# A new proof of Williamson's representation of multiply monotone functions

Nabil Kazi-Tani, Didier Rullière

▶ To cite this version:

Nabil Kazi-Tani, Didier Rullière. A new proof of Williamson's representation of multiply monotone functions. 2020. hal-02984743

HAL Id: hal-02984743
https://hal.science/hal-02984743

Preprint submitted on 31 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A new proof of Williamson's representation of multiply monotone functions.*

Nabil Kazi-Tani †  Didier Rullière ‡

31/10/2020

**Abstract**

This paper provides an alternative proof of the characterization of multiply monotone functions as integrals of simple polynomial-type applications with respect to a probability measure. This constitutes an analogue of the Bernstein-Widder representation of completely monotone functions as Laplace transforms. The proof given here relies on the abstract representation result of Choquet [8] rather than the analytic derivation originally given by Williamson [24]. To this end, we identify the extreme points in the convex set of multiply monotone functions. Our result thus gives a geometric perspective to Williamson's representation.

**Keywords:** Multiply monotone functions, Choquet theorem, Extreme points

**AMS Subject Classification(2020):** 26A48, 26A51, 44A05, 46A55.

## 1 Introduction

Completely monotone functions (c.m. for simplicity), whose signs of successive derivatives alternate, are a standard object of study in analysis and probability. Bernstein proved in [4] that conveniently normalized completely monotone functions on $\mathbb{R}^+$ are Laplace transforms of probability measures on $\mathbb{R}^+$: this is known as the Bernstein-Widder theorem (see Chapter 4, Theorem 12a in [23]). It is natural to weaken the definition of c.m. functions and ask that the derivatives' signs alternate only up to some fixed order $d \geq 1$. Williamson [24] showed that these functions, known as multiply monotone, or $d$-monotone, can also be expressed as expectations of some simple functions with respect to a probability measure (see (3.1)), the simple functions being scale mixtures of Beta distributions. The goal of this paper is to provide an alternative proof for the representation (3.1), that sheds some new light on the geometry of the set of $d$-monotone functions, in particular on the extreme points of this convex set.

*This manuscript version: 31/10/2020 18:35:20.

†Univ Lyon, Université de Lyon 1, Laboratoire de Sciences Actuarielle et Financière, ISFA, 50 Avenue Tony Garnier, F-69007 Lyon, France. nabil.kazi-tani@univ-lyon1.fr.

‡Mines Saint-Etienne, UMR CNRS 6158, LIMOS, 29, rue Ponchardier, 42023 Saint-etienne, France, didier.rulliere@emse.fr

Many known proofs of the Bernstein-Widder representation of c.m. functions use the fact that if $f$ is a c.m. function, then $x^k f^{(k)}(x)$ converges to 0 when $x$ goes to infinity ([4, 23, 15, 7]). In the same vein, Williamson [24] shows that for a multiply monotone function $f$, $x^k f^{(k)}(x)$ also converges to 0 and performs a Taylor expansion to identify the density of the representing measure in (3.1). These proofs are based on the fact that the definition of c.m. or multiply monotone functions requires that these functions are differentiable enough. Note that Bernstein [3] already introduces a definition only involving successive *finite* differences (see Definition 2.1 in this paper) and that Widder ([23], Chapter 4, Theorem 7) shows that the definitions involving derivatives and finite differences are equivalent. The alternative proof presented here has the advantage of not making use of differentiability, since it is based on abstract representation results of Choquet. The Bernstein-Widder representation result is generalized in [8] (Chapter 7, Section 43) to c.m. functions which are defined on an ordered semigroup (an ordered set with an associative operation), by identifying the extreme points of such a set.

In his Bourbaki lecture notes [9], Choquet provides several examples of convex sets whose elements can be represented by simpler preferred points (extreme points): among the examples, one finds c.m. functions on $[0, +\infty)$, or alternating capacities on a compact set. In the first case, the preferred elements are $(\exp(-tx), t \geq 0)$ and in the second case it is elementary capacities, which are $\{0, 1\}$-valued. The proof given in [9] uses abstract representation results, extending the Krein-Milman theorem representing certain convex and compact subsets of topological vector spaces as the convex hull of their extreme points. The proof given here follows this line of reasoning. We first identify the extreme points in the set of $d$-monotone functions, and then apply the representation theorem of Choquet [8, 21], which is recalled in Appendix for completeness.

