



## UvA-DARE (Digital Academic Repository)

### Differentiation by integration using orthogonal polynomials, a survey

Diekema, E.; Koornwinder, T.H.

**DOI**

[10.1016/j.jat.2012.01.003](https://doi.org/10.1016/j.jat.2012.01.003)

**Publication date**

2012

**Document Version**

Author accepted manuscript

**Published in**

Journal of Approximation Theory

[Link to publication](#)

**Citation for published version (APA):**

Diekema, E., & Koornwinder, T. H. (2012). Differentiation by integration using orthogonal polynomials, a survey. *Journal of Approximation Theory*, 164(5), 637-667.  
<https://doi.org/10.1016/j.jat.2012.01.003>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Differentiation by integration using orthogonal polynomials, a survey

Enno Diekema\* and Tom H. Koornwinder†

## Abstract

This survey paper discusses the history of approximation formulas for  $n$ -th order derivatives by integrals involving orthogonal polynomials. There is a large but rather disconnected corpus of literature on such formulas. We give some results in greater generality than in the literature. Notably we unify the continuous and discrete case. We make many side remarks, for instance on wavelets, Mantica's Fourier-Bessel functions and Greville's minimum  $R_\alpha$  formulas in connection with discrete smoothing.

## 1 Introduction

In many applications one needs to estimate or approximate the first or higher derivative of a function which is only given in sampled form or which is perturbed by noise. Good candidates for an approximation of the first derivative  $f'(x)$  are the two expressions

$$\frac{3}{2\delta} \int_{-1}^1 f(x + \delta\xi) \xi d\xi \quad (1.1)$$

and

$$\frac{3}{2N(N + \frac{1}{2})(N + 1)\delta} \sum_{\xi=-N}^N f(x + \delta\xi) \xi \quad (1.2)$$

for  $\delta$  small. The first one is continuous, the second one discrete. These formulas have a long history going back to Cioranescu [11] (1938), Haslam-Jones [29] (1953), Lanczos [42, (5-9.1)] (1956) and Savitzky & Golay [57] (1964). What remains hidden in (1.1) and (1.2) is that the factor  $\xi$  in the integrand or summand can better be considered as an orthogonal polynomial of degree 1 with respect to a constant weight function on  $[-1, 1]$  (in case of (1.1)) or with respect to constant weights on  $\{-N, -N + 1, \dots, N\}$  (in case of (1.2)). With this point of view it can be immediately shown that (1.1) and (1.2) tend to  $f'(x)$  as  $\delta \downarrow 0$ . Moreover the way is opened to a far reaching generalization of (1.1) and (1.2) for the approximation of higher derivatives

---

\*e.diekema@gmail.com

†Korteweg-de Vries Institute, University of Amsterdam, T.H.Koornwinder@uva.nl

and with the involvement of general orthogonal polynomials. Such results were already given by Cioranescu [11] in 1938.

Curiously enough, none of the later papers mentioned above is referring to one of the earlier papers. The results of Cioranescu [11] and Haslam-Jones [29] were hardly taken up by anybody. On the other hand Lanczos [42, (5-9.1)] and in particular Savitzky & Golay [57] had a lot of follow-up by others. One reason for this citation behaviour is probably that Cioranescu and Haslam-Jones were pure analysts, Lanczos was an applied mathematician working in numerical analysis, and Savitzky & Golay were motivated by spectroscopy, considered as a part of chemistry.

It is the aim of the present paper to give a survey of these results, developments and further considerations suggested by them. Moreover, we formulate some results in a more general way than has probably appeared before in literature. It was for us a surprise to see that so many different parts of classical analysis and of applied mathematics are tied together by this theme. All papers until now only treated smaller parts of this wide field. We hope to share with our readers the pleasure to have a comprehensive view.

The present work stems from a long practice by the first author in signal processing in applied situations, where he met the problem of differentiating an analog signal (and later a sampled signal) disturbed by noise, see for instance Strik [63]. The main problem occurring there was the difficulty to build (or program) an ideal differentiator, because the noise of the system will cause an instability. Therefore an integrating factor for the high frequencies to the differentiator is needed. When the signal is sampled the same problem occurs (Hamming [26]). Without being aware of the literature mentioned in the beginning of this Introduction, he tried to use an integrating factor by the method of the least squares and then he independently found special cases of the approximation formulas for higher derivatives by integrals involving all classical orthogonal polynomials as well as the Chebyshev polynomials of a discrete variable. He never published the results, but he used them as material for a course in stochastic system theory at the "Saxion Hogeschool" in Enschede.

The contents of this paper are as follows. In section 2 we give preliminaries on orthogonal polynomials and on the Taylor formula. In section 3 we start formulating the approximation theorem in great generality and next discuss how the contributions of Cioranescu, Haslam-Jones and Lanczos are related to this general theorem. We emphasize the important role of least-square approximation behind this theory. Our discussion gives room for several side observations, for instance on Jacobi type polynomials and on wavelets. Section 4 is focused on the discrete case and the applications to filters. We start with a multi-term extension of the main theorem in section 3. Its special case of constant weights contains the seminal work of Savitzky & Golay. We introduce the characteristic (or transfer) function and we make connection with Mantica's [46] Fourier-Bessel functions. In the smoothing case we discuss the work by Greville [24] (based on older work by Sheppard [61]) on so-called minimum  $R_\alpha$  and minimum  $R_\infty$  formulas. In the Appendix we give new derivations of the characteristic functions for these cases, using Hahn and Krawtchouk polynomials. The  $R_\infty$  case connects with another survey paper [39] by the second author and Schlosser. Finally, in section 5, we discuss log-log plots of transfer functions in some simple cases.

## 2 Preliminaries

### 2.1 Orthogonal polynomials

Let  $\mu$  be a positive Borel measure on  $\mathbb{R}$  with infinite support (or equivalently a nondecreasing function on  $\mathbb{R}$  with an infinite number of points of increase) such that  $\int_{\mathbb{R}} |x|^n d\mu(x) < \infty$  for all  $n \in \mathbb{Z}_{\geq 0}$ . Consider polynomials  $p_n$  ( $n \in \mathbb{Z}_{\geq 0}$ ) of degree  $n$  such that

$$\int_{\mathbb{R}} p_m(x) p_n(x) d\mu(x) = 0 \quad (m \neq n). \quad (2.1)$$

The  $p_n$  are *orthogonal polynomials* with respect to the measure  $\mu$ , see for instance Szegő [64]. Up to constant nonzero factors they are uniquely determined by the above properties. If  $\mu$  has support within some closed interval  $I$  then we say that the  $p_n$  are orthogonal polynomials with respect to  $\mu$  on  $I$ . Typical cases of the orthogonality measure  $\mu$  are:

1.  $d\mu(x) = w(x) dx$  on  $I$  with the *weight function*  $w$  a nonnegative integrable function on  $I$ . Then (2.1) takes the form

$$\int_I p_m(x) p_n(x) w(x) dx = 0 \quad (m \neq n).$$

2.  $\mu$  has discrete infinite support  $\{x_0, x_1, x_2, \dots\}$ . So there are positive numbers  $w_0, w_1, w_2, \dots$  (*weights*) such that (2.1) takes the form

$$\sum_{k=0}^{\infty} p_m(x_k) p_n(x_k) w_k = 0 \quad (m \neq n). \quad (2.2)$$

3. Contrary to what was supposed earlier, we can also consider the case that  $\mu$  has finite support  $\{x_0, x_1, \dots, x_N\}$  with corresponding weights  $w_0, w_1, \dots, w_N$ . Then we have orthogonal polynomials  $p_n$  only for  $n = 0, 1, \dots, N$  and (2.1) takes the form

$$\sum_{k=0}^N p_m(x_k) p_n(x_k) w_k = 0 \quad (m \neq n). \quad (2.3)$$

Special examples of case 1 are given by the *classical orthogonal polynomials* (Jacobi, Laguerre and Hermite polynomials). In particular, we will meet the *Legendre polynomials*  $P_n$ , which are special Jacobi polynomials and where  $I = [-1, 1]$ ,  $w(x) = 1$  and  $P_n(1) = 1$ .

A special example of case 3 are the *Hahn polynomials*  $x \mapsto Q_n(x; \alpha, \beta, N)$  for  $\alpha = \beta = 0$  ( $n = 0, 1, \dots, N$ ). Here  $x_i = i, w_i = 1$  ( $i = 0, 1, \dots, N$ ) and  $Q_n(0; 0, 0, N) = 1$  (see [36, §9.5] and references given there, or [49, Ch. 2], where another notation is used). Hahn polynomials of general parameters were already introduced in 1875 by Chebyshev [10], long before Hahn, but the above special case of constant weights is in particular named after Chebyshev, although in a slightly different notation and normalization. See *Chebyshev's polynomials of a discrete*

variable  $t_n(x) = t_n(x, N)$  ( $n = 0, 1, \dots, N - 1$ ) in [64, §2.8], [18, §10.23]. They are orthogonal on the set  $\{0, 1, \dots, N - 1\}$  with respect to constant weights 1. Hence we must have that  $t_n(x, N) = \text{const. } Q_n(x; 0, 0, N - 1)$ . The constant can be computed by comparing the recurrence relation [36, (9.5.3)] for  $\alpha = \beta = 0$  and  $N$  replaced by  $N - 1$  with the recurrence relation [18, 10.23(6)]. Then we obtain:

$$t_n(x, N) = (-N + 1)_n Q_n(x; 0, 0, N - 1), \quad (2.4)$$

where  $(a)_n := a(a + 1) \dots (a + n - 1)$  is the *Pochhammer symbol*. Thus  $t_n(0, N) = (-N + 1)_n$ . These polynomials are also known as *Gram polynomials*, see [33, §7.13 and §7.16]. This last name we will use in this paper. The polynomials  $x \mapsto t_n(x, N)$  have the shifted Legendre polynomials  $x \mapsto P_n(2x - 1)$  (orthogonal on  $[0, 1]$  with respect to a constant weight function) as a limit case (see [64, (2.8.6)]):

$$\lim_{N \rightarrow \infty} N^{-n} t_n(Nx, N) = P_n(2x - 1) \quad (2.5)$$

For given orthogonal polynomials  $p_n$  define the constants  $h_n$  and  $k_n$  by

$$h_n := \int_{\mathbb{R}} p_n(x)^2 d\mu(x), \quad p_n(x) = k_n x^n + \text{terms of degree less than } n. \quad (2.6)$$

**Lemma 2.1.** *We have*

$$\int_{\mathbb{R}} p_n(x) x^n d\mu(x) = \frac{h_n}{k_n} \quad (2.7)$$

**Proof** We have  $k_n x^n = p_n(x) + q_{n-1}(x)$  with  $q_{n-1}$  a polynomial of degree  $< n$ . Hence

$$k_n \int_{\mathbb{R}} p_n(x) x^n d\mu(x) = \int_{\mathbb{R}} p_n(x)^2 d\mu(x) + \int_{\mathbb{R}} p_n(x) q_{n-1}(x) d\mu(x) = h_n + 0 = h_n \quad \square$$

One of the properties which characterize the classical orthogonal polynomials is that they are given by a (generalized) *Rodrigues formula*

$$p_n(x) = \frac{1}{K_n w(x)} \frac{d^n}{dx^n} (w(x) X(x)^n) \quad (2.8)$$

(see [18, 10.6(1)]). Here  $X$  is a polynomial of degree  $\leq 2$  and

$$K_n = \frac{(-1)^n k_n n!}{h_n} \int_{\mathbb{R}} (X(x))^n d\mu(x). \quad (2.9)$$

For the proof of (2.9) substitute (2.8) and  $d\mu(x) = w(x) dx$  in (2.7) and perform integration by parts  $n$  times.

