# AXIOMATIC ATTRIBUTION FOR MULTILINEAR FUNCTIONS

YI SUN AND MUKUND SUNDARARAJAN

Abstract. We study the attribution problem, that is, the problem of attributing a change in the value of a characteristic function to its independent variables. We make three contributions. First, we propose a formalization of the problem based on a standard cost sharing model. Second, in our most important technical contribution, we show that there is a unique *path attribution method* that satisfies `Affine Scale Invariance` and `Anonymity` if and only if the characteristic function is multilinear. We term this the *Aumann-Shapley-Shubik* method. Third, we study multilinear characteristic functions in detail; we describe a computationally efficient implementation of the Aumann-Shapley-Shubik method and discuss practical applications to pay-per-click advertising and portfolio analysis.

## 1. Introduction

1.1. **The Attribution Problem.** Consider a function $f(r_1, \ldots, r_n)$ of several variables $r_1, \ldots, r_n$. Given a change in the values of these variables, we ask what portion of the overall change is due to the change in each variable $r_i$. In particular, we would like to divide the responsibility for the overall change among the variables in an axiomatic way. We term such problems *attribution problems* and the responsibilities assigned attributions. The attribution to the $i^{\text{th}}$ variable can be more interesting than simply the change $s_i - r_i$ in the variable because the relationship between the magnitude of the change in a variable and the impact it has on $f$ depends on the form of $f$. In particular, a tiny change in a variable could have a huge impact on the the value of the function.

Formally, we are given a real-valued characteristic function $f : \mathbb{R}^n \to \mathbb{R}$ of $n$ variables and initial and final values $r_i$ and $s_i$ for the independent variables. Here, the function $f$ is deterministic, not learned from data, and the values of $r$ and $s$ are known exactly and are not estimates in any sense.[1] Our objective is to find attributions $z_1(r, s, f), \ldots, z_n(r, s, f)$, where we interpret $z_i(r, s, f)$ as the portion of the change in $f$ due to the change in the $i^{\text{th}}$ variable, so that $z_1(r, s, f) + \cdots + z_n(r, s, f) = f(s) - f(r)$, which we call the *completeness* condition on the attribution. We interpret completeness as meaning that all the change in $f$ is accounted for. (We often omit the characteristic function and simply write $z_i(r, s)$ for $z_i(r, s, f)$.)

As we discuss attribution, we will keep the following motivating example in mind.

**Example 1.1.** Consider a firm that repeatedly procures a good from a foreign supplier for use in its manufacturing process. It incurs some expenditure, the product $e = a \cdot p \cdot c$ of the amount $a$ of the good that the buyer purchases, the average cost per unit $p$ of the good in the foreign currency, and the conversion rate $c$ between the foreign and local currencies. We take $e(a, p, c)$ as the characteristic function. The final values of $e$, $a$, $p$, and $c$ may be statistics from a certain quarter, and the initial values may be statistics from the preceding quarter. The attribution problem, then, is to divide responsibility for the change in $e$ among the changes in $a$, $p$, and $c$.

Suppose further that the demand for the good comes from the manufacturing department (so an improvement in the efficiency of manufacturing reduces $a$), that the price for the good is negotiated by the procurement department (so an improvement in the negotiation process decreases $p$), and that the exchange rate is exogenously determined. Such an attribution could then serve to apportion blame between or determine bonuses for the two departments.

How can we attribute the change in the characteristic function $f$ to the various variables? If $f$ were linear, that is, if $f$ takes the form $f(r_1, \ldots, r_n) = \sum_i b_i r_i$, then for a change from $r$ to $s$, it is natural to attribute $b_i \cdot (s_i - r_i)$ to the $i^{\text{th}}$ variable. For non-linear functions such as the one in the Example 1.1, if the

*Date*: February 4, 2011.

[1]This is in contrast to the field of Structural Equation Modeling [5] where the function and the inputs are statistical quantities, and require model fitting and estimation. Some variables may also be latent and may require inference.

independent variables all change slightly, we could replace $b_i$ by the partial derivative with respect to $r_i$ at $s$, giving a linear approximation of $f$ locally at the final value and performing attribution as above.

However, if the changes in the variables are not slight, then this approach would badly violate *completeness*. For instance, in Example 1.1, suppose $a$ changes from 4 to 5, $p$ changes from 1 to 12, and $c$ changes from 1 to 1.5. Using this approach, the attributions to $a$, $p$, and $c$ are $(5-4)\cdot 12\cdot 1.5 = 18$, $5\cdot(12-1)\cdot 1.5 = 82.5$, and $5\cdot 12\cdot(1.5-1) = 30$, respectively. This assigns the two departments and the exogenous currency rate change blame for a total of $18 + 82.5 + 30 = 130.5$ of change; but the total change is only $5\cdot 12\cdot 1.5 - 4\cdot 1\cdot 1 = 86$. This means that the attribution violates completeness, making it difficult to interpret practically.

It might seem in this example that the failure of completeness originated from a poor choice of point approximation for the partial derivative of $f$. In general, no systematic use of such a point approximation suffices for our application. However, in Subsection 1.3, we will examine a principled method of computing attributions along these lines.

1.2. **Axioms for attribution methods.** Our attribution problem (almost trivially) generalizes the cost or surplus sharing problem from the social choice literature (cf. Moulin [15]), where the problem is to axiomatically share the cost of production or surplus among several agents. The characteristic function is either cost or surplus, the independent variables correspond to demands or contributions of agents, and the attributions correspond to cost shares or profit shares. The *completeness* condition is the *budget-balance* condition for cost sharing. We give a more detailed discussion of the relationship between these two problems in Subsection 1.7.

Following the cost sharing literature, we take an axiomatic approach to choosing methods to use for attribution problems. We discuss here the axioms we consider and briefly discuss motivations for them, emphasizing the attribution context; see the cited papers for a longer discussion in the cost sharing context. We note that all of these axioms are commonly considered ones from the cost sharing literature.

- `Additivity`: For all $r, s, f_1, f_2$, we have that $z_i(r, s, f_1 + f_2) = z_i(r, s, f_1) + z_i(r, s, f_2)$.
  This axiom yields a type of procedural invariance. That is, if the system modeled by the characteristic function can be decomposed into several independent sub-processes that interact additively, we can compute the attributions separately for each sub-process.
- `Dummy`: If the value of the characteristic function does not depend on a variable, then the attribution to that variable is zero.
  This axiom is very natural, as it simply requires that variables irrelevant to the outcome be ignored.[2]
- `Scale Invariance` [10]: The attributions are independent of linear rescaling of individual variables. That is, for any $c > 0$, if $g(r_1, \ldots, r_n) = f(r_1, \ldots, r_j/c, \ldots, r_n)$, then for all $i$ we have
  $$z_i(r, s, f) = z_i\Big((r_1, \ldots, cr_j, \ldots, r_n), (s_1, \ldots, cs_j, \ldots, s_n), g\Big).$$
  `Scale Invariance` conveys the idea that the attributions should be independent of the (possibly incomparable) units in which individual variables are measured. It is especially compelling in the context of attribution because the different variables may refer to quantities of entirely different things.
- `Affine Scale Invariance` [24] The attributions are invariant under simultaneous affine transformation of the characteristic function and the variables. That is, for any $c, d > 0$, if $g(r_1, \ldots, r_n) = f(r_1, \ldots, (r_j - d)/c, \ldots, r_n)$, then for all $i$ we have
  $$z_i(r, s, f) = z_i\Big((r_1, \ldots, cr_j + d, \ldots, r_n), (s_1, \ldots, cs_j + d, \ldots, s_n), g\Big).$$
  `Affine Scale Invariance` conveys the idea that both the units and the zero point of individual variables should not affect the value of the attribution. Again, for attribution this is especially compelling, since the variables may represent values without naturally defined units or zero points (For example, temperature is commonly measured in both Celsius and Fahrenheit scales, which are related by an affine transformation.).

[2]In the cost sharing context, it is the bed-rock of the no cross-subsidy (full-responsibility) theory. In such a theory, the variable is deemed responsible for asymmetries in the cost as well as asymmetries in the demand; see Moulin and Sprumont [16] for a discussion.

- `Monotonicity` [10]: Suppose the characteristic function $f$ is non-decreasing in variable $j$. Then, for input pairs $(r, s)$ and $(r, s')$ such that $s_i = s'_i$ for $i \neq j$ and $s_j < s'_j$, we have $z_j(r, s, f) \leq z_j(r, s', f)$. `Monotonicity` precludes attributions with counterintuitive signs.
- `Anonymity`: The attributions are unchanged by relabeling of the variables. More formally, for any permutation $\sigma \in S_n$, if $f_\sigma(r_1, \ldots, r_n) = f(r_{\sigma^{-1}(1)}, \ldots, r_{\sigma^{-1}(n)})$, then for all $i$, we have
$$z_{\sigma^{-1}(i)}(r_{\sigma(1)}, \ldots, r_{\sigma(n)}, s_{\sigma(1)}, \ldots, s_{\sigma(n)}, f_\sigma) = z_i(r, s, f).$$
`Anonymity` conveys the idea that all variables in the characteristic function should be treated equally.

We now briefly describe how we use these axioms in this paper. We include the `Additivity`, `Dummy`, `Scale Invariance`, and `Monotonicity` axioms only to facilitate discussion and comparison with axiomatizations of attribution methods in the cost sharing literature. They play no role in the main results, for which only the `Affine Scale Invariance` and `Anonymity` axioms are used, `Anonymity` to disallow methods which are inherently asymmetric, and `Affine Scale Invariance` to impose continuity on the attributions.