Williamson's result, which is the counterpart for $d$-monotone functions of the Bernstein-Widder representation, has several important applications in probability, statistics and approximation theory. In [19], it is proved that a function $\varphi$ generates a valid $d$-dimensional Archimedean copula if and only if $\varphi$ is $d$-monotone. This allowed to provide new classes of multivariate distributions [18] or new nonparametric statistical estimation procedures for Archimedean copulas [13]. Building $d$-monotone functions by reassembling functions defined on different intervals also allows to modify a copula so that it exhibits a specific tail dependence behavior [10]. Independently of copula theory, there is an interest for the class of $d$-monotone functions in statistics. Gao [12] has computed entropy estimates for $d$-monotone functions, which is used to provide rates of convergence of density estimators, where the density is constrained to be $d$-monotone. Relying on Williamson's representation, the authors of [2] obtain a limit distribution for the maximum likelihood estimator of a $d$-monotone density, where the limit distribution is characterized as a marginal of the lowest $(2d)$-convex process dominating a given Brownian stochastic integral, a process $Z$ being $(2d)$-convex if $(-1)^d Z^{(2d-2)}$ exists and is convex. Multiply monotone functions also have close links with radial functions [14] and approximation theory [6, 5]: we refer the reader to these papers and the references therein for more details.

In the next section, we identify the extreme points of the set of $d$-monotone functions and in Section 3, we prove Williamson's representation.

## 2 Extreme points in the set of $d$-monotone functions

Let $X$ be the space of real valued functions on $\mathbb{R}^+$. For $a > 0$, the difference operator $D_a : X \to X$ is defined by

$$D_a f(t) = f(t+a) - f(t), \quad t \in \mathbb{R}^+.$$

For positive integers $a_1, \dots, a_n$, the product $\prod_{j=1}^{n} D_{a_j}$ denotes the sequential application of the operators $D_{a_1}, \dots, D_{a_n}$.

**Definition 2.1** *Let $d \geq 1$ be an integer. A function $f \in X$ is called $d$-monotone on $(0, +\infty)$ if*

$$(-1)^n \left( \prod_{j=1}^{n} D_{a_j} \right) f(t) \geq 0, \quad \forall t \in (0, +\infty), \tag{2.1}$$

*for all $a_j > 0$ and all $n = 0, \dots, d$, where by convention $n = 0$ corresponds to $f \geq 0$.*

If $d = 1$, the previous definition says that $f$ is 1-monotone on $(0, +\infty)$ if it is nonnegative and nonincreasing there, while 2-monotone functions are nonnegative, nonincreasing and convex.

**Remark 2.1** *Note that unlike the definition given in the previous literature ([24, 2, 12, 19]), we do not require here that $d$-monotone functions are differentiable. We will nonetheless prove, via Theorem 3.1, that for $d \geq 2$, $d$-monotone functions are differentiable up to the order $d-2$. Note that this can also be obtained directly, independently of* (3.1)*, by a reformulation of Theorem 4 in [23] (Chapter 4).*

For $d \geq 2$, let $K_d$ denote the space of functions defined on $[0, +\infty)$, which are $d$-monotone on $(0, +\infty)$ and let $K_d^0 \subset K_d$ be the functions in $K_d$ satisfying

$$f(0) \in [0, 1] \quad \text{and} \quad \lim_{x \to +\infty} f(x) = 0. \tag{2.2}$$

Define $K_1$ as the set of nonnegative, nonincreasing functions on $[0, +\infty)$ which are right-continuous, and $K_1^0$ the functions in $K_1$ satisfying (2.2).