The explicit values of  $h_n$  and  $k_n$  defined by (2.6) can be given in our two main examples:

- Legendre polynomials  $P_n$  (see [36, (9.8.63), ((9.8.65))]):

$$h_n = \frac{2}{2n + 1}, \quad k_n = 2^{-n} \binom{2n}{n}. \quad (2.10)$$

From this we immediately obtain the values of  $h_n$  and  $k_n$  in the case of shifted Legendre polynomials  $x \mapsto P_n(2x - 1)$ :

$$h_n = \frac{1}{2n + 1}, \quad k_n = \binom{2n}{n}. \quad (2.11)$$

- Gram polynomials  $x \mapsto t_n(x, N)$  (combine (2.4) with [36, (9.5.2), (9.5.4)]):

$$h_n = \frac{(N - n)2n + 1}{2n + 1}, \quad k_n = \binom{2n}{n}. \quad (2.12)$$

From this we immediately obtain the values of  $h_n$  and  $k_n$  in the case of centered Gram polynomials  $x \mapsto t_n(x + N, 2N + 1)$ :

$$h_n = \frac{(2N + 1 - n)2n + 1}{2n + 1}, \quad k_n = \binom{2n}{n}. \quad (2.13)$$

The *reproducing kernel* for the space  $\mathcal{P}_n$  of polynomials of degree  $\leq n$  in the Hilbert space  $L^2(\mathbb{R}, \mu)$  is given by

$$\mathbf{K}_n(x, y) := \sum_{k=0}^n \frac{p_k(x)p_k(y)}{h_k} \quad (x, y \in \mathbb{R}). \quad (2.14)$$

Then (see [1, Remark 5.2.2]) the *Christoffel-Darboux formula* gives

$$\mathbf{K}_n(x, y) = \frac{k_n}{k_{n+1} h_n} \frac{p_{n+1}(x)p_n(y) - p_n(x)p_{n+1}(y)}{x - y} \quad (x \neq y). \quad (2.15)$$

The integral operator  $\mathcal{K}_n$  corresponding to (2.14) is given by

$$(\mathcal{K}_n f)(x) := \int_{\mathbb{R}} f(y) K_n(x, y) d\mu(y) \quad (f \in L^2(\mathbb{R}), x \in \mathbb{R}), \quad (2.16)$$

It is the orthogonal projection of the Hilbert space  $L^2(\mathbb{R}, \mu)$  onto  $\mathcal{P}_n$ . In particular,

$$\mathcal{K}_n f = f \quad (f \in \mathcal{P}_n). \quad (2.17)$$

Furthermore, for  $f \in L^2(\mathbb{R}, \mu)$ ,  $\mathcal{K}_n f$  is the element of  $\mathcal{P}_n$  which is on minimal distance to  $f$  (in the metric of the Hilbert space  $L^2(\mathbb{R}, \mu)$ ).

Later we will need the following. For Legendre polynomials formula (2.15) for  $y = 1$  and with  $n$  replaced by  $n - 1$  becomes:

$$\begin{aligned} \mathbf{K}_{n-1}(x, 1) &= \frac{1}{2}n \frac{P_n(x) - P_{n-1}(x)}{x - 1} \\ &= \frac{1}{2}n P_{n-1}^{(1,0)}(x) \end{aligned} \quad (2.18)$$

$$= \frac{1}{2}(P'_n(x) + P'_{n-1}(x)). \quad (2.19)$$

The  $P_{n-1}^{(1,0)}(x)$  in (2.18) is a Jacobi polynomial (see [64, Ch. 4]). We used (2.10) in the first equality, [18, 10.8(32)] in (2.18), and [18, 10.10(13), 10.10(14)] in (2.19).

## 2.2 Taylor formula

Recall a version of Taylor's theorem formulated by Hardy [28, §151]:

**Proposition 2.2.** *Let  $x \in \mathbb{R}$  and let  $I$  be an interval containing  $x$ . Let  $f$  be a continuous function on  $I$  such that its derivatives of order  $1, 2, \dots, n$  at  $x$  exist. Then*

$$f(y) = \sum_{k=0}^n \frac{f^{(k)}(x)}{k!} (y-x)^k + o(|y-x|^n) \quad \text{as } y \rightarrow x \text{ on } I. \quad (2.20)$$

In this proposition the derivatives should be interpreted as right or left derivatives if  $x$  is an endpoint of the interval  $I$ . (Although this special case is not explicit in Hardy's formulation, it is also a consequence of his proof.)

Proposition 2.2 suggests a notion more general than  $n$ -th derivative. Let  $f$  be a continuous function on an interval  $I$  containing  $x$  and let there be constants  $c_0, c_1, \dots, c_n$  such that

$$f(y) = \sum_{k=0}^n \frac{c_k}{k!} (y-x)^k + o(|y-x|^n) \quad \text{as } y \rightarrow x \text{ on } I. \quad (2.21)$$

Then we call  $c_n$  the  $n$ -th *Peano derivative* of  $f$  at  $x$ . This definition goes back to Peano [52] in 1891. By Proposition 2.2 the existence of  $f^{(n)}(x)$  implies the existence of the  $n$ -th Peano derivative, equal to  $f^{(n)}(x)$ . The converse implication is true for  $n = 1$  but not for  $n > 1$ , see a counterexample in [19, Example 1.2].

For later use we restate Proposition 2.2 as follows:

**Proposition 2.3.** *With the assumptions of Proposition 2.2 we have*

$$f(x + \delta) = \sum_{k=0}^n \frac{f^{(k)}(x)}{k!} \delta^k + \delta^n F_{x,n}(\delta) \quad (2.22)$$

with  $F_{x,n}$  continuous on  $I - x$  and  $F_{x,n}(0) = 0$ . Furthermore,  $F_{x,n}$  is bounded on  $I - x$  if  $f$  is bounded on  $I$ . Finally, if  $I$  is unbounded and  $f$  is of polynomial growth on  $I$  then  $F_{x,n}$  is of polynomial growth on  $I - x$ .

We can say more about the remainder term in (2.22) if moreover  $f \in C^n(I)$  (see for instance Apostol [3, Theorem 7.6]):

**Proposition 2.4.** *Keep the assumptions of Proposition 2.2. Assume moreover that  $f \in C^n(I)$ . Then for  $F_{x,n}(\delta)$  in (2.22) we have*

$$F_{x,n}(\delta) = \frac{1}{(n-1)!} \int_0^1 (f^{(n)}(x+t\delta) - f^{(n)}(x)) (1-t)^{n-1} dt \quad (2.23)$$

and the function  $(x, y) \mapsto F_{x,n}(y-x)$  is continuous on  $I \times I$ . If  $f^{(n)}$  is of polynomial growth on  $I$  then  $F_{x,n}(\delta) \rightarrow 0$  as  $\delta \rightarrow 0$  uniformly for  $x$  in compact subsets of  $I$ .

### 3 Higher derivatives approximated by integrals

Let us first state and prove the main theorem and next discuss the many instances of it in the literature, usually more restricted but occasionally more general than our formulation.

**Theorem 3.1.** *For some  $n$  let  $p_n$  be an orthogonal polynomial of degree  $n$  with respect to the orthogonality measure  $\mu$ . Let  $x \in \mathbb{R}$ . Let  $I$  be a closed interval such that, for some  $\varepsilon > 0$ ,  $x + \delta\xi \in I$  if  $0 \leq \delta \leq \varepsilon$  and  $\xi \in \text{supp}(\mu)$ . Let  $f$  be a continuous function on  $I$  such that its derivatives of order  $1, 2, \dots, n$  at  $x$  exist. In addition, if  $I$  is unbounded, assume that  $f$  is of at most polynomial growth on  $I$ . Then*

$$f^{(n)}(x) = \frac{k_n n!}{h_n} \lim_{\delta \downarrow 0} \frac{1}{\delta^n} \int_{\mathbb{R}} f(x + \delta\xi) p_n(\xi) d\mu(\xi), \quad (3.1)$$

where the integral converges absolutely.

**Proof** If  $I$  is bounded then the integral in (3.1) converges absolutely by continuity of  $f$ . If  $I$  is unbounded then, for fixed  $\delta \in [0, \varepsilon]$ , we have for some  $r \geq 0$  that  $f(x + \delta\xi) = O(|\xi|^r)$  as  $\xi \rightarrow \pm\infty$  on  $\text{supp}(\mu)$ . So also in that case the integral in (3.1) converges absolutely.

By substitution of (2.22), by orthogonality and by (2.7) we have:

$$\begin{aligned} \frac{1}{\delta^n} \int_{\mathbb{R}} f(x + \delta\xi) p_n(\xi) d\mu(\xi) &= \sum_{k=0}^n \frac{f^{(k)}(x) \delta^{k-n}}{k!} \int_{\mathbb{R}} \xi^k p_n(\xi) d\mu(\xi) + \int_{\mathbb{R}} F_{x,n}(\delta\xi) \xi^n p_n(\xi) d\mu(\xi) \\ &= \frac{f^{(n)}(x)}{n!} \int_{\mathbb{R}} \xi^n p_n(\xi) d\mu(\xi) + \int_{\mathbb{R}} F_{x,n}(\delta\xi) \xi^n p_n(\xi) d\mu(\xi) \\ &= \frac{h_n}{k_n n!} f^{(n)}(x) + \int_{\mathbb{R}} F_{x,n}(\delta\xi) \xi^n p_n(\xi) d\mu(\xi). \end{aligned}$$

Thus the theorem will be proved if we can show that

$$\lim_{\delta \downarrow 0} \int_{\mathbb{R}} F_{x,n}(\delta\xi) \xi^n p_n(\xi) d\mu(\xi) = 0. \quad (3.2)$$

By the second part of Proposition 2.3 we have the estimate  $|F_{x,n}(h)| \leq C(1 + |h|)^r$  ( $h \in I - x$ ) for some  $C > 0$ ,  $r \geq 0$ . Hence, for  $\delta \in [0, \varepsilon]$  and  $\xi \in \text{supp}(\mu)$  we have the estimate  $|F_{x,n}(\delta\xi)| \leq C(1 + \varepsilon|\xi|)^r$ . Thus, the dominated convergence theorem can be applied to the left-hand side of (3.2). Then, again by Proposition 2.3, it follows that (3.2) is true.  $\square$

Note the following special cases of (3.1).

- **Gram polynomials**  $x \mapsto t_n(x, N)$  (use (2.12)):

$$f^{(n)}(x) = \frac{(2n+1)!(N-n-1)!}{n!(N+n)!} \lim_{\delta \downarrow 0} \frac{1}{\delta^n} \sum_{\xi=0}^{N-1} f(x + \delta\xi) t_n(\xi, N). \quad (3.3)$$



- **Centered Gram polynomials**  $x \mapsto t_n(x + N, 2N + 1)$  on  $2N + 1$  lattice points (use (2.13)):

$$f^{(n)}(x) = \frac{(2n + 1)! (2N - n)!}{n! (2N + n + 1)!} \lim_{\delta \downarrow 0} \frac{1}{\delta^n} \sum_{\xi=-N}^N f(x + \delta\xi) t_n(\xi + N, 2N + 1). \quad (3.4)$$

In particular, for  $n = 1$ :

$$f'(x) = \frac{3}{2N(N + \frac{1}{2})(N + 1)} \lim_{\delta \downarrow 0} \frac{1}{\delta} \sum_{\xi=-N}^N f(x + \delta\xi) \xi. \quad (3.5)$$

Analogues of (3.4) and (3.5) might also be given for the centered Gram polynomials  $x \mapsto t_n(x - N + \frac{1}{2}, 2N)$  on  $2N$  lattice points.

- **Legendre polynomials**  $P_n$  (use (2.10)):

$$f^{(n)}(x) = \frac{(2n + 1)!}{2^{n+1}n!} \lim_{\delta \downarrow 0} \frac{1}{\delta^n} \int_{-1}^1 f(x + \delta\xi) P_n(\xi) d\xi. \quad (3.6)$$

In particular, for  $n = 1$ :

$$f'(x) = \frac{3}{2} \lim_{\delta \downarrow 0} \frac{1}{\delta} \int_{-1}^1 f(x + \delta\xi) \xi d\xi. \quad (3.7)$$

- **Shifted Legendre polynomials**  $x \mapsto P_n(2x - 1)$  (use (2.11)):

$$f^{(n)}(x) = \frac{(2n + 1)!}{n!} \lim_{\delta \downarrow 0} \frac{1}{\delta^n} \int_0^1 f(x + \delta\xi) P_n(2\xi - 1) d\xi. \quad (3.8)$$

### 3.1 Cioranescu's 1938 paper

A variant of Theorem 3.1 was first stated and proved by Cioranescu [11, formula (M')] in 1938 for the case that  $d\mu(x) = w(x) dx$  is absolutely continuous with bounded support within an interval  $[a, b]$ . He showed for  $f \in C^n([a, b])$  that there exists  $\eta \in (a, b)$  such that

$$n! \frac{\int_a^b f(y) p_n(y) w(y) dy}{\int_a^b y^n p_n(y) w(y) dy} = f^{(n)}(\eta). \quad (3.9)$$

Then he took limits for  $b \downarrow a$  in the left-hand side of (3.9) (see [11, formula (9)]) with  $p_n$  remaining an orthogonal polynomial on the shrinking interval  $[a, b]$  with respect to the weight function  $w$  restricted to  $[a, b]$ . The limit on the right-hand side of (3.9) then becomes  $f^{(n)}(a)$ . In general, this limit formula for  $f^{(n)}(a)$  will not be contained in (3.1) since the weight function (after rescaling it to a fixed interval) will not remain the same during the limit process. But Cioranescu's limit result in the case of shifted Legendre polynomials is the same as (3.8). The case  $n = 1$  of (3.8) is explicitly mentioned by Cioranescu (see [11, formula (9')]).