### 1.3. Path and affine path attribution methods.

We now consider a natural, well-studied class of attribution methods, the path attribution methods (see [10, 11]), which assign to each variable its marginal effect along some path from the initial point to the final point. Such attribution methods are analogous to the partial-derivative based approach outlined in the first subsection, and salvage completeness by integrating the partial derivatives along a path instead of taking a naive estimate at a single endpoint. This definition is motivated by Theorem 1.6 from the cost sharing literature, which we discuss in Subsection 1.7.

**Definition 1.2.** For each $r, s \in \mathbb{R}^n$, let $\gamma_{r,s} : [0, 1] \to \mathbb{R}^n$ be a $C^1$-function with $\gamma_{r,s}(0) = r$ and $\gamma_{r,s}(1) = s$, which we interpret as a path from $r$ to $s$. Write $\gamma_{r,s} = (\gamma_{r,s,1}, \ldots, \gamma_{r,s,n})$, and let $\gamma_{r,s,i}$ be non-decreasing if $r_i \leq s_i$ and non-increasing if $r_i \geq s_i$. Then, the attribution method given by

$$z_i(r, s) = \int_0^1 \partial_i f(\gamma_{r,s}(t))\gamma'_{r,s,i}(t)dt$$

is the *single-path attribution method* corresponding to the family of paths $\gamma_{r,s}$. If the method

$$z_i(r, s) = \sum_j c_j z_i^j(r, s) \text{ for } c_j \geq 0 \text{ and } \sum_j c_j = 1$$

is a convex combination of single-path attribution methods $z^j$, we say that $z$ is a *path attribution method*.

For a single-path attribution method, we may check by the gradient theorem that

$$z_1(r, s) + \cdots + z_n(r, s) = \int_0^1 \sum_{i=1}^n \partial_i f(\gamma_{r,s}(t))\gamma'_{r,s,i}(t)dt = \int_{\gamma_{r,s}} \nabla \cdot f = f(s) - f(r),$$

meaning that completeness is satisfied for each single-path attribution method. Completeness is preserved under convex combinations and therefore holds for all path attribution methods. Notice that all path attribution methods satisfy `Additivity` and `Dummy`.

For any attribution method, we would like intuitively that the attributions remain stable when the input changes by a small amount. For instance, we would like the attributions to be robust to small errors in measurement of the inputs. In the case of path attribution methods, this would follow naturally from a continuity condition on the family of paths $(r, s) \mapsto \gamma_{r,s}$. Indeed, we may identify the following class of methods which automatically give this desired continuity.[3]

**Definition 1.3.** Fix a $C^1$ path $\gamma : [0, 1] \to [0, 1]^n$, non-decreasing in each variable, such that $\gamma(0) = 0$ and $\gamma(1) = (1, \ldots, 1)$. Write $\gamma = (\gamma_1, \ldots, \gamma_n)$. Then, the single-path attribution method corresponding to

$$\gamma_{r,s}(t) = r + \Big((s_1 - r_1)\gamma_1(t), \ldots, (s_n - r_n)\gamma_n(t)\Big)$$

is the *affine single-path attribution method* corresponding to $\gamma$. If $z$ is a convex combination of affine single-path attribution methods, we say that $z$ is an *affine path attribution method*.

[3] Of course, there may be other classes of methods which also yield continuity.

Evidently, the paths $\gamma_{r,s}$ in an affine single-path attribution method change smoothly with $r$ and $s$, as they are compatible with affine transformations. More precisely, let $z_i$ be the affine single-path attribution method corresponding to $\gamma$. For any $c, d > 0$, set $g(r_1, \ldots, r_n) = f(r_1, \ldots, (r_j - d)/c, \ldots, r_n)$, $r' = (r_1, \ldots, cr_j + d, \ldots, r_n)$, and $s' = (s_1, \ldots, cs_j + d, \ldots, s_n)$. Then, taking $\tau_{ij}(c) = c$ if $i = j$ and $\tau_{ij}(c) = 1$ otherwise, we have

$$\begin{aligned} z_i\Big(r', s', g\Big) &= \int_0^1 \partial_i g\Big(r' + \big((s_1 - r_1)\gamma_1(t), \ldots, c(s_j - r_j)\gamma_j(t), \ldots, (s_n - r_n)\gamma_n(t)\big)\Big)(s_i - r_i)\tau_{ij}(c)\gamma_i'(t)\, dt \\ &= \int_0^1 \frac{1}{\tau_{ij}(c)}\, \partial_i f(\gamma_{r,s}(t))\, (s_i - r_i)\, \tau_{ij}(c)\, \gamma_i'(t)\, dt \\ &= \int_0^1 \partial_i f(\gamma_{r,s}(t))\gamma_{r,s,i}'(t)\, dt \\ &= z_i(r, s, f). \end{aligned}$$

This shows exactly that any affine single-path attribution method and hence any affine path attribution method satisfies `Affine Scale Invariance`. Therefore, for a fixed characteristic function, any path attribution method satisfying `Affine Scale Invariance` produces the same attributions as some affine path attribution method. As we discuss in the following subsection, several commonly used attribution methods are affine path methods, and we will focus most of our attention in this paper on affine path attribution methods.

1.4. **Some known attribution methods.** In this subsection, we discuss two attribution methods adapted from the two most important cost sharing methods and point out some of their axiomatic properties. We start with the method that is arguably the best known method in the cost sharing literature.

**Definition 1.4.** The *Aumann-Shapley* method [3] is the affine single-path attribution method corresponding to the path $\gamma_i(t) = t$.

This method is a generalization of the Shapley value (see [21]) and was identified by Aumann and Shapley [3] as a 'value' for non-atomic games. We note that it trivially satisfies `Affine Scale Invariance` as an affine path attribution method. Now, we define a different and arguably more direct generalization of the Shapley value.

**Definition 1.5.** The *Shapley-Shubik* method [22, 10] is defined as follows. For any $\sigma \in S_n$, let $\gamma^\sigma$ be the path

$$\gamma_i^\sigma(t) = \begin{cases} 0 & tn < \sigma(i) - 1 \\ (tn - \sigma(i)) & \sigma(i) - 1 \leq tn < \sigma(i) \\ 1 & tn \geq \sigma(i), \end{cases}$$

where $\gamma^\sigma$ walks along edges of the hypercube $[0,1]^n$ in an order determined by $\sigma$. Then, the Shapley-Shubik method is given by the average of the $n!$ affine path attribution methods corresponding to $\gamma^\sigma$. More generally, a *random order method* [18, 10] is any convex combination of the affine path attribution methods corresponding to the $\gamma^\sigma$.

In other words, the Shapley-Shubik attribution from $r$ to $s$ is the expected attribution from a monotone random walk along the edges of the hypercube with opposite vertices at $r$ and $s$.[4] One desirable characteristic of the Shapley-Shubik method is that it satisfies `Monotonicity` for all characteristic functions, while there are characteristic functions for which the Aumann-Shapley method does not satisfy `Monotonicity` (see [10] for specific examples).

[4] The following related characterization of the Aumann-Shapley method is instructive. Subdivide the hypercube with opposite vertices at $r$ and $s$ into smaller hypercubes and consider monotonic random walks in this structure. The density of the resulting walks will be focused on the diagonal. Then, if the characteristic function satisfies some basic regularity conditions, the average of the path attribution methods corresponding to these walks will be equivalent to the Aumann-Shapley method in the limit.

1.5. **Attribution for specific characteristic functions.** When choosing an attribution method, it is very desirable to have a uniqueness result, one which says that there is exactly one method satisfying some axioms, because this identifies a method for use. If an attribution method is the unique method satisfying some axioms, we term this an *axiomatization* for the attribution method.

In this paper, we seek axiomatization results that hold for individual characteristic functions. In particular, we say that two attribution methods $z^1$ and $z^2$ are *equivalent* for a characteristic function $f$ if they give the same attributions for all sets of input variables. For specific characteristic functions, we wish to find attribution methods which are unique up to equivalence, as these are the type of uniqueness results which are useful in practice.

The following classes of characteristic functions will play a prominent role in our results.

- *Additively separable:* A function $f : \mathbb{R}^n \to \mathbb{R}$ is additively separable if there exist $f_i : \mathbb{R} \to \mathbb{R}$ with
$$f(r_1, \ldots, r_n) = f_1(r_1) + \cdots + f_n(r_n).$$
- *Multilinear:* A function $f : \mathbb{R}^n \to \mathbb{R}$ is multilinear if we may write $f$ in the form
$$f(r_1, \ldots, r_n) = \sum_{I \subset [n]} c_I \prod_{i \in I} r_i,$$
that is, as the sum of monomials of degree at most 1 in each variable.

We are now ready to state our results.

1.6. **Statement of results.** We begin by asking which characteristic functions admit a unique path attribution method. In Section 2, we show in Theorem 2.1 that such a unique method exists if and only if the characteristic function is additively separable. The space of such functions is uninteresting from an attribution standpoint, as the overall change in the characteristic function is obviously representable as the sum of the changes due to each independent variable. We view this as a trivial case where it is too much to ask for a unique path attribution method. The proof is an application of Stokes' Theorem from vector calculus (see Appendix A for a review of Stokes' theorem, which is the main technical tool of this paper).

We then restrict our attention to the smaller class of path attribution methods satisfying `Anonymity` and `Affine Scale Invariance`. In Corollary 3.4, we obtain our main result, that there is a unique path attribution method satisfying `Anonymity` and `Affine Scale Invariance` if and only if the characteristic function is the sum of an additively separable function and a multilinear function. We find this result interesting because such functions are non-trivial from an attribution perspective and have applications (see Section 4).