$K_d$ is a convex cone, and $K_d^0$ is a convex set, for which we want to identify the extreme points. To do so, we just start from the extreme points in the set $K_1$, which are well known, and we show that we can retrieve the extreme points in $K_d$ by successive integrations. This is the purpose of the next two lemmas.

Since any $f \in K_d^0$ is convex, it is in particular continuous and the set of points where $f$ is not differentiable has Lebesgue measure zero. Moreover, also because of the convexity of $f$, we can define its right-hand derivative $f'_+(t)$ for every $t \in (0, +\infty)$, which of course coincides with the derivative $f'(t)$ where $f$ is differentiable.

**Lemma 2.1** *Let $d \geq 2$. If $e \in K_d^0$ is an extreme point, then $-e'_+$ is an extreme point in $K_{d-1}$.*

**Proof.**
Step 1: We first show that if $e \in K_d$, then $-e'_+ \in K_{d-1}$. For any $a > 0$, $-D_a e \in K_{d-1}$: indeed, in (2.1), we can take $n = d$ and $a_1 = a$ and write

$$(-1)^{d-1} \left( \prod_{j=1}^{d-1} D_{a_j} \right) (-D_a f) \geq 0.$$

We can use (2.1) this way for any $n = 1, \ldots, d$ and get that $-D_a e \in K_{d-1}$. This implies that $-e'_+ \in K_{d-1}$ since

$$e'_+(t) = \lim_{a \to 0^+} \frac{e(t+a) - e(t)}{a} = \lim_{a \to 0^+} \frac{D_a e(t)}{a}$$

and for $b_1, \ldots, b_n$ positive,

$$(-1)^n \left( \prod_{j=1}^{n} D_{b_j} \right) (-e'_+(t)) = (-1)^{n+1} \lim_{a \to 0^+} \frac{1}{a} \left( \prod_{j=1}^{n} D_{b_j} \right) D_a e(t) \geq 0.$$

Step 2: Let us prove that $e \in K_d^0$ is an absolutely continuous function. Since $e$ is convex, it is also continuous and by monotony, $e$ is of finite variation. By convexity, $e$ has left and right derivatives for every $t \in (0, +\infty)$: Lemma 7.25 in [22] (page 153) implies that $e$ satisfies Luzin's (N) property, i.e. the image by $e$ of every Lebesgue negligible subset of $(0, +\infty)$ is null. By the Banach-Zarecki theorem [20, 11], we can conclude that $e$ is absolutely continuous. In particular

$$e(t) = e(0) - \int_0^t e'_+(u) du.$$

Step 3: Assume now that $-e'_+ = \lambda g + (1 - \lambda) h$ with $g, h \in K_{d-1}$ and $\lambda \in (0, 1)$. By absolute continuity of $e$, we have

$$\begin{aligned} e(t) &= \lambda \left( e(0) - \int_0^t g(v) dv \right) + (1 - \lambda) \left( e(0) - \int_0^t h(v) dv \right) \\ &=: \lambda \left( \widetilde{g}(t) \right) + (1 - \lambda) \left( \widetilde{h}(t) \right). \end{aligned}$$

It can be checked directly that $\widetilde{h}$ and $\widetilde{g}$ belong to $K_d^0$. By extremality of $e$, we get $e(t) = \widetilde{g}(t) = \widetilde{h}(t)$, for $t \in (0, +\infty)$, and by differentiation, which is authorized since $g$ and $h$ are continuous, we get $-e'_+ = g = h$.

□

**Lemma 2.2** *Let $d \geq 1$, $f_d$ be an extreme point in $K_d^0$ and define*

$$f_{d+1}(t) = 1 - \int_0^t f_d(v) dv.$$

*Then $f_{d+1}$ is an extreme point in $K_{d+1}^0$.*

**Proof.** Assume that $f_{d+1} = \lambda g + (1-\lambda)h$, with $g, h \in K^0_{d+1}$ and $\lambda \in (0,1)$. By Lemma 2.1, $-g'_+ \in K_d$ and $-h'_+ \in K_d$. By continuity of $f_d$, we have