### 3.2 Substitution of the Rodrigues formula

For classical orthogonal polynomials  $p_n$  and for  $f \in C^n(I)$  with  $f^{(n)}$  of polynomial growth on  $I$  we can also prove (3.1) by substituting  $d\mu(x) = w(x)$  and the Rodrigues formula (2.8) together with (2.9), and by performing integration by parts  $n$  times:

$$\frac{k_n n!}{h_n \delta^n} \int_I f(x + \delta\xi) p_n(\xi) w(\xi) d\xi = \frac{\int_I f^{(n)}(x + \delta\xi) X(\xi)^n w(\xi) d\xi}{\int_I X(\xi)^n w(\xi) d\xi} \rightarrow f^{(n)}(x) \quad \text{as } \delta \downarrow 0. \quad (3.10)$$

In the Legendre case  $w(x) = 1$  this was already observed by Cioranescu [11, p.296].

### 3.3 Haslam-Jones' 1953 paper

Next Theorem 3.1, for the case that  $\mu$  has bounded support, was observed (with proof omitted as being easy) in 1953 by Haslam-Jones [29, p.192], who was apparently not aware of Cioranescu's result. In fact, in his formulation the measure  $\mu$  only has to be real, not necessarily positive. Furthermore,  $f$  only has to be continuous with an  $n$ -th Peano derivative at  $x$  (see (2.21)). Note that our proof of Theorem 3.1 can be used without essential changes under the weaker hypotheses of Haslam-Jones.

In fact, the assumptions in [29] are still weaker. Haslam-Jones assumes, for given  $n > 0$ , a real, not necessarily positive measure  $\nu$  with bounded support (or equivalently a function  $\nu$  of bounded variation) on a finite interval  $J$  such that  $\int_J x^k d\nu(x) = 0$  for  $k = 0, 1, \dots, n-1$  and  $\int_J x^n d\nu(x) = \kappa \neq 0$ . Then for a function  $f$  which is continuous on a neighbourhood of  $x$  and has  $n$ -th Peano derivative  $c_n$  in  $x$  we have

$$c_n = \frac{n!}{\kappa} \lim_{\delta \downarrow 0} \frac{1}{\delta^n} \int_J f(x + \delta\xi) d\nu(\xi). \quad (3.11)$$

Again this can be proved as we did for Theorem 3.1, without essential changes.

### 3.4 A special case of Haslam-Jones' results

We will consider here a special case of (3.11) which is essentially different from Theorem 3.1. Let  $\{p_m\}$  be a system of orthogonal polynomials on  $[-1, 1]$  with respect to a positive Borel measure  $\mu$ . Let  $K_m(x, y)$  be the corresponding Christoffel-Darboux kernel given by (2.14), (2.15). Fix  $n$  and define the measure  $\nu$  in (3.11) by

$$\int_{-1}^1 f(\xi) d\nu(\xi) := f(1) - \int_{-1}^1 f(\xi) K_{n-1}(\xi, 1) d\mu(\xi). \quad (3.12)$$

Indeed, by the reproducing kernel property the right-hand side of (3.12) equals 0 if  $f$  is a polynomial of degree  $< n$ , while for  $f(\xi) := \xi^n$  the right-hand side of (3.12) becomes

$$- \int_{-1}^1 \xi^{n-1} (\xi - 1) K_{n-1}(\xi, 1) d\mu(\xi) = \frac{k_{n-1} p_n(1)}{k_n h_{n-1}} \int_{-1}^1 \xi^{n-1} p_{n-1}(\xi) d\mu(\xi) = \frac{p_n(1)}{k_n} \neq 0.$$

Thus for this case (3.11) becomes

$$c_n = \frac{n! k_n}{p_n(1)} \lim_{\delta \downarrow 0} \frac{1}{\delta^n} \left( f(\delta) - \int_{-1}^1 f(x + \delta\xi) K_{n-1}(\xi, 1) d\mu(\xi) \right). \quad (3.13)$$

In particular, take  $d\mu(\xi) := d\xi$  on  $[-1, 1]$ . Then substitute (2.19), by which (3.13) takes the form

$$c_n = 2^n \left(\frac{1}{2}\right)_n \lim_{\delta \downarrow 0} \frac{1}{\delta^n} \left( f(\delta) - \frac{1}{2} \int_{-1}^1 f(x + \delta\xi) (P'_n(\xi) + P'_{n-1}(\xi)) d\xi \right). \quad (3.14)$$

For this case Haslam-Jones [29] showed that if the limit on the right of (3.14) exists then the  $n$ -th Peano derivative of  $f$  at  $x$  exists and it equals  $c_n$  given by (3.14). A different proof of this result was given by Gordon [21].

### 3.5 Connection with Jacobi type orthogonal polynomials

For a larger family of examples than (3.14) consider formula (3.12) with  $d\mu(x) := (1-x)^\alpha(1+x)^{\beta+1} dx$  ( $\alpha, \beta > -1$ ). Then  $p_m(x) = P_m^{(\alpha, \beta+1)}(x)$ , a *Jacobi polynomial* (see [64, Ch. 4]). From [64, (4.5.3)] we obtain that

$$\mathbf{K}_{n-1}(x, 1) = \frac{\Gamma(n + \alpha + \beta + 2)}{2^{\alpha+\beta+2}\Gamma(\alpha + 1)\Gamma(n + \beta + 1)} P_{n-1}^{(\alpha+1, \beta+1)}(x).$$

Then the vanishing of the right-hand side of (3.12) for polynomials  $f$  of degree  $< n$  can be written more explicitly as

$$\begin{aligned} \frac{\Gamma(\alpha + \beta + 2)}{2^{\alpha+\beta+1}\Gamma(\alpha + 1)\Gamma(\beta + 1)} \int_{-1}^1 f(x) \frac{(1+x) P_{n-1}^{(\alpha+1, \beta+1)}(x)}{2P_{n-1}^{(\alpha+1, \beta+1)}(1)} (1-x)^\alpha(1+x)^\beta dx \\ - \frac{(\beta + 1)_n(n-1)!}{(\alpha + \beta + 2)_n(\alpha + 2)_{n-1}} f(1) = 0. \end{aligned}$$

Hence, for fixed  $n$ , the polynomial  $(1+x)P_{n-1}^{(\alpha+1, \beta+1)}(x)$  is the  $n$ -th degree orthogonal polynomial with respect to a measure on  $[-1, 1]$  consisting of the weight function  $(1-x)^\alpha(1+x)^\beta$  and a negative constant times a delta weight at  $x = 1$ . On comparing with [37, Theorem 3.1] (extended to negative multiples of delta weights by analytic continuation) we can identify this orthogonal polynomial with a *Jacobi type polynomial*:

$$P_n^{(\alpha, \beta; 0, N)}(x) = c(1+x)P_{n-1}^{(\alpha+1, \beta+1)}(x),$$

where

$$N = -\frac{(\beta + 1)_n(n-1)!}{(\alpha + \beta + 2)_n(\alpha + 2)_{n-1}} \quad \text{and} \quad c = \frac{P_n^{(\alpha, \beta; 0, N)}(1)}{2P_{n-1}^{(\alpha+1, \beta+1)}(1)} = \frac{\alpha + 1}{2n}.$$

This corresponds correctly with [37, (2.1)], which simplifies for  $M = 0$  and the above choice of  $N$  to

$$P_n^{(\alpha, \beta; 0, N)}(x) = \frac{\alpha + 1}{n(n + \alpha + \beta + 1)} (1 + x) \frac{d}{dx} P_n^{(\alpha, \beta)}(x).$$

The above formulas extend by continuity to the case  $\beta = -1$ , which occurs if  $d\mu(x) = dx$  (the case considered in (3.14)).

### 3.6 Lanczos' 1956 work

In a book published in 1956 Lanczos [42, (5-9.1)], apparently unaware of the earlier work by Cioranescu [11] and Haslam-Jones [29], rediscovered formula (3.7). He called this *differentiation by integration*. His work got quite a lot of citations, see for instance [25], [60], [32], [54], [8], [9], [66]. The name *Lanczos derivative*, notated as  $f'_L(x)$ , became common for a value obtained from the right-hand side of (3.7). In [54] and [8] also (3.6) (the Legendre case for general  $n$ ) was rediscovered.

As an important new aspect Lanczos observed that

$$\frac{3}{2\delta} \int_{-1}^1 f(x + \delta\xi) \xi d\xi = \frac{\int_{-1}^1 f(x + \delta\xi) \xi d\xi}{\delta \int_{-1}^1 \xi^2 d\xi},$$

as an approximation of  $f'(x)$ , is the limit as  $N \rightarrow \infty$  of the quotient of Riemann sums

$$\frac{N^{-1} \sum_{\xi=-N}^N f(x + \delta\xi/N) \xi/N}{\delta N^{-1} \sum_{\xi=-N}^N (\xi/N)^2} = \frac{3}{2N(N + \frac{1}{2})(N + 1)} \frac{N}{\delta} \sum_{\xi=-N}^N f(x + N^{-1}\delta\xi) \xi.$$

We have seen this last expression already in (3.5) (the special case of (3.1) with a centered Gram polynomial of degree 1). The expression approximates  $f'(x)$  for  $N^{-1}\delta$  small. In fact, (3.5) was the starting point of Lanczos, see [42, (5-8.4)].

### 3.7 Interpretation by least-square approximation

Lanczos [42, (5-8.4)] arrived at (3.5) by a least-square minimization (a technique invented by Gauss and Legendre, see for instance [16]). For this compare the function  $\xi \mapsto f(x + \delta\xi)$  with a linear function  $g(\xi) := a_0 + a_1\xi$  such that the squared distance

$$S(a_0, a_1) := \sum_{\xi=-N}^N (f(x + \delta\xi) - g(\xi))^2 = \sum_{\eta=-\delta N, -\delta(N-1), \dots, \delta N} (f(x + \eta) - (a_0 + \delta^{-1}a_1\eta))^2$$

is minimal. Then the slope  $\delta^{-1}a_1$  of the straight line  $\eta \mapsto a_0 + \delta^{-1}a_1\eta$  minimizing the distance will approximate  $f'(x)$  for small  $\delta$ . The minimum is achieved for a unique  $(a_0, a_1)$ , where one finds  $a_1$  in this simple case already by solving  $\frac{\partial}{\partial a_1} S(a_0, a_1) = 0$ . Thus Lanczos obtained

$$\delta^{-1}a_1 = \frac{\sum_{\xi=-N}^N f(x + \delta\xi)\xi}{\delta \sum_{\xi=-N}^N \xi^2} = \frac{3}{2\delta N(N + \frac{1}{2})(N + 1)} \sum_{\xi=-N}^N f(x + \delta\xi) \xi,$$

and he thus arrived at (3.5).

We can interpret (3.1) as a more general least-square approximation. By the assumptions on  $f$  in the Theorem 3.1 the function  $\xi \mapsto f(x + \xi\delta)$  is in  $L^2(\mathbb{R}, \mu)$  for each  $\delta \in [0, \varepsilon]$ . Let  $\xi \mapsto P_{n,x,\delta}[f](\xi)$  be the polynomial of degree  $\leq n$  which is on minimal distance from the function  $\xi \mapsto f(x + \xi\delta)$  in the Hilbert space  $L^2(\mathbb{R}, \mu)$ . Then

$$P_{n,x,\delta}[f](\eta) = \sum_{k=0}^n \frac{1}{h_k} \left( \int_{\mathbb{R}} f(x + \delta\xi) p_k(\xi) d\mu(\xi) \right) p_k(\eta). \quad (3.15)$$

Then

$$\delta^{-n} (P_{n,x,\delta}[f])^{(n)}(0) = \frac{p_n^{(n)}(0)}{\delta^n h_n} \int_{\mathbb{R}} f(x + \delta\xi) p_n(\xi) d\mu(\xi) = \frac{k_n n!}{\delta^n h_n} \int_{\mathbb{R}} f(x + \delta\xi) p_n(\xi) d\mu(\xi)$$

approximates  $f^{(n)}(x)$  as  $\delta \downarrow 0$ . Thus we arrive at (3.1).