The proof involves two steps, both of which are of independent interest. First, we show in Theorem 2.2 that every affine path attribution method is equivalent to a random order method for multilinear functions via a flow argument (Recall that for a fixed characteristic function every path attribution method satisfying `Affine Scale Invariance` is equivalent to an affine path attribution method). This places a large class of attribution methods into the very structured class of random order methods, which are in particular computable in finite time. To complete the proof in the forward direction, we simply note that that the Shapley-Shubik method (Definition 1.5) is the unique random-order method that satisfies `Anonymity` (Corollary 2.3).

Interestingly, Theorem 3.1 shows that the Aumann-Shapley (Definition 1.4) and Shapley-Shubik methods coincide for these characteristic functions. This makes clear that the unique method that we identify for these characteristic functions is equivalent to both the Aumann-Shapley and Shapley-Shubik methods. We call it the *Aumann-Shapley-Shubik* method and give an efficient algorithm to compute it in Theorem 3.5 and Corollary 3.6.

For the reverse direction, we show in Theorem 3.3 that the Aumann-Shapley and Shapley-Shubik methods are equivalent only if the characteristic function is the sum of a multilinear function and an additively separable function, which completes the proof because both the Aumann-Shapley and Shapley-Shubik methods satisfy `Anonymity` and `Affine Scale Invariance`. Finally, we illustrate the practical relevance of our results. We discuss practical applications to pay-per-click advertising and portfolio analysis and give several examples of useful insights which can be obtained by applying attribution with multilinear characteristic functions.

1.7. **Attribution versus cost sharing.** In this subsection, we discuss the relation between our attribution problem and the classical cost sharing problem.

1.7.1. *Cost sharing as attribution.* Cost sharing models come in various flavors depending on whether the demands are binary, integral, or real-valued and whether the cost function is homogeneous or not (see Moulin [15] for a classification). In this sense our model resembles the $rr$, heterogeneous cost sharing model (both the characteristic function and the independent variables are real-valued, and the characteristic function is not homogeneous in the variables). More precisely, for a monotonically increasing cost function, $rr$ heterogeneous cost sharing is equivalent to attribution from 0 to the final demand.

There are two immediate differences between attribution and cost sharing. First, in the attribution problem, variables change from one set of values to another, while in cost sharing there is just a single set of demands or contributions. Secondly, attribution relaxes the requirement that the cost function be monotone. Therefore, while negative cost shares do not make sense, negative attributions can make sense in some contexts.

1.7.2. *Axiomatics for cost sharing.* Here we discuss axiomatization results from the cost sharing literature, both as motivation for some of our assumptions and as context for our results. We begin with a result identifying the analogue of the path attribution methods in cost sharing as exactly those methods satisfying the most basic of the axioms we introduced in Subsection 1.2. This motivates us to restrict our attention to such methods in this paper.

**Theorem 1.6** (Theorem 1 of [9])**.** Any cost sharing method satisfying `Dummy` and `Additivity` is a path cost sharing method.

We now give two axiomatizations of the Aumann-Shapley and Shapley-Shubik methods in the cost sharing context. The Aumann-Shapley method was axiomatized by Billera and Heath [4] and Mirman [14] in the following theorem.

**Theorem 1.7** ([4, 10, 14])**.** The Aumann-Shapley method is the unique cost sharing method that satisfies `Additivity`, `Dummy`, `Scale Invariance`, and `Average Cost for Homogeneous Goods`, which states that, for cost functions that are a function of the sum of the demands, the cost shares should be proportional to the demands.

For the Shapley-Shubik method, we have the following axiomatization given by Friedman and Moulin [10].

**Theorem 1.8** (Theorem 1 of [10])**.** Any cost sharing method that satisfies `Additivity`, `Dummy`, `Monotonicity`, `Scale Invariance`, and `Continuity at Zero`, which states that the cost shares are continuous in each variable near 0, is a random order method. The Shapley-Shubik method is the unique cost sharing method that satisfies `Anonymity` in addition to `Additivity`, `Dummy`, `Monotonicity`, `Scale Invariance`, and `Continuity at Zero`.

**Remark.** In the attribution context, the Aumann-Shapley and Shapley-Shubik methods satisfy the axioms of Theorems 1.7 and 1.8, but it is not clear if the uniqueness properties continue to hold. We suspect that these results should also carry over to the attribution framework (with very similar proofs) after some appropriate modification.

1.7.3. *Axiomatization for attribution versus axiomatization for cost sharing.* Our approach to the axiomatic study of attribution methods differs from that taken in the cost sharing literature. A typical axiomatic result in the cost sharing literature (like Theorems 1.7 and 1.8) identifies a certain cost sharing method as the unique method that satisfies certain axioms for *all* cost functions (cf. [10, 21]). This does not preclude the existence of multiple methods that satisfy the same set of axioms for a certain subclass of cost functions. For instance, Redekop [17] notes that the Aumann-Shapley cost sharing method satisfies the axioms mentioned in the uniqueness result for the Shapley-Shubik method (Theorem 1.8) when the cost function has increasing marginal costs.[5]

In our model, the characteristic function is known when the attribution method is selected, and so general uniqueness results similar to Theorems 1.7 and 1.8 are not necessarily sufficient to guide the selection of an attribution method. For instance, for a specific convex characteristic function, analogues of these theorems would not be enough to select a unique method, illustrating an advantage of our framework.

[5]It is easy to show directly that all path attribution methods satisfying `Scale Invariance` are monotone for convex functions, of which multilinear functions with positive coefficients are an instance. Redekop [17] notices this for the Aumann-Shapley attribution method.

In addition, having a known and fixed characteristic function allows us to be more parsimonious with axioms. In the case of multilinear functions, our main result allows us to characterize the Aumann-Shapley method without using `Average Cost for Homogeneous Goods`, a 'partial domain axiom,' which, as Friedman and Moulin [10] argue, is not very natural.

1.8. **Notations.** We write $[n]$ for the set $\{1, 2, \dots, n\}$. For a set of variables $c_1, \dots, c_n$ and a subset $I \subset [n]$, write $c$ for the $n$-tuple $(c_1, \dots, c_n)$ and $c_I$ for the product $\prod_{i \in I} c_i$ over the indices in $I$. For two sets of variables $c_1, \dots, c_n$ and $d_1, \dots, d_n$, we write $c < d$ (resp. $c \leq d$) if $c_i < d_i$ (resp. $c_i \leq d_i$) for all $i$. We write $[c, d]$ for the closed box $I_1 \times \cdots \times I_n$, where $I_i$ is the closed interval bounded by $c_i$ and $d_i$. We use 0 to denote the vector $(0, \dots, 0)$ containing all 0's. The length of this vector will always be clear from context. For a function $f : \mathbb{R}^n \to \mathbb{R}$ and a multi-set of indices $\alpha$, we denote by $\partial_\alpha f$ the mixed partial derivative with respect to the indices in $\alpha$. In all cases where this construction appears, we will assume that $f$ is chosen so that Young's theorem on the equality of mixed partial derivatives holds.

## 2. Uniqueness of path methods and affine path methods

In this section, we consider the question of when there exists a unique path attribution method for a particular characteristic function. We begin with a negative result showing that if all path attribution methods agree for a particular $f$, then $f$ is additively separable, which we view as a trivial case. We then enrich this class of characteristic functions by allowing $f$ to be the sum of an additively separable function and a multilinear one. In this case, we find that there exists a unique path attribution method satisfying `Anonymity` and `Affine Scale Invariance`, where we view `Anonymity` as removing non-uniqueness stemming from inherent asymmetry of paths rather than the characteristic function itself and `Affine Scale Invariance` as removing non-uniqueness stemming from a lack of continuity among paths.

2.1. **Comparing path methods.** We begin by summarizing our general approach to the problem of comparing path attribution methods, which will consist of two elementary but crucial technical steps. Consider a path attribution method corresponding to a family of paths $\gamma_{r,s}$. Letting $I_{r,s}$ be the (closed) image of $\gamma_{r,s}$ in $[r, s]$, we see that the attributions are given by

$$z_i(r, s) = \int_0^1 \partial_i f(\gamma_{r,s}(t)) \gamma'_{r,s,i}(t) dt = \int_{I_{r,s}} \partial_i f(r) dr_i, \tag{2.1}$$

where we view $\partial_i f(r)\, dr_i$ as a differential form on $I_{r,s}$. From this perspective, it is clear that $z_i(r, s)$ depends only on the underlying set $I_{r,s}$ of the path and not on the choice of parametrization $\gamma_{r,s}$.

We can now use this viewpoint to compare methods. For the sake of clarity, we restrict to the case where $n = 2$ in this preliminary discussion. Let $\gamma^1_{r,s}$ and $\gamma^2_{r,s}$ be two families of paths and consider the corresponding path attribution methods. If these methods coincide for some characteristic function $f$, then for all $r, s$, we have for all $i$ that

$$z_i(r, s, f) = \int_{I^1_{r,s}} \partial_i f dr_i = \int_{I^2_{r,s}} \partial_i f dr_i,$$

where $I^1_{r,s}$ and $I^2_{r,s}$ are the images in $[r, s]$ of $\gamma^1_{r,s}$ and $\gamma^2_{r,s}$, respectively. Write $I^1_{r,s} \cup I^2_{r,s}$ for the closed curve formed by first traversing $\gamma^1_{r,s}$ and then traversing $\gamma^2_{r,s}$. Because $n = 2$, $I^1_{r,s} \cup I^2_{r,s}$ is the boundary of some closed 2-dimensional submanifold with boundary $A_{r,s}$ of $[r, s]$. More precisely, there are closed 2-dimensional submanifolds with boundary $B_{r,s}$ and $C_{r,s}$ of $[r, s]$ such that

- $B_{r,s} \cup C_{r,s} = A_{r,s}$,
- the winding number of $I^1_{r,s} \cup I^2_{r,s}$ about each point in $B_{r,s}$ is 1,
- and the winding number of $I^1_{r,s} \cup I^2_{r,s}$ about each point in $C_{r,s}$ is $-1$.