$$f_d(t) = \lambda(-g'_+(t)) + (1-\lambda)(-h'_+(t)).$$

From the inequalities $0 \leq f_d(0) \leq 1$ and the fact that $-g'_+(0)$ and $h'_+(0)$ are both nonnegative, we get $-g'_+(0) \leq 1$ and $h'_+(0) \leq 1$. Since $\lim_{t\to\infty} f_d(t) = 0$ and that both functions $-g'_+$ and $-h'_+$ are nonnegative, we get that $-h'_+$ and $-g'_+$ are in $K^0_d$. $f_d$ being an extreme point in $K^0_d$, we have $-g'_+ = -h'_+ = f_d$, which in turn implies that $f_{d+1} = g = h$, by definition of $f_{d+1}$ and by absolute continuity of $g$ and $h$ (see step 2 of the proof of Lemma 2.1).

□

For $u > 0$, we define the following sequence of functions

$$e^u_1(t) = \begin{cases} 1 & \text{if } 0 \leq t < \frac{1}{u} \\ 0 & \text{otherwise,} \end{cases} \tag{2.3}$$

$$e^u_{d+1}(t) = \begin{cases} 1 - u\,d\int_0^t e_d(v)dv & \text{if } 0 \leq t < \frac{1}{u} \\ 0 & \text{otherwise,} \quad d \geq 2, \end{cases} \tag{2.4}$$

for $t \in [0, +\infty)$.
We have of course the explicit expression

$$e_d(t) = \begin{cases} (1-ut)^{d-1} & \text{if } 0 \leq t < \frac{1}{u} \\ 0 & \text{if } \ t \geq \frac{1}{u}. \end{cases} \tag{2.5}$$

These functions are scale mixtures of Beta distributions. The result below says that these are the extreme points in the set of $d$-monotone functions.

**Theorem 2.1** *The extreme points of the set $K^0_d$ are given by*

$$Ext(K_d) = \{e^u_d,\ u > 0\}.$$

**Proof.** The functions in $K^0_1$, which are right-continuous by definition, are distribution functions of probability measures on $\mathbb{R}^+$, whose extreme points are unit masses concentrated on a single point (see Theorem 15.9 in [1]). Said differently, we have $Ext(K^0_1) = \{e^u_1,\ u > 0\}$. Using Lemma 2.2 by induction, we get that

$$\{e^u_d,\ u > 0\} \subset Ext(K^0_d),$$

since the multiplication by a positive constant do not alter $d$-monotony.
Assume now that there exists $f \in Ext(K^0_d)$ with $f \notin \{e^u_d,\ u > 0\}$. Using Lemma 2.1 by induction, we would have $(-1)^{d-1} f^{(d-1)}_+ \in Ext(K_1)$. The set of extreme points of $K_1$ is equal to the set of distribution functions of positive point masses, not necessarily equal to one, concentrated on one point: this is exactly the set of right-derivatives of order $d-1$ of functions in $\{e^u_d,\ u > 0\}$. So $f \notin \{e^u_d,\ u > 0\}$ is not possible and we have $Ext(K^0_d) \subset \{e^u_d,\ u > 0\}$. □

## 3 Integral representation result

We are now in a position to provide an alternative proof of Williamson's representation result.

**Theorem 3.1** *$f \in K_d^0$ if and only if there exists a probability measure $\mu$ on $\mathbb{R}^+$ such that*

$$f(t) = \int_0^{+\infty} e_d^u(t)\mu(du). \tag{3.1}$$

The representation (3.1) is given in Theorem 1 in [24]. The main argument used there is a control of the rate at which $x^k f^{(k)}(x)$ goes to 0 when $x$ goes to infinity ($k \leq d-2$). These analytic arguments are very similar to the ones used later by Lévy in [17], where the derivatives are just assumed to be monotone. Here, we do not assume differentiability, but obtain it as a consequence of Theorem 3.1. The second part of next proof follows the lines of the proof of Theorem 3 (Section 14.3) in [16].