Also observe that clearly

$$\int_{\mathbb{R}} f(x + \delta\xi) p_n(\xi) d\mu(\xi) = \int_{\mathbb{R}} P_{n,x,\delta}[f](\xi) p_n(\xi) d\mu(\xi).$$

Hence, we can rewrite (3.1) as

$$f^{(n)}(x) = \frac{k_n n!}{h_n} \lim_{\delta \downarrow 0} \frac{1}{\delta^n} \int_{\mathbb{R}} P_{n,x,\delta}[f](\xi) p_n(\xi) d\mu(\xi). \quad (3.16)$$

For the Legendre case (3.6) this interpretation by least-square approximation was given in [8]. But earlier, in 1990, Kopel & Schramm [41], apparently unaware of any predecessors, arrived at the case  $n = 1$  of (3.8) while they were guided by least-square approximation.

### 3.8 Even orthogonality measures

We can refine Theorem 3.1 if we assume that the orthogonality measure  $\mu$  considered there is *even*, i.e., that  $d\mu(-x) = d\mu(x)$ . Then the corresponding orthogonal polynomials  $p_n$  are even or odd according to whether  $n$  is even or odd, respectively. The simplest example is given by the Legendre polynomials. Now also modify the assumptions about  $f^{(n)}(x)$  in Theorem 3.1. Only assume that the right  $n$ -th derivative  $f_+^{(n)}(x)$  and left  $n$ -th derivative  $f_-^{(n)}(x)$  exist. Then by an easy adaptation of the proof of Theorem 3.1 we get (3.1) with on the left-hand side the symmetric  $n$ -th derivative of  $f$  at  $x$ :

$$\frac{1}{2} (f_+^{(n)}(x) + f_-^{(n)}(x)) = \frac{k_n n!}{h_n} \lim_{\delta \downarrow 0} \frac{1}{\delta^n} \int_{\mathbb{R}} f(x + \delta\xi) p_n(\xi) d\mu(\xi). \quad (3.17)$$

The special case of this result for Legendre polynomials (see (3.6), (3.7)) was observed in [25, Proposition 1] for  $n = 1$  and in [8, Theorem 2] for general  $n$ . The special case of (3.17) for centered Gram polynomials (see (3.4)) was observed in [9, Theorem 1.1].

Moreover, in [8, pp. 370–371] and [66, §4] examples were given for the Legendre case with  $n = 1$ , where the limit on the right-hand side of (3.7) exists, but the left and right derivative of  $f$  at  $x$  do not exist. Earlier, in [41, pp. 231–232] an example was given where the limit of the right-hand side of (3.8) for  $n = 1$  exists, while the right derivative of  $f$  at  $x$  does not exist.

Consider the proof of Theorem 3.1 if we know that  $f^{(k)}(x)$  exists for  $k$  up to some  $m > n$ . Then

$$\begin{aligned} \frac{1}{\delta^n} \int_{\mathbb{R}} f(x + \delta\xi) p_n(\xi) d\mu(\xi) &= \frac{h_n}{k_n n!} f^{(n)}(x) + \sum_{k=m+1}^n \frac{f^{(k)}(x) \delta^{k-n}}{k!} \int_{\mathbb{R}} \xi^k p_n(\xi) d\mu(\xi) \\ &+ \delta^{m-n} \int_{\mathbb{R}} F_{x,m}(\delta\xi) \xi^m p_n(\xi) d\mu(\xi) = \frac{h_n}{k_n n!} f^{(n)}(x) + \mathcal{O}(\delta) \quad \text{as } \delta \downarrow 0. \end{aligned}$$

We can say more if moreover  $\mu$  is an even measure. Then  $\int_{\mathbb{R}} \xi^{n+1} p_n(\xi) d\mu(\xi) = 0$  and thus we have, as  $\delta \downarrow 0$ ,

$$\frac{1}{\delta^n} \int_{\mathbb{R}} f(x + \delta\xi) p_n(\xi) d\mu(\xi) = \frac{h_n}{k_n n!} f^{(n)}(x) + \begin{cases} o(\delta) & \text{if } f^{(n+1)}(x) \text{ exists,} \\ \mathcal{O}(\delta^2) & \text{if } f^{(n+2)}(x) \text{ exists.} \end{cases} \quad (3.18)$$

The Legendre case of (3.18) was observed for  $n = 1$  in [42, (5-9.3)] and for general  $n$  in [54, (4)].

### 3.9 Generalized Taylor series

Rewrite (3.1) as

$$f^{(n)}(x) = \frac{k_n n!}{h_n} \lim_{\delta \downarrow 0} \frac{1}{\delta^n} \int_{\mathbb{R}} f(x + \xi) p_n(\delta^{-1}\xi) d\mu_{\delta}(\xi). \quad (3.19)$$

Here  $\mu_{\delta}(E) := \mu(\delta^{-1}E)$ . Then the polynomials  $x \mapsto p_n(\delta^{-1}x)$  are orthogonal with respect to the measure  $\mu_{\delta}$ . The formal Taylor series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!} \eta^n \quad (3.20)$$

can accordingly be seen as a termwise limit of the formal generalized Fourier series

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{1}{h_n} \left( \int_{\mathbb{R}} f(x + \xi) p_n(\delta^{-1}\xi) d\mu_{\delta}(\xi) \right) p_n(\eta) \\ = \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{k_n n!}{\delta^n h_n} \int_{\mathbb{R}} f(x + \xi) p_n(\delta^{-1}\xi) d\mu_{\delta}(\xi) \right) \frac{\delta^n p_n(\delta^{-1}\eta)}{k_n}. \end{aligned} \quad (3.21)$$

Indeed, use (3.19) and the limit  $\delta^n p_n(\delta^{-1}\eta)/k_n \rightarrow \eta^n$  as  $\delta \downarrow 0$ .

While for a big class of orthogonal polynomials and for  $f$  moderately smooth, the series (3.21) converges with sum  $f(x + \eta)$  because of equiconvergence theorems given in Szegő [64, Ch. 9 and 13], we will need analyticity of  $f$  in a neighbourhood of  $x$  for convergence of (3.20) to

$f(x + \eta)$ . In the case of Jacobi polynomials and for  $f$  analytic on a neighbourhood of  $x$  we can use Szegő [64, Theorem 9.1.1]. Then an open disk around  $x$  of radius less than the convergence radius of (3.20) is in the interior of the ellipse of convergence (3.21) for  $\delta$  small enough. This was discussed for Legendre polynomials by Fishback [20].

A different limit from orthogonal polynomials to monomials is discussed by Askey & Haimo [4]. This involves Gegenbauer polynomials, which we write as Jacobi polynomials  $P_n^{(\alpha, \alpha)}(x) = k_n^{(\alpha, \alpha)} x^n + \dots$ :

$$\lim_{\alpha \rightarrow \infty} \frac{P_n^{(\alpha, \alpha)}(x)}{k_n^{(\alpha, \alpha)}} = x^n. \quad (3.22)$$

Consider now the formal expansion of  $f(x + \eta)$  for  $\eta \in [-1, 1]$  in terms of the polynomials  $P_n^{(\alpha, \alpha)}(\eta)$ :

$$\begin{aligned} & \sum_{n=0}^{\infty} \frac{1}{h_n^{(\alpha, \alpha)}} \left( \int_{-1}^1 f(x + \xi) P_n^{(\alpha, \alpha)}(\xi) (1 - \xi^2)^\alpha d\xi \right) P_n^{(\alpha, \alpha)}(\eta) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{k_n^{(\alpha, \alpha)} n!}{h_n^{(\alpha, \alpha)}} \int_{-1}^1 f(x + \xi) P_n^{(\alpha, \alpha)}(\xi) (1 - \xi^2)^\alpha d\xi \right) \frac{P_n^{(\alpha, \alpha)}(\eta)}{k_n^{(\alpha, \alpha)}} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\int_{-1}^1 f^{(n)}(x + \xi) (1 - \xi^2)^{n+\alpha} d\xi}{\int_{-1}^1 (1 - \xi^2)^{n+\alpha} d\xi} \frac{P_n^{(\alpha, \alpha)}(\eta)}{k_n^{(\alpha, \alpha)}}, \end{aligned}$$

where we used the identity for  $\delta = 1$  in (3.10). The last form of the above formal series tends termwise to the formal Taylor series (3.20). This is seen from (3.22) and the fact that the measure  $(1 - \xi^2)^{n+\alpha} d\xi / \int_{-1}^1 (1 - \xi^2)^{n+\alpha} d\xi$  tends to the delta measure as  $\alpha \rightarrow \infty$  (see [4, p.301]). As observed in [4, p.303], the function  $f$  has to be increasingly smooth as  $\alpha$  grows in order to have convergence in the expansion of  $f(x + \eta)$  in terms of  $P_n^{(\alpha, \alpha)}(\eta)$ .

### 3.10 Connection with the continuous wavelet transform

The *continuous wavelet transform*  $\Phi_g$  (see for instance [14], [38]) is defined by

$$(\Phi_g f)(a, b) := |a|^{-\frac{1}{2}} \int_{\mathbb{R}} f(t) \overline{g(a^{-1}(t - b))} dt \quad (f \in L^2(\mathbb{R}), a, b \in \mathbb{R}, a \neq 0). \quad (3.23)$$

Here we will take the *wavelet*  $g$  as a nonzero function in  $(L^1 \cap L^2)(\mathbb{R})$  such that  $\int_{\mathbb{R}} g(t) dt = 0$ .

For orthogonal polynomials  $p_n(x)$  let the orthogonality measure have the form  $d\mu(x) = w(x) dx$  (essentially item 1 in §2.1) with  $w(x) \geq 0$  for  $x \in \mathbb{R}$  and with the functions  $x \mapsto x^n w(x)$  in  $(L^1 \cap L^2)(\mathbb{R})$  for all  $n = 1, 2, \dots$ . Put

$$g_n(x) := p_n(x) w(x). \quad (3.24)$$

Then, for  $n = 1, 2, \dots$  the functions  $g_n$  are in  $(L^1 \cap L^2)(\mathbb{R})$  and satisfy  $\int_{\mathbb{R}} g_n(t) dt = 0$ . Now consider the continuous wavelet transform (3.23) for  $g$  equal to such  $g_n$  and compare with (3.1).

Then

$$\int_{\mathbb{R}} f(x + \xi\delta) p_n(\xi) w(\xi) d\xi = \delta^{-1} \int_{\mathbb{R}} f(t) g_n(\delta^{-1}(t - x)) dt = \delta^{-\frac{1}{2}} (\Phi_{g_n} f)(\delta, x). \quad (3.25)$$

Hence, (3.1) can now be written as

$$f^{(n)}(x) = \frac{k_n n!}{h_n} \lim_{\delta \downarrow 0} \frac{1}{\delta^{n-\frac{1}{2}}} (\Phi_{g_n} f)(\delta, x). \quad (3.26)$$

A similar observation about the continuous wavelet transform approximating the  $n$ -th derivative was made by Rieder [55, (5)] in the case of (3.23) with  $g$  having its first  $n$  moments equal to zero. In fact, he recovers formula (3.11), first obtained by Haslam-Jones [29], for  $\nu$  absolutely continuous,  $f$ ,  $c_n$  being the  $n$ -th distributional derivative of  $f$ , and the limit taken in a suitable Sobolev norm (see also [55, Theorem 2.3]). An example of a wavelet  $g$  having its first  $n$  moments equal to zero is Daubechies' wavelet  $n\psi$  of compact support, see [13, p.984].

The continuous wavelet transform  $\Phi_{g_n}$  with  $g_n$  given by (3.24) and  $w(x)$  being a weight function for one of the classical orthogonal polynomials (Jacobi, Laguerre, Hermite) was considered by Moncayo & Yáñez [48].