In other words, $B_{r,s}$ and $C_{r,s}$ are the regions enclosed by $I^1_{r,s} \cup I^2_{r,s}$ with positive and negative signed area, respectively. Notice that $I^1_{r,s} \cup I^2_{r,s}$ cannot have a winding number other than 1 or $-1$ about any point in $A_{r,s}$ because $\gamma^1_{r,s}$ and $\gamma^2_{r,s}$ are both non-decreasing. Therefore, we find from (2.1) that

$$0 = \int_{I^1_{r,s}} \partial_2 f dr_2 - \int_{I^2_{r,s}} \partial_2 f dr_2 = \int_{B_{r,s}} \partial_{12} f dr_1 dr_2 - \int_{A_{r,s}} \partial_{12} f dr_1 dr_2, \tag{2.2}$$

where the final equality follows from Stokes' Theorem. We have therefore translated condition (2.1) involving line integrals to condition (2.2) involving area integrals. This conversion will feature prominently in our later approach.

2.2. **A negative result: When are all path attribution methods equivalent?** We now consider when all path attribution methods are equivalent, finding in Theorem 2.1 that this occurs only for the trivial case of additively separable characteristic functions.

**Theorem 2.1.** A characteristic function $f$ is additively separable if and only if every path attribution method agrees on $f$.

*Proof.* First, suppose that $f$ is additively separable. Then, for any map $(r,s) \mapsto \gamma_{r,s}$, the path attribution method for $\gamma_{r,s}$ gives the attributions

$$z_i = \int_0^1 \partial_i f(\gamma_{r,s}(t))\gamma'_{r,s,i}(t)dt = \int_0^1 f'_i(\gamma_{r,s,i}(t))\gamma'_{r,s,i}(t)dt = \int_{\gamma_{r,s,i}(0)}^{\gamma_{r,s,i}(1)} f'_i(y)dy = f_i(r_i) - f_i(s_i),$$

which are independent of the choice of $\gamma_{r,s}$. Hence, every path attribution method agrees on $f$.

Now, suppose every path attribution method agrees on $f$. We first claim that $f$ is additively separable if and only if $\partial_{ij} f = 0$ for $i,j$ distinct. The "if" direction is clear. For the "only if" direction, the condition implies that we may write $\partial_i f = g_i$ for $g_i$ a function of $r_i$ only. This means that

$$f(r_1,\dots,r_n) - f(r_1,\dots,r_{i-1},0,r_{i+1},\dots,r_n) = \int_0^{r_i} g_i(t)dt$$

is a function of $r_i$ alone. Thus, $f$ takes the desired form with $f_i(r_i) = \int_0^{r_i} g_i(t)dt$.

Now, we claim that $\partial_{ij} f = 0$. Consider two path attribution methods that for $r = (r_1,\dots,r_n)$ and $s = (r_1,\dots,r_{i-1},s_i,r_{i+1},\dots,r_{j-1},s_j,r_{j+1},\dots,r_n)$ correspond to the paths

$$\gamma^1_{r,s}(t) = \begin{cases} (r_1,\dots,r_i + 2t(s_i - r_i),\dots,r_j,\dots,r_n) & t \le 1/2 \\ (r_1,\dots,s_i,\dots,r_j + (2t-1)(s_j - r_j),\dots,r_n) & t > 1/2 \end{cases}$$

and

$$\gamma^2_{r,s}(t) = \begin{cases} (r_1,\dots,r_i,\dots,r_j + 2t(s_j - r_j),\dots,r_n) & t \le 1/2 \\ (r_1,\dots,r_i + (2t-1)(s_i - r_i),\dots,s_j,\dots,r_n) & t > 1/2. \end{cases}$$

Let $I^1_{r,s}$ and $I^2_{r,s}$ be the images of $\gamma^1_{r,s}$ and $\gamma^2_{r,s}$, respectively. The two methods agree by assumption, so we find for all $r$ and $s$ that

$$0 = z^1_j(r,s) - z^2_j(r,s) = \int_{I^1_{r,s}} \partial_j f dr_j - \int_{I^2_{r,s}} \partial_j f dr_j = \int_{A_{r,s}} \partial_{ij} f dr_i dr_j \tag{2.3}$$

by Stokes' Theorem, where $A_{r,s}$ is the rectangle with opposite corners at $r$ and $s$.

Now, suppose for contradiction that $\partial_{ij} f$ is not identically zero. Then, without loss of generality, we can find some $z$ with $\partial_{ij} f(z) > 0$ and, by continuity of $\partial_{ij} f$, an open neighborhood $U$ of $z$ such that $\partial_{ij} f|_U > 0$. Then, we may find $r < s \in U$ such that $[r,s] \subset U$ and $r_k = s_k = z_k$ for $k \neq i,j$. For these values of $r,s$, we have that $\int_{A_{r,s}} \partial_{ij} f dr_1 dr_2 > 0$, a contradiction. So $\partial_{ij} f$ must be identically zero, as needed. □

Theorem 2.1 suggests that asking for a unique path attribution method is too strong a condition to ask for a characteristic function that is non-trivial (which we interpret to mean not additively separable). We wish to obtain uniqueness for a wider class of characteristic functions, meaning that we must restrict the space of path attribution methods that we consider. Let us analyze the proof of Theorem 2.1 to give some motivation for the restrictions we will later impose. The proof proceeds by taking a family of pairs of explicitly defined path attribution methods (corresponding to $\gamma^1_{r,s}$ and $\gamma^2_{r,s}$) and showing that they are equal only when the characteristic function $f$ is additively separable. The key point to the proof is that $\gamma^1_{r,s}$ and $\gamma^2_{r,s}$ treat variables $r_i$ and $r_j$ in essentially different ways, suggesting that we might try imposing `Anonymity` to eliminate this possibility.

However, `Anonymity` alone is insufficient because it still allows too much freedom in our choice of path attribution method. We illustrate this with an example in the case $n = 2$. Specify a family of paths

$\gamma_{r,s}$ as follows. First, for $s_1 - r_1 < s_2 - r_2$, choose arbitrary paths for $\gamma_{r,s}$. For $s_1 - r_1 = s_2 - r_2$, take $\gamma_{r,s}(t) = (1-t)r + ts$, the straight-line path between $r$ and $s$. Then, for $s_1 - r_1 > s_2 - r_2$, we set

$$\gamma_{(r_1,r_2),(s_1,s_2),1}(t) = \gamma_{(r_2,r_1),(s_2,s_1),2}(t) \text{ and } \gamma_{(r_1,r_2),(s_1,s_2),2}(t) = \gamma_{(r_2,r_1),(s_2,s_1),1}(t).$$

This gives a well-defined family of paths which yields a path attribution method which trivially satisfies `Anonymity`. However, this family may include *arbitrary* paths $\gamma_{r,s}$ for $s_1 - r_1 < s_2 - r_2$! In particular, if we are careful with our choice of $r$ and $s$, the argument of Theorem 2.1 will work verbatim for these path attribution methods.

Why was the imposition of `Anonymity` not sufficient? The problem lay in the discontinuity allowed in our constructed family of paths $\gamma_{r,s}$. In our construction, $\gamma_{0,(1,1+\varepsilon)}$ and $\gamma_{0,(1,1)}$ can differ quite radically, with $\gamma_{0,(1,1)}$ being a relatively symmetric path but $\gamma_{0,(1,1+\varepsilon)}$ an arbitrary one. To address this issue, we impose in the next subsection a continuity condition on our paths.

2.3. **Affine path attribution methods for multilinear functions.** Following our axiomatic approach, we would like to impose this continuity condition on paths via an axiom on our attribution methods. A natural candidate, then, is `Affine Scale Invariance`, as it is a continuity condition on attributions and has a very natural interpretation in the attribution context. On any given characteristic function, any path attribution method satisfying `Affine Scale Invariance` is obviously equivalent to an affine path attribution method, as the single-path attribution method for the family of paths $\gamma_{r,s}$ is equivalent in this case to the affine single-path attribution method for $\gamma_{0,1}$.

We restrict our attention for now to these affine path attribution methods. Further, allowing our characteristic function $f$ to be the sum of a multilinear function and an additively separable function, the following theorem shows that the affine path attribution methods on $f$ take a canonical form, that of the random order methods.

**Theorem 2.2.** Suppose that $f$ is the sum of a multilinear function and an additively separable function. For $\gamma : [0,1] \to [0,1]^n$ with $\gamma_i$ monotonically increasing for each $i$, the affine path attribution method for $\gamma$ is a random order method (Definition 1.5) on $f$.

