition formation. In *Handbook of Game Theory with Economic Applications*, pages 239–326. Elsevier.

[20] D. P. rez-Castrillo and D. Wettstein. Bidding for the surplus: a non-cooperative approach to the Shapley value. *Journal of Economic Theory*, 100(2):274–294, 2001.

[21] L. S. Shapley. A value for n-person games. In H. Kuhn and A. Tucker, editors, *Contributions to the Theory of Games*, volume II, pages 307–317. Princeton University Press, Princeton, 1953.

[22] Y. Sprumont. Population monotonic allocation schemes for cooperative games with transferable utility. *Games and Economic Behavior*, 2:378–394, 1990.

[23] A. Stern and A. Tettenhorst. Hodge decomposition and the shapley value of a cooperative game. *Games and Economic Behavior*, 113:186–198, 2019.

[24] H. P. Young. Monotonic solutions of cooperative games. *International Journal of Game Theory*, 14:65–72, 1985.