### 3.11 A special case: the $n$ -th order finite difference as approximation of the $n$ -th derivative

For  $N = n + 1$  we can see that (3.3) specializes as an  $n$ -th order finite difference approximating the  $n$ -th derivative. Indeed, write (3.3) as

$$f^{(n)}(x) = \lim_{\delta \downarrow 0} (D_{n,\delta,N} f)(x),$$

where (see also (2.4) and [36, (9.5.1)])

$$\begin{aligned} (D_{n,\delta,N} f)(x) &:= \frac{(2n+1)!(N-n-1)!}{n!(N+n)!} \frac{1}{\delta^n} \sum_{\xi=0}^{N-1} f(x + \delta\xi) t_n(\xi, N) \\ &= \frac{(2n+1)!(N-1)!(-1)^n}{n!(N+n)!\delta^n} \sum_{\xi=0}^{N-1} f(x + \delta\xi) {}_3F_2\left(\begin{matrix} -n, n+1, -\xi \\ 1, -N+1 \end{matrix}; 1\right). \end{aligned}$$

Now put  $N := n + 1$  and use that for  $\xi = 0, 1, \dots, n$  we have by [50, (15.2.4)]:

$${}_3F_2\left(\begin{matrix} -n, n+1, -\xi \\ 1, -n \end{matrix}; 1\right) = \sum_{k=0}^{\xi} \frac{(-\xi)_k (n+1)_k}{k! k!} = {}_2F_1\left(\begin{matrix} -\xi, n+1 \\ 1 \end{matrix}; 1\right) = \frac{(-n)_\xi}{(1)_\xi} = (-1)^\xi \binom{n}{\xi}.$$

Hence

$$(D_{n,\delta,n+1} f)(x) = \delta^{-n} \sum_{\xi=0}^n (-1)^{n-\xi} \binom{n}{\xi} f(x + \delta\xi) = (\Delta_\delta^n f)(x),$$

where

$$(\Delta_\delta f)(x) := \frac{f(x + \delta) - f(x)}{\delta}.$$



## 4 Filters for higher derivatives

The following theorem is a multi-term variant of Theorem 3.1.

**Theorem 4.1.** *Let  $\{p_k\}_{k=0,1,2,\dots}$  be a system of orthogonal polynomials with respect to the orthogonality measure  $\mu$ . Let  $m, n$  be integers such that  $0 \leq m \leq n$ . Let  $x \in \mathbb{R}$ . Let  $I$  be a closed interval such that, for some  $\varepsilon > 0$ ,  $x + \delta\xi \in I$  if  $0 \leq \delta \leq \varepsilon$  and  $\xi \in \text{supp}(\mu)$ . Let  $f$  be a continuous function on  $I$  such that its derivatives of order  $1, 2, \dots, n$  at  $x$  exist. In addition, if  $I$  is unbounded, assume that  $f$  is of at most polynomial growth on  $I$ . Then*

$$\frac{1}{\delta^m} \sum_{j=m}^n \frac{1}{h_j} \left( \int_{\mathbb{R}} f(x + \delta\xi) p_j(\xi) d\mu(\xi) \right) p_j^{(m)}(0) = f^{(m)}(x) + o(\delta^{n-m}) \quad \text{as } \delta \downarrow 0. \quad (4.1)$$

If  $f \in C^n(I)$  and  $f^{(n)}$  is of polynomial growth on  $I$  then (4.1) holds uniformly for  $x$  in compact subsets of  $I$ .

Note that (4.1) turns down to (3.1) if  $m = n$ .

**Proof** With the notation (2.14) we can rewrite the left-hand side of (4.1) as

$$\delta^{-m} \left( \frac{\partial}{\partial \eta} \right)^m \int_{\mathbb{R}} f(x + \delta\xi) \mathbf{K}_n(\xi, \eta) d\mu(\xi) \Big|_{\eta=0}. \quad (4.2)$$

By substitution of (2.22) and by application of (2.17) the expression (4.2) is equal to

$$\begin{aligned} \sum_{l=0}^{n-m} \frac{f^{(m+l)}(x)}{l!} \delta^l \eta^l \Big|_{\eta=0} + \delta^{n-m} \left( \frac{\partial}{\partial \eta} \right)^m \int_{\mathbb{R}} \xi^n F_{x,n}(\delta\xi) \mathbf{K}_n(\xi, \eta) d\mu(\xi) \Big|_{\eta=0} \\ = f^{(m)}(x) + \delta^{n-m} \int_{\mathbb{R}} \xi^n F_{x,n}(\delta\xi) \left( \frac{\partial}{\partial \eta} \right)^m K_n(\xi, \eta) \Big|_{\eta=0} d\mu(\xi). \end{aligned}$$

Now use the same argument as at the end of the proof of Theorem 3.1, calling Proposition 2.3 and using dominated convergence, in order to show that

$$\lim_{\delta \downarrow 0} \int_{\mathbb{R}} \xi^n F_{x,n}(\delta\xi) \left( \frac{\partial}{\partial \eta} \right)^m \mathbf{K}_n(\xi, \eta) \Big|_{\eta=0} d\mu(\xi) = 0.$$

For the proof of the last statement also use Proposition 2.4. □

**Remark 4.2.** In view of (3.15) we can rewrite (4.1) as

$$f^{(m)}(x) = \frac{1}{\delta^m} \frac{d^m}{d\eta^m} P_{n,x,\delta}[f](\eta) \Big|_{\eta=0} + o(\delta^{n-m}) \quad \text{as } \delta \downarrow 0 \quad (m = 0, 1, \dots, n). \quad (4.3)$$

If  $f$  is a polynomial of degree  $\leq n$  then (4.3) holds exactly without the term  $o(\delta^{n-m})$  because

$$P_{n,x,\delta}[f](\eta) = f(x + \delta\eta).$$

For  $f$  having derivatives at  $x$  up to order  $n + 1$  we can refine (4.1) as follows.

**Proposition 4.3.** *Keep the assumptions of Theorem 4.1. Moreover assume that  $f^{(n+1)}(x)$  exists. Then*

$$\begin{aligned} & \frac{1}{\delta^m} \sum_{j=m}^n \frac{1}{h_j} \left( \int_{\mathbb{R}} f(x + \delta\xi) p_j(\xi) d\mu(\xi) \right) p_j^{(m)}(0) \\ &= f^{(m)}(x) - \frac{p_{n+1}^{(m)}(0) f^{(n+1)}(x)}{k_{n+1} (n+1)!} \delta^{n-m+1} + o(\delta^{n-m+1}) \quad \text{as } \delta \downarrow 0. \end{aligned} \quad (4.4)$$

If  $f \in C^{m+1}(I)$  and  $f^{(n+1)}$  is of polynomial growth on  $I$  then (4.4) holds uniformly for  $x$  in compact subsets of  $I$ .

**Proof** Write the left-hand side of (4.4) as (4.2). Then substitute (2.22) with  $n$  replaced by  $n + 1$  and apply (2.17). Then the expression (4.2) becomes

$$\begin{aligned} f^{(m)}(x) + \frac{f^{(n+1)}(x)}{(n+1)!} \delta^{n-m+1} \left( \frac{\partial}{\partial \eta} \right)^m \int_{\mathbb{R}} \xi^{n+1} \mathbf{K}_n(\xi, \eta) d\mu(\xi) \Big|_{\eta=0} \\ + \delta^{n-m+1} \int_{\mathbb{R}} \xi^{n+1} F_{x,n+1}(\delta\xi) \left( \frac{\partial}{\partial \eta} \right)^m \mathbf{K}_n(\xi, \eta) \Big|_{\eta=0} d\mu(\xi). \end{aligned}$$

By a similar argument as in the proof of Theorem 4.1 we see that the last term equals  $o(\delta^{n-m+1})$  as  $\delta \downarrow 0$ . By application of (2.17) the second term becomes

$$\begin{aligned} \frac{f^{(n+1)}(x)}{(n+1)!} \delta^{n-m+1} \left( \frac{\partial}{\partial \eta} \right)^m \int_{\mathbb{R}} \xi^{n+1} \left( \mathbf{K}_{n+1}(\xi, \eta) - \frac{p_{n+1}(\xi) p_{n+1}(\eta)}{h_{n+1}} \right) d\mu(\xi) \Big|_{\eta=0} \\ = - \frac{p_{n+1}^{(m)}(0)}{h_{n+1} (n+1)!} \delta^{n-m+1} \int_{\mathbb{R}} \xi^{n+1} p_{n+1}(\xi) d\mu(\xi). \end{aligned}$$

Then (4.4) follows by using (2.7). □

#### 4.1 Even orthogonality measures

If in Proposition 4.3 the orthogonality measure  $\mu$  is even and if  $n - m$  is odd then  $p_j^{(m)}(0) = 0$  whenever  $j - m$  is odd, so the sum on the left-hand side of (4.4) then runs over  $j = m, m + 2, \dots, n - 1$ . Moreover, for  $p_{n+1}^{(m)}(0)/k_{n+1}$  occurring on the right-hand side of (4.4) we then have that

$$(-1)^{(n-m+1)/2} \frac{p_{n+1}^{(m)}(0)}{k_{n+1}} > 0. \quad (4.5)$$

In order to prove (4.5) we can assume without loss of generality that  $k_n > 0$ . Then, for all  $j$  and for  $x$  large,  $p_{n+1}^{(j)}(x) > 0$ . Also note that  $p_{n+1}^{(j)}$  is a polynomial of degree  $n + 1 - j$  which is even or odd according to whether  $n + 1 - j$  is even or odd. Since  $p_{n+1}$  belongs to a family of

orthogonal polynomials, it has  $n + 1$  simple real zeros. The number of positive zeros is  $(n + 1)/2$  if  $n + 1$  is even and  $n/2$  if  $n + 1$  is odd. Now it follows by induction with respect to  $j$  that  $p_{n+1}^{(j)}$  has  $(n + 1 - j)/2$  positive zeros if  $n + 1 - j$  is even and  $(n - j)/2$  if  $n + 1 - j$  is odd. So we arrive also at this property for  $j = m$ , and then (4.5) readily follows.

Thus, for  $\mu$  an even measure and  $n - m$  an odd number we can rewrite (4.4) as follows for  $\delta \downarrow 0$ .

$$\begin{aligned} & \frac{1}{\delta^m} \sum_{j=m, m+2, \dots, n-1} \left( \int_{\mathbb{R}} f(x + \delta\xi) p_j(\xi) d\mu(\xi) \right) \frac{p_j^{(m)}(0)}{h_j} = f^{(m)}(x) \\ & - (-1)^{(n-m+1)/2} \frac{|p_{n+1}^{(m)}(0)| f^{(n+1)}(x)}{|k_{n+1}| (n+1)!} \delta^{n-m+1} + \begin{cases} o(\delta^{n-m+1}) & \text{if } f^{(n+1)}(x) \text{ exists,} \\ o(\delta^{n-m+2}) & \text{if } f^{(n+2)}(x) \text{ exists,} \\ \mathcal{O}(\delta^{n-m+3}) & \text{if } f^{(n+3)}(x) \text{ exists.} \end{cases} \quad (4.6) \end{aligned}$$

This is proved by a slight adaptation of the proof of Proposition 4.3. Furthermore, (4.6) holds uniformly for  $x$  in compact subsets of  $I$  if the appropriate derivative of  $f$  of order  $n + 1$ ,  $n + 2$  or  $n + 3$  is continuous and of polynomial growth on  $I$ .

**Remark 4.4.** Note that for fixed  $m$  the approximation of the left-hand side of (4.6) (and earlier formulas (4.1), (4.3) and (4.4)) to  $f^{(m)}(x)$  becomes better as  $n$  increases. However, this observation disregards the frequency spectrum of the signal  $f$  and the effect of noise. See Remark 4.5 for a discussion of these aspects.