*Proof.* Our proof is based upon an explicit computation. By `Additivity`, `Dummy`, and Theorem 2.1, we may reduce to the case where $f(r_1, \ldots, r_n) = r_1 \cdot \cdots \cdot r_n$. In this case, the attributions are given by

$$\begin{aligned}
z_i(r,s) &= \int_0^1 \partial_i f(\gamma_{r,s}(t))\gamma'_{r,s,i}(t)dt \\
&= \int_0^1 (s_i - r_i) \prod_{j \neq i} [r_j + (s_j - r_j)\gamma_j(t)]\, \gamma'_i(t)dt \\
&= (s_i - r_i) \int_0^1 \gamma'_i(t) \sum_{K \subset [n]-\{i\}} \prod_{j \in K} r_j \prod_{j \in [n]-\{i\}-K} (s_j - r_j)\gamma_j(t)dt \\
&= (s_i - r_i) \int_0^1 \gamma'_i(t) \sum_{J \subset [n]-\{i\}} \prod_{j \in J} s_j \prod_{j \in [n]-J-\{i\}} r_j \sum_{J \subset K \subset [n]-\{i\}} (-1)^{|K|-|J|} \prod_{j \in K} \gamma_j(t)dt \\
&= (s_i - r_i) \sum_{J \subset [n]-\{i\}} \prod_{j \in J} s_j \prod_{j \in [n]-J-\{i\}} r_j \int_0^1 \gamma'_i(t) \prod_{j \in J} \gamma_j(t) \prod_{j \in [n]-J-\{i\}} (1-\gamma_j(t))dt,
\end{aligned}$$

which is of the form

$$z_i(r,s) = (s_i - r_i) \sum_{J \subset [n]-\{i\}} c_{i,J} \prod_{j \in J} s_j \prod_{j \in [n]-J-\{i\}} r_j$$

for the constants

$$c_{i,J} = \int_0^1 \gamma'_i(t) \prod_{j \in J} \gamma_j(t) \prod_{j \in [n]-J-\{i\}} (1-\gamma_j(t))dt$$

which depend only on $\gamma$, $J$, and $i$. Because $\gamma'_i(t) \geq 0$ for all $t$ and $\gamma_j(t) \in [0,1]$, we see that $c_{i,J} \geq 0$.

Now, consider the directed graph $\Gamma$ with vertex set $\{0,1\}^n$ and an edge from a vertex $u$ to a vertex $v$ if and only if $v \geq u$ and the coordinates of $u$ and $v$ differ in exactly one coordinate.[6] Notice that a vertex $v$ in

[6] These vertices correspond to the vertices of the hypercube $[0,1]^n$ in Definition 1.5.

$\Gamma$ may be labeled by the subset $J \subset [n]$ of coordinates which have value 1. Then, all edges are between $v_J$ and $v_{J\cup\{i\}}$ for some $i \notin J$, so we may label each edge by the subset $J \subset [n]$ corresponding to its source and the additional coordinate $i$ which has value 1 in its target. Assign flow $c_{i,J}$ to this edge.

Now, completeness implies that $\sum_{i=1}^n z_i = \prod_i s_i - \prod_i r_i$ and hence that

$$\prod_i s_i - \prod_i r_i = \sum_{i=1}^n z_i = \sum_{i=1}^n (s_i - r_i) \sum_{J\subset[n]-\{i\}} c_{i,J} \prod_{j\in J} s_j \prod_{j\in[n]-J-\{i\}} r_j$$
$$= \sum_{J\subset[n]} \prod_{j\in J} s_j \prod_{j\notin J} r_j \left( \sum_{j\notin J} c_{j,J} - \sum_{j\in J} c_{j,J-\{j\}} \right)$$

for all choices of $r \le s$. Hence, we have an equality of polynomials. Equating coefficients, we see that $\sum_{j\notin J} c_{j,J} = \sum_{j\in J} c_{j,J-\{j\}}$ for $J$ other than $\emptyset$ and $[n]$, that $c_{j,\emptyset} = 1$, and that $\sum_{j\in[n]} c_{j,[n]-\{j\}} = 1$. Together, these imply that the non-negative flows $c_{i,J}$ give rise to a feasible flow network with total flow 1 from $v_\emptyset$ to $v_{[n]}$. Therefore, this flow may be decomposed as the sum of flows along paths from $v_\emptyset$ to $v_{[n]}$, giving exactly the desired decomposition of our attribution method $z$ as the convex combination of the methods $z^\sigma$. So $z$ is a random-order method, as desired. □

This result is somewhat surprising because random order methods are inherently combinatorial and may be evaluated using the values of the characteristic function at a finite set of points, while affine path methods require a continuous evaluation of the characteristic function. Thus we see that the form of the characteristic function in Theorem 2.2 is key in reducing the latter continuous evaluation to a discrete one. As a corollary, we now have a unique anonymous affine path attribution method for this class of characteristic functions.

**Corollary 2.3.** If the characteristic function $f$ is the sum of a multilinear function and an additively separable function, then there is a unique anonymous affine path attribution method for $f$.

*Proof.* By Theorem 2.2, any affine path attribution method for $f$ is a random order method. But there is a unique anonymous random order method for $f$, namely the Shapley-Shubik method (Definition 1.5), giving the result. □

**Remark.** We note that the condition that $\gamma$ be monotonically increasing in each variable is necessary (but not sharp) for Theorem 2.2 to hold. Indeed, observe that if any $c_{i,J}$ is negative, then $z$ cannot be a random order method, as in the flow interpretation, this would require the convex combination of non-negative flows to produce a negative flow along some edge. We now exhibit a case where this occurs. Let $n = 3$ and define $\gamma$ by

$$\gamma_1(t) = \begin{cases} 2t & 0 \le t < \frac{1}{3} \\ 1-t & \frac{1}{3} \le t < \frac{2}{3} \\ 2t-1 & \frac{2}{3} \le t \le 1 \end{cases} \text{ and } \gamma_2(t) = \begin{cases} 0 & 0 \le t < \frac{1}{3} \\ 1 & \frac{1}{3} \le t \le 1 \end{cases} \text{ and } \gamma_3(t) = \begin{cases} 0 & 0 \le t < \frac{2}{3} \\ 1 & \frac{2}{3} \le t \le 1. \end{cases}$$

Now, for each $i$, let $\gamma_i^k$ be a sequence of smooth functions $[0,1] \to [0,1]$ which differ from $\gamma_i$ on a set $A_k$ of measure at most $\frac{1}{k}$ and such that $|(\gamma_1^k)'(t)| < 3$ for all $t$. This means that

$$\int_{A_k} (\gamma_1^k)'(t)\gamma_2^k(t)(1-\gamma_3^k(t))dt \to 0$$

because $(\gamma_1^k)'(t)\gamma_2^k(t)(1-\gamma_3^k(t))$ is bounded uniformly in $k$. Then, notice that

$$c_{1,\{2\}}^k = \int_{[0,1]-A_k} \gamma_1'(t)\gamma_2(t)(1-\gamma_3(t))dt + \int_{A_k} (\gamma_1^k)'(t)\gamma_2^k(t)(1-\gamma_3^k(t))dt \to \int_{1/3}^{2/3} -1\,dt = -\frac{1}{3} < 0,$$

hence the affine path attribution method corresponding to $\gamma^k$ for sufficiently large $k$ is not a random order method. On the other hand, if $\gamma$ is not monotonically increasing, it is possible to obtain affine path attribution methods which are random order methods. This occurs exactly when none of the corresponding flows $c_{i,J}$ are negative.

## 3. The Aumann-Shapley-Shubik method

For characteristic functions which are the sum of a multilinear function and an additively separable function, we saw in Theorem 2.2 that every affine path method was a random order method. We will see in this section that the random order method that corresponds to the Aumann-Shapley method is in fact the Shapley-Shubik method. We term the resulting single attribution method the *Aumann-Shapley-Shubik* method. We then characterize this method axiomatically and find the characteristic functions for which it exists.

### 3.1. Existence and axiomatization of Aumann-Shapley-Shubik for multilinear functions.

We explicitly compute the random order method corresponding to the Aumann-Shapley method in Theorem 2.2 to show that it is the Shapley-Shubik method.

**Theorem 3.1.** If $f$ is the sum of a multilinear function and an additively separable function, then the Aumann-Shapley (Definition 1.4) and Shapley-Shubik (Definition 1.5) attribution methods agree for $f$.

*Proof.* By additivity of the attributions with respect to $f(r)$, it is enough to consider $f(r) = r_{i_1} r_{i_2} \cdots r_{i_k}$, since Lemma 2.1 shows that the two methods agree for additively separable functions. Further, if $f(r)$ does not depend on the value of $r_i$, then the attribution to variable $r_i$ is always 0, so in fact it is enough to consider $f(r) = r_1 \cdot \cdots \cdot r_n$.

Our proof will build on the computations done in the proof of Theorem 2.2. The Aumann-Shapley method is the affine path attribution method for $\gamma_i(t) = t$, so it gives attributions

$$z_i^{AS}(r,s) = (s_i - r_i) \sum_{J \subset [n]-\{i\}} c_{i,J} \prod_{j \in J} s_j \prod_{j \in [n]-J-\{i\}} r_j$$

for

$$c_{i,J} = \int_0^1 \gamma_i'(t) \prod_{j \in J} \gamma_j(t) \prod_{j \in [n]-J-\{i\}} (1-\gamma_j(t)) dt = \int_0^1 t^{|J|}(1-t)^{n-1-|J|} dt.$$

On the other hand, each affine path attribution method for $\gamma^\sigma$ assigns to variable $i$ the attribution

$$z_i^\sigma(r,s) = (s_i - r_i) \prod_{\sigma(j)<\sigma(i)} r_j \prod_{\sigma(j)>\sigma(i)} s_j.$$

Therefore, the attribution assigned to variable $i$ under Shapley-Shubik is

$$z_i^{SS} = \frac{1}{n!}(s_i - r_i) \sum_{\sigma \in S_n} \prod_{\sigma(j)<\sigma(i)} r_j \prod_{\sigma(j)>\sigma(i)} s_j = \frac{1}{n!}(s_i - r_i) \sum_{J \subset [n]-\{i\}} |J|!(n-1-|J|)! \prod_{j \in J} s_J \prod_{j \in [n]-\{i\}-J} r_j,$$

so it suffices for us to show that

$$\int_0^1 t^{|J|}(1-t)^{n-1-|J|} dt = \frac{|J|!(n-1-|J|)!}{n!}, \tag{3.1}$$

which follows by taking $i = |J|$ and $j = n-1-|J|$ in Lemma B.1. □

The result of Theorem 3.1 is not surprising, as both the Aumann-Shapley and Shapley-Shubik methods are affine path attribution methods satisfying `Anonymity`. Indeed, it is simply a special case of Corollary 2.3. We may now define the *Aumann-Shapley-Shubik* method for characteristic functions that are the sum of a multilinear and an additively separable function as the method equivalent to both the Aumann-Shapley and Shapley-Shubik methods. Summarizing the conclusions of Theorem 3.1 and Corollary 2.3, we obtain the following axiomatic characterization of the Aumann-Shapley-Shubik method.