**Proof. of Theorem 3.1.** First, to prove that $f$ given by (3.1) is $d$-monotone, just write

$$(-1)^n \left(\prod_{j=1}^n D_{a_j}\right) f(t) = \int_{\mathbb{R}^+} (-1)^n \left(\prod_{j=1}^n D_{a_j}\right) e_d^u(t)\mu(du) \geq 0.$$

Secondly, to prove that a $d$-monotone function takes the form given in (3.1), we introduce the topology on $X$, which is the coarsest for which all the linear functionals $\ell_t$:

$$\ell_t(f) := f(t), \quad t \geq 0$$

are continuous. This makes $X$ a locally convex topological linear space. $K_d^0$ is a convex subset of $X$. Since the values of $f \in K_d^0$ lie between 0 and 1, $K_d^0$ is a subset of $\prod_{0 \leq t} [0,1]$, the set of functions from $(0, \infty)$ to $[0,1]$, which is compact by Tychonov's theorem. So to prove that $K_d^0$ is compact, it suffices to prove that it is closed. For a fixed integer $n$, fixed $a^{(n)} = (a_1, \dots, a_n)$ and $t \geq 0$, the set $K_d^{a^{(n)},t}$ of functions satisfying (2.1) and the normalization $f(0) \in [0,1]$ and $\lim_{t\to\infty} f(t) = 0$ is closed. Hence the set

$$K_d^0 = \bigcap_{n=1}^d \bigcap_{a^{(n)}>0,\, t\geq 0} K_d^{a^{(n)},t}$$

is also closed. By Theorem 2.1, $Ext(K_d) = \{e_d^u,\ u > 0\}$. Let us prove that $\{e_d^u,\ u > 0\}$ is closed. Note that this set contains the function $e_d^\infty(t)$, equal to 1 if $t = 0$ and equal to 0 otherwise. Take a sequence $e_d^{(n)} \in Ext(K_d^0)$ such that for every $t \geq 0$, $e_d^{(n)}(t)$ converges to some $e(t)$, when $n$ goes to infinity. By (2.5), there is some sequence $(u_n)$ of positive numbers such that $e_d^{(n)}(t) = (1 - u_n t)_+^{d-1}$. Since $e_d^{(n)}$ converges pointwise, then $(u_n)$ has a limit. We claim that this limit is different from 0. Indeed,

$$(1 - u_n t)_+^{d-1} = \gamma\left((u_n t, +\infty)\right),$$

where $\gamma$ is a Beta$(1, d)$ distribution. Since for every $t \geq 0$, $\gamma\left((u_n t, +\infty)\right)$ converges to $e(t)$, the map $t \mapsto e(t)$ is the survival function of some probability distribution. Hence $\lim_{n\to\infty} u_n = 0$ is not possible, since it would imply $e(t) \equiv 1$, which is not an admissible survival function. So $\lim_{n\to\infty} u_n = u$, with $0 < u \leq +\infty$ and $e(t) \in \{e_d^u,\ u > 0\}$.
It remains to apply Theorem A.1, with $K = K_d^0$ and $\ell_t(f) = f(t)$. □

**Corollary 3.1** *For $d \geq 1$, the functions in $K_d^0$ are right-continuous at 0 and differentiable on $(0, +\infty)$ up to the order $d-2$.*

**Proof.** From representation (3.1), right-continuity at 0 and derivability result from standard applications of the dominated convergence theorem and the rule for derivation under the integral sign (see for instance [24]). □

# Appendices

## A Choquet's integral representation result

The following theorem of Choquet, that we recall here for completeness, generalizes the Krein-Milman theorem. Informally, it gives sufficient conditions for a point in a convex set to be represented as a barycenter of extreme points.