## 4.2 Filters

We can consider the left-hand side of (4.1) as a filter (continuous or discrete depending on the choice of  $\mu$ ) for  $m$ -th order differentiation at  $x$ . In general, a continuous respectively analog filter sends an input function  $f$  to an output function  $g$  by convolution with a fixed real-valued function  $\rho$ :

$$g(y) = \sum_{x=M}^N f(y-x) \rho(x) \quad (y \in \mathbb{Z}), \quad (4.7)$$

$$g(y) = \int_M^N f(y-x) \rho(x) dx \quad (y \in \mathbb{R}). \quad (4.8)$$

Here  $M$  may be  $-\infty$  and  $N$  may be  $\infty$ . Filters are widely used in electrical engineering, with analog filters being continuous and digital filters being discrete. There  $y$  is usually the time variable  $t$  and instead of  $\rho$  one writes  $h$ , the *unit impulse response*. If  $M$  and  $N$  are finite in (4.12), one speaks about a *finite impulse response* (FIR) filter, otherwise about an *infinite impulse response* (IIR) filter. The Fourier transform  $\phi$  of  $\rho$  is called the *characteristic function*

or *transfer function* (also denoted by  $H$ ) of the filter:

$$\phi(\omega) := \sum_{x=-M}^N \rho(x) e^{-ix\omega}, \quad (4.9)$$

$$\phi(\omega) := \int_{-M}^N \rho(x) e^{-ix\omega} dx. \quad (4.10)$$

Equivalently,  $\phi$  equals the quotient  $g/f$  of the input function  $g$  and the output function  $f$  if  $f(y) := e^{i\omega y}$ . Standard books of digital filter theory are for instance [2] and [27]. In [2, p.306] methods are described for the design of digital differentiators.

### 4.3 The characteristic function

We continue with the left-hand side of (4.1) considered as a filter. Let us assume that  $\mu$  is an even measure and that  $n-m$  is odd, so that we can work with (4.6). We obtain the *characteristic function*  $\phi$  of the filter defined by the left-hand side of (4.6) if we put there  $f(\xi) := e^{i\omega\xi}$  with  $\omega \in \mathbb{R}$  and take  $x := 0$ :

$$\begin{aligned} \phi(\omega\delta) &:= \frac{1}{\delta^m} \sum_{j=m, m+2, \dots, n-1} \left( \int_{\mathbb{R}} e^{i\omega\delta\xi} p_j(\xi) d\mu(\xi) \right) \frac{p_j^{(m)}(0)}{h_j} \\ &= (i\omega)^m \left( 1 - \frac{|p_{n+1}^{(m)}(0)|}{|k_{n+1}|(n+1)!} (\delta\omega)^{n-m+1} + (\delta\omega)^{n-m+3} G(\delta\omega) \right), \end{aligned} \quad (4.11)$$

with  $G$  a bounded function on  $\mathbb{R}$  (for the proof of the second equality use Proposition 2.3 with  $f$  bounded).

**Remark 4.5.** From the last part of (4.11) we see that for differentiation with fixed order  $m$  the first term gives the characteristic function  $(i\omega)^m$  of the ideal differentiator. The second term has degree  $n+1$  in  $\omega$  and gives rise to a falling down of the characteristic function, since the coefficient has negative sign. So for high frequencies the filter will be a low pass filter and for low frequencies the filter works well for differentiation. Increase of  $n$  brings the filter in a sense closer to the ideal differentiator (see also Remark 4.4 for approximation to  $f^{(m)}(x)$  in the  $x$ -domain), but the pass band will also increase, causing more high frequency noise (see section 5.2 and Figure 1 for an example in the Legendre case). In the practice of the construction of a differentiating filter one has to decide to what frequency the differentiation must do the job and how much noise one accepts. This all depends on the frequency contents of the signal and the noise. See also the discussion for the case of constant weights by Barak [5, p.2761] (for  $m=0$ ) and by Luo et al. [44, §5] (for general  $m$ ).

### 4.4 Smoothing filters

For  $m=0$  the filters given by the left-hand sides of (4.1) and (4.6) are examples of *smoothing filters*. These have a very long history, see Schoenberg [58], [59] and references given there, which

go back as far as De Forest's work in 1878. We put  $M := -N$  in (4.7) and  $M = -N = -1$  in (4.8):

$$g(y) = \sum_{x=-N}^N f(y-x) \rho(x) \quad (y \in \mathbb{Z}), \quad (4.12)$$

$$g(y) = \int_{-1}^1 f(y-x) \rho(x) dx \quad (y \in \mathbb{R}), \quad (4.13)$$

and we usually take  $\rho$  symmetric:  $\rho(x) = \rho(-x)$ .

We say that (4.12) or (4.13) is *exact* for the degree  $j$  (where  $j < 2N$  in case of (4.12)) if  $g = f$  whenever  $f$  is a polynomial of degree  $\leq j$ , but  $g \neq f$  for some polynomial  $f$  of degree  $j + 1$ . Because of symmetry of  $\rho$ , such  $j$  will always be odd.

Exactness of (4.12) for degree at least  $2n + 1 < 2N$  can equivalently be stated as

$$\rho(x) = \sum_{k=0}^N c_k t_{2k}(x + N, 2N + 1) \quad (x \in \{-N, -N + 1, \dots, N\})$$

with

$$c_k = t_{2k}(N, 2N + 1)/h_{2k} \quad \text{if } k = 0, 1, \dots, n,$$

where  $y \mapsto t_{2k}(y, 2N + 1)$  is a Gram polynomial (see (2.4)) and  $h_{2k}$  is the corresponding constant given by (2.6). For such  $\rho$  the sum of squares  $\sum_{x=-N}^N \rho(x)^2$  is minimal if and only if  $c_k = 0$  for  $n < k \leq N$ . Then

$$\rho(x) = \mathbf{K}_{2n}(x + N, N) \quad (x \in \{-N, -N + 1, \dots, N\}), \quad (4.14)$$

where  $\mathbf{K}_{2n}$  is the Christoffel-Darboux kernel of degree  $2n$  (see (2.14)) for the orthogonal polynomials  $y \mapsto t_k(y, 2N + 1)$ .

Similarly, in case of (4.13) and assuming that  $\rho$  is a polynomial, the requirements that the formula is exact for the degree  $2n + 1$  and that  $\int_{-1}^1 \rho(x)^2 dx$  is minimal are equivalent to

$$\rho(x) = \mathbf{K}_{2n}(x, 0) \quad (x \in [-1, 1]), \quad (4.15)$$

where  $\mathbf{K}_{2n}$  is the Christoffel-Darboux kernel of degree  $2n$  for the Legendre polynomials  $P_k$ .

More generally than (4.14) we can work with orthogonal polynomials  $p_n$  satisfying (2.2) or (2.3) for equidistant points running over a symmetric set and with symmetric weights:

$$\sum_{x=-N}^N p_m(x) p_n(x) w(x) = 0 \quad (m \neq n),$$

where  $w(x) = w(-x)$ . In terms of these polynomials  $p_n$  exactness of (4.12) for degree at least  $2n + 1 < 2N$  can equivalently be stated as

$$\rho(x) = \sum_{k=0}^N c_k p_{2k}(x) w(x) \quad (x \in \{-N, -N + 1, \dots, N\}) \quad (4.16)$$

with

$$c_k = p_{2k}(0)/h_{2k} \quad \text{if } k = 0, 1, \dots, n. \quad (4.17)$$

In particular, the choice

$$\rho(x) := \mathbf{K}_{2n}(x, 0) w(x) = \frac{k_{2n} p_{2n}(0)}{k_{2n+1} h_{2n}} \frac{p_{2n+1}(x) w(x)}{x}, \quad (4.18)$$

with  $\mathbf{K}_{2n}$  the Christoffel-Darboux kernel and with (2.15) used for the second equality, will make (4.12) exact for the degree  $2n + 1$ .

The *characteristic function* for (4.12) respectively (4.13) is defined by (4.9) with  $M = -N$  respectively (4.10) with  $M = -N = -1$ . The condition that (4.12) or (4.13) is exact for the degree  $2n + 1$  is equivalent with the condition that  $\phi$  has power series of the form

$$\phi(\omega) = 1 - a\omega^{2n+2} + \dots \quad (4.19)$$

with  $a \neq 0$ .

An  $m$ -fold iteration of (4.12) (with  $N = \infty$  for convenience) yields

$$g_m(y) = \sum_{x=-\infty}^{\infty} f(y-x) \rho_m(x) \quad (y \in \mathbb{Z}),$$

where  $\rho_m = \rho * \dots * \rho$  is an  $m$ -fold convolution product. De Forest (1878) raised the question for which choices of  $\rho$  the asymptotic behaviour of  $\rho_m(x)$  for large  $m$  can be described. Schoenberg [58, Theorem 1 and Remark 1 on p.358] showed that this is possible precisely if the characteristic function satisfies

$$|\phi(\omega)| < 1 \quad \text{for } 0 < \omega < 2\pi. \quad (4.20)$$

If (4.20) is satisfied then the smoothing is called *stable*. Clearly (4.20) will imply that (4.19) holds with  $a > 0$ . For  $\rho$  given by (4.18) we see from (4.11) that  $a > 0$  is satisfied for any choice of the weights and that  $a$  is explicitly given by

$$a = \frac{|p_{2n+2}(0)|}{|k_{2n+2}|(2n+2)!}.$$

The stability condition (4.20) can also be considered for the continuous smoothing formula (4.13), where now  $\omega \in \mathbb{R} \setminus \{0\}$  in (4.20). For the Legendre case where  $\rho$  is given by (4.15), stability was proved by Trench [65] and Lorch & Szego [43].

## 4.5 Fourier-Bessel functions

As a common generalization of (4.12) and (4.13) with  $\rho$  given by (4.15) and (4.18), respectively, we can consider a smoothing formula

$$g(y) = \int_{\mathbb{R}} f(y-x) r(x) d\mu(x) \quad (4.21)$$

with  $\mu$  an even positive orthogonality measure for the orthogonal polynomials  $p_n$  and with  $r$  given by

$$r(x) := \mathbf{K}_{2n}(x, 0) = \sum_{j=0}^n \frac{p_{2j}(0) p_{2j}(x)}{h_{2j}} = \frac{k_{2n} p_{2n}(0)}{k_{2n+1} h_{2n}} \frac{p_{2n+1}(x)}{x}. \quad (4.22)$$

(Such usage of the Christoffel-Darboux formula was emphasized in [53] for the cases of Gram and Legendre polynomials.) Then we can define the corresponding characteristic function by

$$\begin{aligned} \phi(\omega) &:= \int_{\mathbb{R}} r(x) e^{-ix\omega} d\mu(x) \\ &= \sum_{j=0}^n \frac{p_{2j}(0)}{h_{2j}} \int_{\mathbb{R}} p_{2j}(x) e^{-ix\omega} d\mu(x) \\ &= \frac{k_{2n} p_{2n}(0)}{k_{2n+1} h_{2n}} \int_{\mathbb{R}} \frac{p_{2n+1}(x)}{x} e^{-ix\omega} d\mu(x). \end{aligned} \quad (4.23)$$

Thus

$$\phi'(\omega) = \frac{k_{2n} p_{2n}(0)}{i k_{2n+1} h_{2n}} \int_{\mathbb{R}} p_{2n+1}(x) e^{-ix\omega} d\mu(x). \quad (4.24)$$

In the Legendre case  $d\mu(x) = dx$  with support  $[-1, 1]$  we can evaluate integrals occurring above in terms of (spherical) Bessel functions as follows:

$$\int_{-1}^1 P_n(x) e^{-ix\omega} dx = i^{-n} \sqrt{\frac{2\pi}{\omega}} J_{n+\frac{1}{2}}(\omega) = 2i^{-n} j_n(\omega), \quad (4.25)$$

see [50, (18.17.19), (10.47.3)] or [6, (4)]. In the Chebyshev case  $d\mu(x) = (1-x^2)^{-1/2} dx$ ,  $T_n(\cos \theta) := \cos(n\theta)$  we similarly obtain (see [50, (10.9.2)])

$$\pi^{-1} \int_{-1}^1 T_n(x) e^{-ix\omega} dx = i^{-n} J_n(\omega). \quad (4.26)$$

Formula (4.26) was the reason for Mantica [45], [46] to call the functions

$$\mathcal{J}_n(\omega; \mu) := \int_{\mathbb{R}} p_n(x) e^{-ix\omega} d\mu(x) \quad (4.27)$$

*Fourier-Bessel functions* (where he took  $p_n$  orthonormal and  $\mu$  a probability measure). The same functions occur in Ignjatović [34] as the functions  $\mathcal{K}^n[m]$  (in the notation of [34, §2.1]). Formula (4.25) played an important role in the proof of the stability result in the Legendre case, see [43]. It also occurred in Rangarajan et al. [54, (14), (15)] for a formal operational calculus in connection with the right-hand side of (3.6) before taking limits. In the Appendix we will compute the Fourier-Bessel functions for the case of the shifted symmetric Hahn polynomials (4.28).