**Theorem 3.2.** For characteristic functions $f$ which are the sum of a multilinear function and an additively separable function, the Aumann-Shapley-Shubik method is the unique path attribution method satisfying `Anonymity` and `Affine Scale Invariance`.

**Remark.** Corollary 2.3 and the fact that the Shapley-Shubik method satisfies `Monotonicity` together imply that the Aumann-Shapley-Shubik method satisfies `Monotonicity`. Further, Sprumont and Wang [24] show that the Shapley-Shubik method satisfies a property stronger than `Affine Scale Invariance` called

`Ordinal Invariance`, meaning that the Shapley-Shubik method is invariant to all order-preserving (monotone) reparameterizations of the variables. Corollary 2.3 implies that this carries over to the Aumann-Shapley-Shubik method.

3.2. **When do Aumann-Shapley and Shapley-Shubik agree?** Having now identified the Aumann-Shapley-Shubik method as a uniquely desirable one for characteristic functions which are the sum of a multilinear function and an additively separable function, we now consider when it exists. As we show in the following Theorem 3.3, this will occur only if the characteristic function $f$ takes this form.

**Theorem 3.3.** If the Aumann-Shapley and Shapley-Shubik attribution methods agree for some cost function $f(r)$, then $f$ is the sum of a multilinear function and an additively separable function.

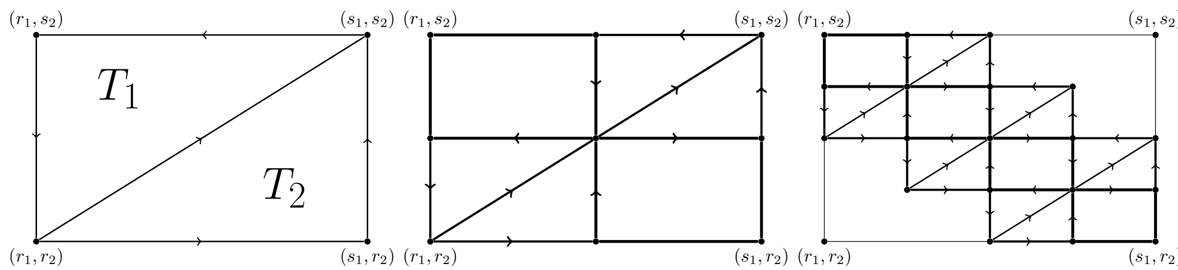
*Proof.* By Lemma B.2, it suffices for us to show that $\partial_{iij} f = 0$ for distinct $i, j$. We first consider the case $n = 2$, in which case we wish to show that $\partial_{12} f$ is constant. Then, for any $r = (r_1, r_2)$ and $s = (s_1, s_2)$ with $r \leq s$, the Aumann-Shapley attribution to the second variable is

$$z_2^{AS}(r, s, f) = \int_0^1 \partial_2 f(\gamma(t))\gamma_2'(t)dt$$

with $\gamma(t) = (1-t)r + ts$. On the other hand, the Shapley-Shubik attribution is

$$z_2^{SS}(r, s, f) = \frac{1}{2}[f(s_1, s_2) - f(s_1, r_2)] + \frac{1}{2}[f(r_1, s_2) - f(r_1, r_2)].$$

Now, subdivide the rectangle $R$ with vertices at $(r_1, r_2)$, $(r_1, s_2)$, $(s_1, r_2)$, and $(s_1, s_2)$ into the triangular regions $T_1$ lying above the path of $\gamma$ and $T_2$ lying below the path of $\gamma$ as shown in Figure 1(a) below.




(a) Regions $T_1$ and $T_2$ (b) Removing two pairs of triangles. (c) Rectangles to apply (3.3) on.

FIGURE 1. Steps in the proof of Theorem 3.3

Then, by Stokes' Theorem, we have

$$\int_{T_1} \partial_{12} f(x_1, x_2)dx_1 dx_2 = \int_{\partial T_1} \partial_2 f(x_1, x_2)dx_2 = \int_0^1 \partial_2 f(\gamma(t))\gamma_2'(t)dt - [f(r_1, s_2) - f(r_1, r_2)]$$

and

$$\int_{T_2} \partial_{12} f(x_1, x_2)dx_1 dx_2 = \int_{\partial T_2} \partial_2 f(x_1, x_2)dx_2 = [f(s_1, s_2) - f(s_1, r_2)] - \int_0^1 \partial_2 f(\gamma(t))\gamma_2'(t)dt.$$

Because $z_2^{SS}(r, s, f) = z_2^{AS}(r, s, f)$ by assumption, subtracting the two previous equations and applying our previous computations gives that

$$\int_{T_1} \partial_{12} f(z_1, z_2)dz_1 dz_2 = \int_{T_2} \partial_{12} f(z_1, z_2)dz_1 dz_2 \tag{3.2}$$

for any choice of $r, s$. In particular, applying (3.2) for the pairs $(r, s)$, $\left(r, \frac{r+s}{2}\right)$, and $\left(\frac{r+s}{2}, s\right)$ and subtracting the result of the latter two from the first, we obtain

$$\int_{\left[r_1, \frac{r_1+s_1}{2}\right] \times \left[\frac{r_2+s_2}{2}, s_2\right]} \partial_{12} f = \int_{\left[\frac{r_1+s_1}{2}, s_1\right] \times \left[r_2, \frac{r_2+s_2}{2}\right]} \partial_{12} f \tag{3.3}$$

for all $r, s$. The results of this process are shown in Figure 1(b). Now, for any $x = (x_1, x_2)$, set $x' = (x_1, -x_2)$. Applying (3.3) to the pairs $(r, r+2x), (r+x', r+x'+2x), \ldots, (r+nx', r+nx'+2x)$, we find that for any $n$ we have

$$\int_{[r_1,r_1+x_1]\times[r_2+x_2,r_2+2x_2]} \partial_{12} f = \int_{[r_1+(n+1)x_1,r_1+(n+2)x_1]\times[r_2-nx_2,r_2-(n-1)x_2]} \partial_{12} f. \tag{3.4}$$

This process is shown in Figure 1(c).

Suppose now for the sake of contradiction that $\partial_{12} f$ were not constant. Then, there must exist some $r < s$ such that $\partial_{12} f(r) \neq \partial_{12} f(s)$. Suppose without loss of generality that $\partial_{12} f(r) > \partial_{12} f(s)$. Because $\partial_{12} f$ is continuous, we may find open neighborhoods $U$ of $r$ and $V$ of $s$ such that $\partial_{12} f(x) > \partial_{12} f(y)$ for $x \in U, y \in V$. Now, choose $x = (x_1, x_2)$ and $n$ so that $[r_1, r_1 + x_1] \times [r_2 + x_2, r_2 + 2x_2] \subset U$ and that $[r_1 + (n+1)x_1, r_1 + (n+2)x_1] \times [r_2 - nx_2, r_2 - (n-1)x_2] \subset V$, in which case (3.4) provides a contradiction. Therefore, $\partial_{12} f$ is constant, which completes the proof in the case $n = 2$.

For the general case, choose any two variables $r_i$ and $r_j$. Restricting to attributions between points with all other variables held fixed, the $n = 2$ case tells us that $\partial_{ij} f$ is independent of $r_i$ and $r_j$, which means exactly that $\partial_{iij} f = 0$ and $\partial_{ijj} f = 0$. This holds for all $i, j$, so $f$ takes the desired form. $\square$

**Corollary 3.4.** There is a unique affine path attribution method satisfying `Anonymity` and `Affine Scale Invariance` for a characteristic function $f$ if and only if $f$ is the sum of a multilinear function and an additively separable function. In this case, the method corresponds to both the Aumann-Shapley and Shapley-Shubik methods.

*Proof.* Immediate from Corollary 2.3 and Theorem 3.3. $\square$

3.3. **Computing Aumann-Shapley-Shubik.** In this subsection, we discuss the efficient computation of the Aumann-Shapley-Shubik method for multilinear functions (We ignore additively separable functions because the attribution assigned to a variable is simply the change in the function in which it appears). As discussed at the end of Subsection 2.3, if $f$ is a multilinear function, then this method is computable in finite time because it coincides with the Shapley-Shubik method; recall that in this case we need to average over the marginal impact of changing a variable over all possible orderings, which is finitely computable. However, the Shapley-Shubik method does not always have an efficient (polynomial time) algorithm to compute it; see for instance the hardness results in [7, 13].

Now, for $f(r) = r_1 \cdot \cdots \cdot r_n$, the most basic example of a multilinear function, the Aumann-Shapley-Shubik attributions $z_i(r, s, f)$ are computable in finite time, as to compute the Shapley-Shubik attributions in this case it suffices to evaluate $f$ a finite number of times. In principle, this may involve $\Theta(2^n)$ evaluations, one for each of the vertices of $[r, s]$. However, Theorem 3.5 below implies that in this case we may compute attributions in time quadratic in the number of variables. If we instead consider general multilinear functions, iterating the algorithm of Theorem 3.5 in Corollary 3.6 yields runtime quadratic in the number of variables and linear in the number of non-zero monomials in the characteristic function. These two results together ensure that our attribution theory is not impractical for computational reasons.