**Theorem A.1 (Choquet [8])** *Let $X$ be a locally convex topological linear space, $K \subset X$ a convex and compact subset of $X$, $K_e = Ext(K)$ and $\overline{K}_e$ the closure of $K_e$. For all $x \in K$, there exists a probability measure $\mu_x$ on $K_e$ such that*

$$x = \int_{\overline{K}_e} e d\mu_x(e), \quad \textit{in the weak sense.}$$

In Theorem A.1 above, the *weak sense* means that for every continuous linear functional $\ell$ on $X$,

$$\ell(u) = \int_{\overline{K}_e} \ell(e) d\mu_x(e).$$

## References


[1] C. D. Aliprantis and C. Kim, *Border. 2006. infinite dimensional analysis: A hitchhikers guide.*

[2] F. Balabdaoui, J. A. Wellner, et al., *Estimation of a k-monotone density: limit distribution theory and the spline connection*, The Annals of Statistics, 35 (2007), pp. 2536–2564.

[3] S. Bernstein, *Sur la définition et les propriétés des fonctions analytiques d'une variable réelle*, Mathematische Annalen, 75 (1914), pp. 449–468.

[4] ——, *Sur les fonctions absolument monotones*, Acta Math., 52 (1929), pp. 1–66.

[5] M. Buhmann and J. Jäger, *Multiply monotone functions for radial basis function interpolation: Extensions and new kernels*, Journal of Approximation Theory, (2020), p. 105434.

[6] M. D. Buhmann and C. A. Micchelli, *Multiply monotone functions for cardinal interpolation*, Advances in Applied Mathematics, 12 (1991), pp. 358–386.

[7] D. Chafaï, *The bernstein theorem on completely monotone functions, blog entry.* https://djalil.chafai.net/blog/2013/03/23/the-bernstein-theorem-on-completely-monotone-functions. Accessed: 2020-10-30.

[8] G. Choquet, *Theory of capacities*, in Annales de l'institut Fourier, vol. 5, 1954, pp. 131–295.

[9] ——, *Existence et unicité des représentations intégrales au moyen des points extrémaux dans les cônes convexes*, Séminaire Bourbaki, 4 (1956), pp. 33–47.

[10] E. Di Bernardino and D. Rullière, *A note on upper-patched generators for archimedean copulas*, ESAIM: Probability and Statistics, 21 (2017), pp. 183–200.

[11] J. Duda and L. Zajíček, *The banach-zarecki theorem for functions with values in metric spaces*, Proceedings of the American Mathematical Society, (2005), pp. 3631–3633.

[12] F. Gao et al., *Entropy estimate for k-monotone functions via small ball probability of integrated brownian motions*, Electronic Communications in Probability, 13 (2008), pp. 121–130.

[13] C. Genest, J. Nešlehová, and J. Ziegel, *Inference in multivariate archimedean copula models*, Test, 20 (2011), p. 223.

[14] T. Gneiting, *Radial positive definite functions generated by euclid's hat*, Journal of Multivariate Analysis, 69 (1999), pp. 88–119.

[15] B. Korenblum, R. O'Neil, K. Richards, and K. Zhu, *Totally monotone functions with applications to the bergman space*, Transactions of the American Mathematical Society, 337 (1993), pp. 795–806.

[16] P. Lax, *Functional Analysis*, Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts, Wiley, 2002.

[17] P. Lévy, *Extensions d'un théorème de d. dugué et m. girault*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 1 (1962), pp. 159–173.

[18] A. J. McNeil and J. Nešlehová, *From archimedean to liouville copulas*, Journal of Multivariate Analysis, 101 (2010), pp. 1772–1790.

[19] A. J. McNeil and J. Nešlehová, *Multivariate archimedean copulas, d-monotone functions and $\ell_1$-norm symmetric distributions*, Ann. Statist., 37 (2009), pp. 3059–3097.

[20] I. P. Natanson, *Theory of functions of a real variable*, Courier Dover Publications, 2016.

[21] R. R. Phelps, *Lectures on Choquet's theorem*, Springer Science & Business Media, 2001.

[22] W. Rudin, *Real and complex analysis*, Tata McGraw-hill education, 2006.

[23] D. V. Widder, *Laplace transform (PMS-6)*, Princeton university press, 2015.

[24] R. E. Williamson, *Multiply monotone functions and their laplace transforms*, Duke Math. J., 23 (1956), pp. 189–207.