## 4.6 Stability of smoothing in case of symmetric Hahn and Krawtchouk polynomials

The shifted symmetric Hahn polynomials

$$p_n(x) := Q_n(N+x; \alpha, \alpha, 2N) \quad (n = 0, 1, \dots, 2N) \quad (4.28)$$

are orthogonal polynomials on  $\{-N, N+1, \dots, N\}$  with respect to the symmetric weights

$$w_x := \frac{(\alpha+1)_{N+x}}{(N+x)!} \frac{(\alpha+1)_{N-x}}{(N-x)!}, \quad (4.29)$$

see [36, (9.5.1) and (9.5.2)]. Assume that  $\alpha$  is a nonnegative integer. Consider in terms of these polynomials  $p_n$  formulas (4.16) and (4.17) characterizing exactness for degree at least  $2n+1$ . Now observe that, by [36, (9.5.9)], we have

$$\Delta_x^\alpha(w_x p_n(x)) = \frac{(2N+1)_\alpha}{\alpha!} Q_{n+\alpha}(N+x+\alpha; 0, 0, 2N+\alpha),$$

where  $\Delta_x(f(x)) = (\Delta f)(x) := f(x+1) - f(x)$ . It follows that

$$\sum_{x=-N-\alpha}^N ((\Delta^\alpha \rho)(x))^2 \quad (4.30)$$

is minimal for  $\rho$  given by (4.16) and (4.17) if and only if  $c_k = 0$  for  $n < k \leq N$ . Greville [24, §3] (1966) denotes (4.30) by  $R_\alpha^2$  (after division by  $\binom{2\alpha}{\alpha}$ ). Therefore he calls the smoothing formula (4.12) the *minimum  $R_\alpha$  formula* if  $\rho$  is taken such that (4.30) is minimal. Greville [24, (4.2)] gives an explicit formula for the characteristic function  $\phi$  in case of a minimum  $R_\alpha$  formula. He ascribes this formula to Sheppard [61] (1913). We will derive this formula in the Appendix. Greville [24, §5] next proves the stability property (4.20) for these cases. Curiously, Greville does not mention Hahn polynomials in any way. Hahn polynomials in this context seem to come up first in Bromba & Ziegler [7, §3.2].

As a special case of [36, (9.5.16)] there is the limit formula

$$\lim_{\alpha \rightarrow \infty} Q_n(x+N; \alpha, \alpha, 2N) = K_n(x+N; \frac{1}{2}, 2N) = {}_2F_1 \left( \begin{matrix} -n, -N-x \\ -2N \end{matrix}; 2 \right) \quad (4.31)$$

where the polynomials  $x \mapsto K_n(x; p, N)$  are *Krawtchouk polynomials* (see [36, §9.11]). The corresponding weights (4.29), suitably normalized, tend for  $\alpha \rightarrow \infty$  to the symmetric weights

$$w_x := \binom{2N}{N+x} \quad (x = -N, -N+1, \dots, N),$$

with respect to which the polynomials  $p_n(x) = K_n(x+N; \frac{1}{2}, 2N)$  are orthogonal. The smoothing formula with  $\rho$  given by (4.18) for this  $p_n$  and  $w$  is called the *minimum  $R_\infty$  formula* by Greville [24, §6]. He obtains the characteristic function  $\phi$  for this case explicitly as a limit case of his



formula in the minimum  $R_\alpha$  case (see also (A.9)), not working with Krawtchouk polynomials at all. (Krawtchouk polynomials seem to come up first in this context in Bromba & Ziegler [7, §3.3].) But Greville also obtains in some way that,

$$\phi^{(k)}(\pi) = 0 \quad (k = 0, 1, \dots, 2N - 2n - 1). \quad (4.32)$$

Let us prove this by observing from (4.31) that

$$p_{2N}(x) = {}_2F_1\left(\begin{matrix} -2N, -N - x \\ -2N \end{matrix}; 2\right) = \sum_{j=0}^{N+x} \binom{N+x}{j} (-2)^j = (-1)^{N+x}.$$

Hence

$$\phi^{(k)}(\pi) = \sum_{x=-N}^N \rho(x) (-ix)^k (-1)^x = i^{2N-k} \sum_{x=-N}^N \mathbf{K}_{2n}(x, 0) x^k p_{2N}(x) w_x = 0$$

for  $k < 2N - 2n$  by orthogonality. Now Greville concludes from (4.19) and (4.32) that

$$\phi(\omega) = 1 - (\sin^2(\omega/2))^{n+1} P(\sin^2(\omega/2)) = (\cos^2(\omega/2))^{N-n} Q(\sin^2(\omega/2)) \quad (4.33)$$

for certain polynomials  $P$  of degree  $N - n - 1$  and  $Q$  of degree  $n$ . From that he immediately derives that  $Q(z)$  is equal to the power series of  $(1 - z)^{-N+n}$  truncated after the term with  $z^n$ , i.e.,

$$Q(z) = \sum_{k=0}^n \frac{(N-n)_k}{k!} z^k > 0 \quad (0 \leq z \leq 1). \quad (4.34)$$

By a similar argument we see that

$$P(z) = \sum_{k=0}^{N-n-1} \frac{(n+1)_k}{k!} (1-z)^k > 0 \quad (0 \leq z \leq 1). \quad (4.35)$$

Hence, by (4.33),

$$0 \leq \phi(\omega) < 1 \quad (0 < \omega < 2\pi),$$

which is even stronger than the stability condition (4.20).

Identity (4.33) with  $Q$  and  $P$  given by (4.34), (4.35) has a long history which is surveyed in Koornwinder & Schlosser [39]. However, this paper missed Greville's paper and the connection with Krawtchouk polynomials. A sequel [40] to [39], tracing (4.33)–(4.35) back to 1713, is in preparation.

By a short chain of identities, using (4.33) and (4.34),  $\phi(\omega)$  can be expressed as

$$\phi(\omega) = (N-n) \binom{N}{n} \int_0^{\cos^2(\omega/2)} s^{N-n-1} (1-s)^n ds,$$

see [39, top of p.249]. Hence  $\phi$  is monotonically decreasing on  $[0, \pi]$  from 1 to 0. Such filters without ripples are called *maximally flat* by Herrmann [31] (see also Samadi and Nishihara [56]). Herrmann gave the same argument as above for solving (4.33), apparently unaware of Greville [24].

## 4.7 The Savitzky-Golay paper and its follow-up

The first instance of an approximation of first and higher derivatives by formula (4.1) with possibly  $n > m$  was given by Savitzky & Golay [57] in 1964. They only dealt with the case of constant weights on  $\{-N, -N + 1, \dots, N\}$ , they used only very special  $N$ ,  $n$  and  $m$ , and they did not explicitly mention or use the corresponding orthogonal polynomials. They were motivated by applications in spectroscopy. Their paper had an enormous impact, for instance 5432 citations in Google Scholar in January 2012. Some corrections to [57] were given by Steinier et al. [62] in 1972.

Probably, Gorry [22] (1990) was the first who gave (4.1) in a more structural form in the case of constant weights on an equidistant set using centered Gram polynomials. Next, in [23] (1991) he considered (4.1) on a finite non-equidistant set, still with constant weights. Meer & Weiss [47] (1992) gave (4.1) for orthogonal polynomials on a set  $\{-N, -N + 1, \dots, N\}$  with respect to general weights. They made this more explicit in the cases of centered Gram polynomials and of centered Krawtchouk polynomials with symmetric weights.

We recommend Luo et al. [44] as a relatively recent survey of follow-up to the Savitzky-Golay paper.

## 5 Filter properties in the frequency domain: some examples

In section 4 many so-called *linear* filters for derivatives were mentioned. In electrical engineering one uses *the transfer function*  $H$  for understanding the properties of the filter. This is the Fourier transform of the unit impulse response of the filter, see (4.9), (4.10) where we wrote  $\phi$  instead of  $H$ . In general, this function is complex-valued. We can show the properties of the filter in the frequency domain by a log-log plot of the modulus of the transfer function (which may be complemented by a phase plot). In general, for a differentiation filter of order  $n$ ,  $H(\omega)$  should behave for low frequencies like  $\omega^n$ , and for high frequencies like a constant (equal to zero in the ideal case). When the behaviour is different, the filter is called *unstable*.

### 5.1 The Lanczos derivative

The (analog) filter corresponding to the Lanczos derivative is given by (3.7) ignoring the limit:

$$g(x) = \frac{3}{2\delta} \int_{-1}^1 f(x + \xi\delta) \xi d\xi. \quad (5.1)$$

The output function  $g$  can be considered as a continuous (i.e. unsampled) approximation of the first derivative of the input function  $f$ . The transfer function  $H(\omega)$  is equal to the quotient of  $g$  and  $f$  with  $f(y) := e^{i\omega y}$ . A short computation gives

$$\begin{aligned} H(\omega) &= \frac{3}{2\delta} \int_{-1}^1 e^{i\delta\omega\xi} \xi d\xi = i\omega \frac{3}{(\delta\omega)^3} \left( \sin(\delta\omega) - \delta\omega \cos(\delta\omega) \right) \\ &= i\omega(1 + \mathcal{O}(\delta^2\omega^2)) \quad \text{as } \delta\omega \downarrow 0, \end{aligned} \quad (5.2)$$

compatible with (4.11) for  $m = 1, n = 2$ . For small  $\omega\delta$  we have  $\arg(H(\omega)) = \pi/2$ . The modulus of  $H(\omega)$  for  $\delta = 1$  is given in Figure 1, case  $n = 1$  as a log-log plot.

## 5.2 Multi-term variant of the Lanczos derivative

To get a better approximation we can use equation (4.1) with  $p_n$  a Legendre polynomial:

$$g(x) = \frac{1}{\delta^m} \sum_{k=0}^{[(n-m)/2]} \frac{P_{m+2k}^{(m)}(0)}{h_{m+2k}} \int_{-1}^1 f(x + \delta\xi) P_{m+2k}(\xi) d\xi. \quad (5.3)$$

Here

$$\frac{P_{m+2k}^{(m)}(0)}{h_{m+2k}} = \frac{2^m (-1)^k (\frac{1}{2})_{m+k} (m + 2k + \frac{1}{2})}{k!}$$

by [18, 10.10(26), 10.9(19)] and (2.10). For  $m = n = 1$  (5.3) reduces to (5.1). For the transfer function we obtain (see also (4.11)):

$$\begin{aligned} H_{m,n}(\omega) &= \frac{2^m}{\delta^m} \sum_{k=0}^{[(n-m)/2]} \frac{(-1)^k (\frac{1}{2})_{m+k} (m + 2k + \frac{1}{2})}{k!} \int_{-1}^1 e^{i\omega\delta\xi} P_{m+2k}(\xi) d\xi \\ &= \frac{2^{m+1} i^m}{\delta^m} \sum_{k=0}^{[(n-m)/2]} \frac{(\frac{1}{2})_{m+k} (m + 2k + \frac{1}{2})}{k!} j_{m+2k}(\delta\omega), \end{aligned}$$

where the spherical Bessel functions  $j_{m+2k}$  entered by (4.25). An explicit formula for spherical Bessel functions is given in [50, (10.49.2)]. In particular,

$$\begin{aligned} j_1(z) &= \frac{1}{z^2} (\sin z - z \cos z), \\ j_3(z) &= \frac{1}{z^4} ((15 - 6z^2) \sin z - (15 - z^2) z \cos z). \end{aligned}$$

Thus  $H_{1,1}(\omega)$  is given by (5.2). After some computation we obtain

$$H_{1,3}(\omega) = \frac{15i\omega}{2} \frac{(21 - 8(\delta\omega)^2) \sin(\delta\omega) + (-21\delta\omega + (\delta\omega)^3) \cos(\delta\omega)}{(\delta\omega)^5}. \quad (5.4)$$

The modulus of  $H_{1,3}(\omega)$  for  $\delta = 1$  is also given in figure 1, case  $n = 3$ . It is clear that for  $n = 3$  the plot stays close to a straight line until higher values of  $\omega$  than for  $n = 1$ .