**Theorem 3.5.** Let $f(r) = r_1 \cdots r_n$. Then, for any $r, s$ and each $i$, the Aumann-Shapley-Shubik attribution $z_i(r, s, f)$ is computable in $O(n^2)$ time and $O(n)$ memory.

*Proof.* From the calculations in the proof of Theorem 3.1, the attributions are given by

$$\begin{aligned} z_i(r, s, f) &= \frac{1}{n!}(s_i - r_i) \sum_{K \subset [n]-\{i\}} |K|!(n-1-|K|)! s_K r_{[n]-\{i\}-K} \\ &= \frac{1}{n!}(s_i - r_i) \sum_{k=0}^{n-1} k!(n-1-k)! \sum_{\substack{K \subset [n]-\{i\} \\ |K|=k}} s_K r_{[n]-\{i\}-K}, \end{aligned}$$

so it suffices to compute this value. The computation is invariant under relabeling of coordinates, so we may assume for convenience of notation that $i = n$. In this case, we have

$$z_n(r, s, f) = \frac{1}{n!}(s_n - r_n) \sum_{k=0}^{n-1} k!(n-1-k)! \sum_{\substack{K \subset [n-1] \\ |K|=k}} s_K r_{[n-1]-K}.$$

Our approach is to compute the sums

$$X_{k,m} := \sum_{\substack{K \subset [m] \\ |K|=k}} s_K r_{[m]-K}$$

for $m \leq n-1$ and $0 \leq k \leq m$ using dynamic programming. Computing $z_i(r,s,f)$ then requires only a simple summation. Algorithm 1 formalizes this idea.

**Algorithm 1** Computing the Aumann-Shapley-Shubik attribution $z_n(r,s,f)$.

$X_{0,0} \leftarrow 1$
**for** $m = 1$ to $n-1$ **do**
  $X_{0,m} \leftarrow r_m \cdot X_{0,m-1}$
  **for** $k = 1$ to $m-1$ **do**
    $X_{k,m} \leftarrow s_m \cdot X_{k-1,m-1} + r_m \cdot X_{k,m-1}$
  **end for**
  $X_{m,m} \leftarrow s_m \cdot X_{m-1,m-1}$
**end for**
**return** $\frac{1}{n!}(s_n - r_n)\sum_{k=0}^{n-1} k!(n-1-k)! \cdot X_{k,n-1}$

The correctness of Algorithm 1 follows from the evident recursion

$$X_{k,m} = \begin{cases} r_m \cdot X_{0,m-1} & k=0 \\ s_m \cdot X_{k-1,m-1} + r_m \cdot X_{k,m-1} & 1 \leq k \leq m-1 \\ s_m \cdot X_{m-1,m-1} & k=m \end{cases}$$

and the expression for $z_i(r,s,f)$ obtained at the beginning of the proof. There are $O(n^2)$ iterations of the loop, each taking $O(1)$ time to update $X_{k,m}$, giving a total runtime of $O(n^2)$. Further, at each step, only the values of $X_{k,m}$ for $0 \leq k \leq m$ and $X_{k,m-1}$ for $0 \leq k \leq m-1$ are required, so storing only these yields a memory requirement of $O(n)$. □

**Corollary 3.6.** Let $f$ be a multilinear characteristic function with $N$ non-zero monomial terms. Then, the Aumann-Shapley-Shubik attribution $z_i(r,s,f)$ is computable in $O(n^2 \cdot N)$ time and $O(n)$ memory.

*Proof.* By `Additivity` and `Dummy`, we may simply run the algorithm of Theorem 3.5 $N$ times, once for each non-zero monomial in $f$, and sum the resulting contributions. This trivially gives the desired runtime and memory costs. □

## 4. Applications

In this section we list several examples of applications for which attribution can yield practically relevant insights. In each of these applications, the characteristic function is known, deterministic, and multilinear, and the values of the variables at the initial and final points are known. Therefore, our approach in this paper is applicable. We begin with a few examples motivated by the Internet.

**Example 4.1** (Pay-per-click advertising [8])**.** The characteristic function is the spend $s$ of an advertiser, which can be expressed as the product of the number of clicks ($c$) that an advertiser's advertisement received and the average cost per click ($p$). The final values $s$, $c$, and $p$ may be statistics from a certain week, and the initial values may be statistics from the preceding week. The problem then is to identify to what extent the advertiser's change in spend is due to a change in the number of clicks versus a change in the cost per click.

A more granular spend model applicable in a specific form of pay-per-click advertising called sponsored search advertising is

$$f_{\text{spend}} = q \cdot b \cdot \sum_i p_i \cdot \text{CTR}_i \cdot \text{CPC}_i.$$

Here, $q$ is the number of ad-views that the advertiser is eligible for, $b$ is the probability that the ads have sufficient budget to show, $p_i$ is the probability that an ad appears in the $i^{\text{th}}$ auction position, and $(\mathrm{CTR_i}, \mathrm{CPC_i})$

are the click through rate and the cost per click for the $i^{\text{th}}$ auction position. (Recall that all major search engines place some ads based on the results of an auction.)

**Example 4.2** (e-Commerce website analysis)**.** Consider an online retailer's website. We can model the website as a directed acyclic graph with a single sink $t$, which is the page displayed on a successful transaction (see Immorlica et al. [12] and Archak et al. [2] for similar models). For every page, let $s_j$ denote the number of times that a surfer starts on page $j$. For every hyperlink directed from page $i$ to page $j$, let $p_{ij}$ denote the probability on average that a surfer follows this link given that she is at page $i$. The expected number of successful transactions is

$$\sum_{i\in V} s_i \sum_{P \text{ a path from } i \text{ to } t} \prod_{(r,s)\in P} p_{rs},$$

which is multilinear. The attributions to the variables $\{s_j\}$ and $\{p_{ij}\}$ may then yield insight into changes in traffic patterns that impact sales.

We now have an example motivated by investment.

**Example 4.3** (Performance attribution)**.** The performance of a portfolio can be expressed as the following sum of products

$$\sum_{i\in S} r_i \cdot w_i,$$

where $r_i$ is the return within an asset class $i$, and $w_i$ the amount invested within this asset class. Performance attribution [6] attempts to explain why the performance of a portfolio (the final variables) deviates from the performance of a benchmark portfolio (the initial variables). In particular, it asks whether it is due to a difference in the allocation of investments across asset classes (the attributions to the $w_i$'s) or to the selection of assets within an asset classes (the attributions to the $r_i$'s).

The standard way of doing performance attribution involves considering an active allocation term $r_i^1 \cdot (w_i^2 - w_i^1)$, a security selection term $w_i^1 \cdot (r_i^2 - r_i^1)$, and a slack term $(r_i^2 - r_i^1) \cdot (w_i^2 - w_i^1)$ for each asset class; the latter term is necessary for completeness, but does not yield any insight. In contrast, our approach yields completeness automatically.

Finally, here is an example from performance analysis of basketball statistics.

**Example 4.4** (Sports analysis)**.** Suppose the coaching staff of a basketball team wants insight into the change in offensive performance of the team from last year (the initial version of the variables) to this year (the final version of the variables); such studies are currently done in other frameworks as in [20] or [19]. Then, letting $n_i$, $m_i$, $a_i$, and $p_i$ be the number of games per season, the number of minutes per game, the number of attempts per minute, and the field goal percentage of each player, the total number of points scored is

$$f_{\text{points}} = \sum_i n_i \cdot m_i \cdot a_i \cdot \frac{p_i}{100}.$$

Using attributions for $f_{\text{points}}$ in combination with other information can help the coaches understand and refine the performance of the team.

**Remark.** Let us reiterate the benefit of a method satisfying `Affine Scale Invariance` in light of the above examples. For many attribution problems, the units in which variables are measured are a matter of convention and are not canonical in any sense. For instance, in pay-per-click advertising (Example 4.1), the cost of advertising may be measured as the cost per thousand impressions or the cost per million impressions (see [8]), and, in basketball statistics, field goal accuracy is popularly expressed as a percentage between 0 and 100 rather than an accuracy rate between 0 and 1.

In these examples, it critical that a different scaling of the units does not change the attribution. Specifying variables can be even more difficult than this, however. Consider a characteristic function which depends on a dimensionless physical quantity such as the Reynolds number of a chaotic fluid or the Prandtl number of a material. Such quantities lack natural units or even canonical reference points; for them, we would like the attribution to be invariant not only under rescaling of units but also changes of the zero points of these units. Such changes are exactly affine transformations, leading us to the `Affine Scale Invariance` axiom.

## References


[1] Apostol, T. M. *Calculus, Vol. 2: Multi-Variable Calculus and Linear Algebra with Applications*. Wiley, 1969.
[2] Archak, N., Mirrokni, V. S., and Muthukrishnan, S. Mining advertiser-specific user behavior using adfactors. In *Proceedings of the 19th International Conference on World Wide Web* (New York, NY, USA, 2010), WWW '10, ACM, pp. 31–40.
[3] Aumann, R. J., and Shapley, L. S. *Values of non-atomic games*. Princeton University Press, 1974.
[4] Billera, L. J., and Heath, D. C. Allocation of shared costs: A set of axioms yielding a unique procedure. *Math. Oper. Res. 7*, 1 (1982), pp. 32–39.
[5] Bollen, K. A. *Structural Equations with Latent Variables*, $1^{st}$ ed. Wiley-Interscience, April 1989.
[6] Brinson, G. P., Singer, B. D., and Beebower, G. L. Determinants of portfolio performance II: An update. *Financial Analysts Journal 47*, 3 (1991), pp. 40–48.
[7] Deng, X., and Papadimitriou, C. H. On the complexity of cooperative solution concepts. *Math. Oper. Res. 19* (May 1994), 257–266.
[8] Fain, D. C., and Pedersen, J. O. Sponsored search: A brief history. *Bul. Am. Soc. Info. Sci. Tech. 32*, 2 (2006), 12–13.
[9] Friedman, E. Paths and consistency in additive cost sharing. *Int. J. Game Theory 32*, 4 (August 2004), 501–518.
[10] Friedman, E., and Moulin, H. Three methods to share joint costs or surplus. *J. Econ. Theory 87*, 2 (August 1999), 275–312.
[11] Haimanko, O. Partially symmetric values. *Math. Oper. Res. 25* (November 2000), 573–590.
[12] Immorlica, N., Jain, K., and Mahdian, M. Game-theoretic aspects of designing hyperlink structures. In *Internet and Network Economics (WINE)* (Patras, Greece, December 2006), vol. 4286, Springer LNCS, pp. 150–161.
[13] Matsui, Y., and Matsui, T. NP-completeness for calculating power indices of weighted majority games. *Theor. Comput. Sci. 263* (July 2001), 306–310.
[14] Mirman, L. J., and Tauman, Y. Demand compatible equitable cost sharing prices. *Math. Oper. Res. 7*, 1 (1982), 40–56.
[15] Moulin, H. Axiomatic cost and surplus sharing. In *Handbook of Social Choice and Welfare*, K. J. Arrow, A. K. Sen, and K. Suzumura, Eds., vol. 1 of *Handbook of Social Choice and Welfare*. Elsevier, April 2002, ch. 6, pp. 289–357.
[16] Moulin, H., and Sprumont, Y. Responsibility and cross-subsidization in cost sharing. Cahiers de recherche 19-2002, Centre interuniversitaire de recherche en l'économie quantitative, CIREQ, 2002.
[17] Redekop, J. Increasing marginal cost and the monotonicity of aumann-shapley pricing. mimeo, University of Waterloo, Ontario, Canada, 1996.
[18] Roth, A. E. *Probabilistic values for games*. Cambridge University Press, 1988.
[19] Sampaio, J., Ibánez, S., Lorenzo, A., and Gómez, M. Discriminative game-related statistics between basketball starters and nonstarters when related to team quality and game outcome. *Perceptual and Motor Skills 103* (October 2006), 486–494.
[20] Sampaio, J., and Janeira, M. Statistical analyses of basketball team performance: understanding teams wins and losses according to a different index of ball possessions. *International Journal of Performance Analysis in Sport 3* (1 April 2003), 40–49(10).
[21] Shapley, L. S. A value for $n$-person games. *Contributions to the theory of games 2* (1953), 307–317.
[22] Shapley, L. S., and Shubik, M. A method for evaluating the distribution of power in a committee system. *Am. Polit. Sci. Rev. 48*, 3 (1954), 787–792.
[23] Warner, F. W. *Foundations of Differentiable Manifolds and Lie Groups*. Springer, 1983.
[24] Y. Sprumont, Y. W. Ordinal additive cost-sharing methods must be random order values. mimeo, Université de Montréal, 1996.


## Appendix A. A review of Stokes' theorem

In this appendix, we give a brief intuitive introduction to Stokes' theorem as it relates to our paper for the unfamiliar reader. To minimize technical difficulties, we restrict ourselves to the case of dimension two, where Stoke's Theorem coincides with Green's Theorem, and suppress technical assumptions. First, we state a basic version of the theorem.

**Theorem A.1** (Stokes' Theorem)**.** Let $A$ be the region enclosed by a smooth closed curve in the plane. Let $f$ be a differentiable function defined on an open neighborhood of $A$, and let $\partial A$ be the (oriented) boundary of $A$. Then, we have

$$\int_{\partial A} f dx_2 = \int_A \partial_1 f dx_1 dx_2. \tag{A.1}$$

Let us explain intuitively the meaning of Theorem A.1. It relates the path integral of the 1-dimensional differential form $f dx_1$ along the boundary $\partial A$ of $A$ to the double integral of its exterior derivative $d(f dx_1) = \partial_2 f dx_1 dx_2$ on the interior of $A$. We may visualize this in Figure 2 below.

It may be instructive to consider an analogy between Stokes' theorem and the fundamental theorem of calculus (which is actually Stokes' theorem in dimension 1). For a differentiable function $F$, the fundamental theorem of calculus relates the integral of $F'(x)$ along an interval to the difference in values of $F$ on the

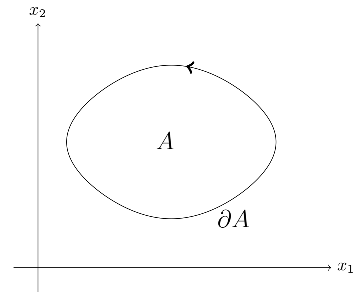



FIGURE 2. A region $A$ and its boundary $\partial A$ in Stokes' theorem.

endpoints of this interval. That is, it states that

$$F(b) - F(a) = \int_a^b F'(x)dx.$$

Stokes' theorem generalizes the fundamental theorem of calculus in the sense that it replaces the concept of an interval with a simple region, and the endpoints of the interval (which form its boundary) with the closed curve that forms the boundary of the region. Its proof is also ultimately an application of the fundamental theorem of calculus. We refer the interested reader to Chapter 11 of [1] or to [23] for more detailed expositions of Stokes' theorem, which also appears in various engineering applications such as electrostatics and fluid dynamics.

In this paper, Stokes' theorem is particularly convenient because it allows manipulation of line integrals of 1-dimensional differential forms. We see that the attributions given by path attribution methods take exactly this form for differential forms involving the characteristic function. Applying Stokes' theorem now yields conditions on the area integral of a mixed partial which we use as a starting point for further considerations.

In this paper, we require a slightly more general version of Stokes' theorem, stated below, which is proved by repeated application of Theorem A.1.

**Theorem A.2** (Stokes' Theorem for non-simple curves)**.** Let $\gamma : S^1 \to \mathbb{R}^2$ be a smooth closed curve (which may self-intersect). Let $\gamma$ divide $\mathbb{R}^2 - \gamma(S^1)$ into open regions $A_i$ such that the winding number of $\gamma$ about each point in $A_i$ is $i$. Then, for any differentiable function $f$ on $\mathbb{R}^2$, we have

$$\int_\gamma f dx_2 = \sum_{i \in \mathbb{Z}} i \int_{A_i} \partial_1 f dx_1 dx_2.$$

## APPENDIX B. TECHNICAL LEMMAS

In this appendix we state and prove some technical results in mathematics which are used in our proofs.

**Lemma B.1.** For non-negative integers $i, j$, we have

$$\int_0^1 x^i (1-x)^j dx = \frac{1}{(i+j+1)\binom{i+j}{i}} dx.$$

*Proof.* We induct on $i$. For $i = 0$, the result is clear. Now, suppose that the result holds for some $i-1$. In this case, integration by parts gives that

$$\begin{aligned}
\int_0^1 x^i (1-x)^j dx &= \left[\frac{1}{j+1} x^i (1-x)^{j+1}\right]_0^1 + \int_0^1 i x^{i-1} \frac{1}{j+1} (1-x)^{j+1} dx \\
&= \frac{i}{j+1} \int_0^1 x^{i-1} (1-x)^{j+1} dx \\
&= \frac{i}{j+1} \frac{1}{(i+j+1)\binom{i+j}{i-1}} \\
&= \frac{1}{(i+j+1)\binom{i+j}{i}},
\end{aligned}$$

which completes the proof. □

**Lemma B.2.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is the sum of a multilinear function and an additively separable function if and only if $\partial_{iij} f = 0$ for all $i \neq j$.

*Proof.* The "if" direction is evident, so we focus on the "only if" direction. We proceed by induction on $n$, with the base case $n = 1$ trivial. Now, if $n > 1$, we may write $\partial_{11} f = g_1$ as a function of $q_1$ only, hence we see that

$$\partial_1 f(r) = \int_0^{r_1} g_1(t)dt + h(r)$$

and

$$f(r) = \int_0^{r_1} \int_0^{t_2} g_1(t_1)dt_1dt_2 + r_1h(r) + p(r),$$

where $\partial_1 h = \partial_1 p = 0$. It remains to show that $h$ is multilinear and that $p$ is the sum of a multilinear function and an additively separable function. Now, for any distinct $i, j \neq 1$, we have that

$$0 = \partial_{iij} f = r_1 \partial_{iij} h + \partial_{iij} p,$$

so taking $r_1 = 0$ shows that $\partial_{iij} p = 0$. Hence $p$ is the sum of a multilinear function and an additively separable function by the inductive hypothesis. Now, notice that for $i \neq 1$, we have

$$0 = \partial_{ii1} f = \partial_{ii} h,$$

so $h$ is multilinear. This completes the induction. □

**Remark.** The condition in Lemma B.2 is a mixture of the conditions for $f$ to be multilinear ($\partial_{ii} f = 0$) and to be additively separable ($\partial_{ij} = 0$).

Churchill College, Cambridge, CB3 0DS, United Kingdom
*E-mail address*: yi.sun@post.harvard.edu

Google, 1600 Amphitheatre Parkway, Mountain View, CA 94043
*E-mail address*: mukunds@google.com