## 5.3 First order Savitzky-Golay filter

When the input signal of the filter is given by a vector of (for convenience) odd dimension  $2N + 1$  obtained by sampling a function  $f$  on equidistant points  $x - N\delta, x - (N - 1)\delta, \dots, x + N\delta$ , then

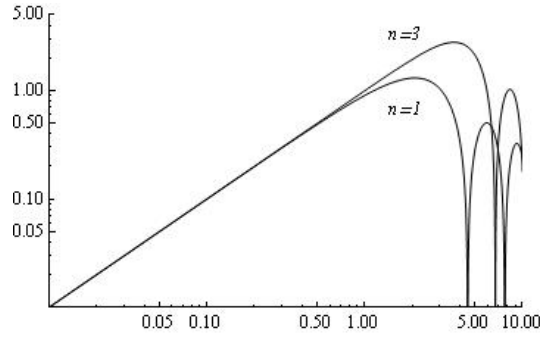


Figure 1: Modulus of transfer function for the first order analog filter,  $n = 1$  and 3

we may use (3.4) ignoring the limit as a discrete filter for the  $n$ -th derivative of  $f$  at  $x$ . In particular, for  $n = 1$ , we can use (3.5):

$$g(x) = \frac{3}{2N(N + \frac{1}{2})(N + 1)\delta} \sum_{\xi=-N}^N f(x + \delta\xi) \xi.$$

For the transfer function

$$H(\omega) = \frac{3}{2N(N + \frac{1}{2})(N + 1)\delta} \sum_{\xi=-N}^N e^{i\delta\omega\xi} \xi \quad (5.5)$$

we obtain by straightforward computation that

$$H(\omega) = \frac{3i}{2(2N + 1)\delta} \left( \sin(\frac{1}{2}\delta\omega) \right)^{-2} \left( \frac{\sin(N\delta\omega)}{N} - \frac{\sin((N + 1)\delta\omega)}{N + 1} \right). \quad (5.6)$$

Note that the phase shift is exactly  $\pi/2$ .

The modulus of  $H(\omega)$  for  $\delta = 1$  and for  $N = 1$  and 2 is given in Figure 2. See [44, Figure 1] for similar pictures.

#### 5.4 Butterworth filter

If one needs a filter that does differentiation for low frequencies very well and has a good suppression for high frequencies then there are better filters than the ones discussed in this paper until here. For example there are the so-called Tchebyshev, inverse Tchebyshev, Elliptic, Butterworth and Bessel filters. These filters all differentiate, but the choice of the most suitable filter depends on the properties one needs, for instance a constant phase response, a good amplitude response, less side-lobes etc. We mention here the so-called  $n$ -th order *Butterworth filter*. The square of the modulus of the transfer function of an  $n$ -th order analog Butterworth filter that differentiates with order  $m$  is given by

$$\left| H_{m,n}(\omega) \right|^2 = \frac{\omega^{2m}}{1 + (\omega/\omega_0)^{2n}} = \omega^{2m} |H_{0,n}(\omega)|^2 \quad (n > m). \quad (5.7)$$

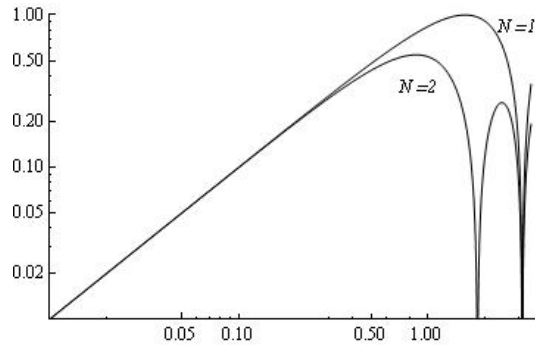


Figure 2: Modulus of transfer function for the first order discrete filter,  $N = 1$  and  $2$

Here  $\omega_0$  is the so-called *cutoff frequency*. It is at this frequency  $\omega_0$  where the the asymptotics of the low frequency part and the high frequency part of the transfer function meet. The factor  $\omega^{2m}$  is the square of the modulus of the transfer function  $(i\omega)^m$  of the ideal  $m$ -th order differentiator.

As an example see Figure 3 showing the transfer function of a seventh order Butterworth filter with  $\omega_0 = 1$  (see how the side lobes differ from those of Figure 2).

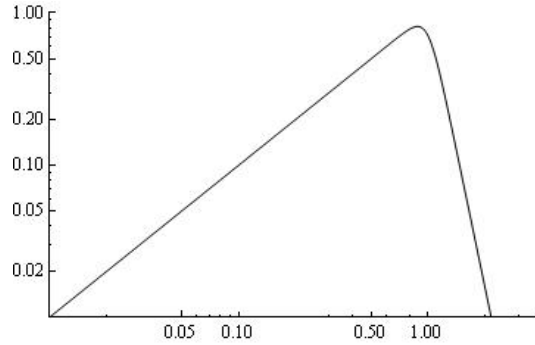


Figure 3: Transfer function for seventh order Butterworth filter

In (5.7) one has to make a choice of  $H_{m,n}(\omega)$  as follows:

$$H_{m,n}(\omega) = \frac{(i\omega)^m}{p_n(i\omega)} \quad \text{with } |p_n(i\omega)|^2 = 1 + (\omega/\omega_0)^{2n}$$

such that  $p_n$  is a polynomial of degree  $n$  with real coefficients for which all (possibly complex) roots have negative real part. Then (4.10) and (4.8) take the form

$$H_{0,n}(\omega) = \int_0^\infty \rho(t) e^{-i\omega t} dt, \quad g(t) = \int_0^\infty f(t - \tau) \rho(\tau) d\tau.$$

$\rho(t)$  is called the *impulse response* of the filter. It follows that the output function  $g$  satisfies a differential equation with the input function  $f$  as inhomogeneous part:

$$p_n(d/dt) g(t) = f(t).$$

For instance, for  $n = 1$  we have

$$p_1(i\omega) = 1 + i\omega/\omega_0, \quad \rho(t) = \omega_0 e^{-\omega_0 t}, \quad \omega_0^{-1} g'(t) + g(t) = f(t).$$

For  $n = 2$  we have

$$p_2(i\omega) = 1 + 2^{1/2} (i\omega/\omega_0) + (i\omega/\omega_0)^2, \quad \rho(t) = 2^{1/2} \omega_0 e^{-2^{-1/2}\omega_0 t} \sin(2^{-1/2}\omega_0 t).$$

One can obtain the transfer function for the Butterworth filter in the digital case from  $H_{0,n}(\omega)$  in the analog case by so-called frequency warping: replace  $\omega$  by  $2T^{-1} \tan(\omega T/2)$ , where  $T$  is the length of the sampling interval. Then some linear combination of finitely many output values  $g(x), g(x-T), \dots$  will be equal to some linear combination of finitely many input values  $f(x), f(x-T), \dots$  (a so-called recursive filter).

There are important differences for practical applications between filters obtained from orthogonal polynomials, as amply considered in this paper, and the Butterworth filter. In the analog case the Butterworth filter can be much easier constructed physically. But in the discrete case the filters obtained from orthogonal polynomials are much easier to handle in the time domain than the Butterworth filter.

For more details about the Butterworth filter for  $m = 0$  see Oppenheim & Schaffer [51, §5.2.1] (both analog and digital), Johnson [35, §3.2] (analog) and Hamming [27, §12.6] (digital). See Herely & Vetterli [30, §IV-A] and Cohen & Daubechies [12, §6.2] for Butterworth filters in connection with wavelets.

## A Appendix

Here we derive an explicit expression for the Fourier-Bessel functions (see (4.27)) associated with the shifted symmetric Hahn polynomials (see (4.28)). First observe that by the duality between Hahn polynomials and dual Hahn polynomials (see the formula after (9.6.16) in [36]) the generating function [36, (9.6.12)] for dual Hahn polynomials can be rewritten in terms of Hahn polynomials:

$$(1-t)^n {}_2F_1 \left( \begin{matrix} n-N, n+\alpha+1 \\ -\beta-N \end{matrix}; t \right) = \frac{N!}{(\beta+1)_N} \sum_{x=0}^N w_x Q_n(x; \alpha, \beta, N) t^x, \quad (\text{A.1})$$

where

$$w_x := \binom{\alpha+x}{x} \binom{\beta+N-x}{N-x}, \quad (\text{A.2})$$

i.e., the weight occurring in the orthogonality relation [36, (9.5.2)] for Hahn polynomials.

Next, in (A.1) take  $\beta = \alpha$ ,  $t := e^{-i\theta}$ , replace  $N$  by  $2N$ , shift  $x$  to  $x+N$  and apply Pfaff's identity [1, (2.3.14)]. Then

$$\frac{(2\alpha+2)_{2N+n}}{2^{2n}(2N)! (\alpha + \frac{3}{2})_n} e^{iN\theta} (1-e^{-i\theta})^n {}_2F_1 \left( \begin{matrix} n-2N, n+\alpha+1 \\ 2n+2\alpha+2 \end{matrix}; 1-e^{-i\theta} \right) = \sum_{x=-N}^N w_x p_n(x) e^{-ix\theta}, \quad (\text{A.3})$$

where  $p_n(x)$  and  $w_x$  are given by (4.28) and (4.29), respectively.

Now use the quadratic transformation [17, 2.11(30)] and the expression [18, 10.9(20)] of Gegenbauer polynomials in terms of hypergeometric functions:

$$\begin{aligned} {}_2F_1\left(\begin{matrix} n-2N, n+\alpha+1 \\ 2n+2\alpha+2 \end{matrix}; 1-e^{-i\theta}\right) &= e^{i(\frac{1}{2}n-N)\theta} {}_2F_1\left(\begin{matrix} n-2N, 2N+n+2\alpha+2 \\ n+\alpha+\frac{3}{2} \end{matrix}; \frac{1}{2}(1-\cos(\frac{1}{2}\theta))\right) \\ &= \frac{(2N-n)!}{(2n+2\alpha+2)_{2N-n}} e^{i(\frac{1}{2}n-N)\theta} C_{2N-n}^{n+\alpha+1}(\cos(\frac{1}{2}\theta)). \end{aligned}$$

Thus we can rewrite (A.3) as

$$\frac{(\alpha+1)_n}{(-2N)_n} i^{-n} (2\sin(\frac{1}{2}\theta))^n C_{2N-n}^{n+\alpha+1}(\cos(\frac{1}{2}\theta)) = \sum_{x=-N}^N w_x p_n(x) e^{-ix\theta}. \quad (\text{A.4})$$

The left-hand side of (A.4) gives an expression for the Fourier-Bessel function (4.27) associated with the shifted symmetric Hahn polynomials (4.28).

Now let  $\phi$  be defined by (4.23), (4.22) with  $p_n$  given by (4.28). Then combination of (4.24), (A.4) and [18, 10.9(22)] yields that

$$\frac{d\phi(\omega)}{d(\sin^2(\frac{1}{2}\omega))} = C (\sin^2(\frac{1}{2}\omega))^n {}_2F_1\left(\begin{matrix} -N+n+1, N+n+\alpha+2 \\ 2n+\alpha+\frac{5}{2} \end{matrix}; \sin^2(\frac{1}{2}\omega)\right). \quad (\text{A.5})$$

where

$$\begin{aligned} C &= (-1)^{n+1} 2^{2n+1} \frac{k_{2n} p_{2n}(0)}{k_{2n+1} h_{2n}} \frac{(\alpha+1)_{2n+1}}{(-2N)_{2n+1}} \frac{(4n+2\alpha+4)_{2N-2n-1}}{(2N-2n-1)!} \\ &= (-1)^n \frac{(N+\alpha+1)_{n+1} (-N)_{n+1}}{(n+\alpha+\frac{3}{2})_{n+1} n!}. \end{aligned} \quad (\text{A.6})$$

In the last equality we used [36, (9.5.2), (9.5.4)] for  $h_n$  and  $k_n$  and [15, (2.4)] together with [36, (9.5.3)] for getting

$$p_{2n}(0) = \frac{(\frac{1}{2})_n (N+\alpha+1)_n}{(-N+\frac{1}{2})_n (\alpha+1)_n}. \quad (\text{A.7})$$

Formulas (A.5), (A.6) coincide with formulas (4.3), (4.4) in Greville [24] if we replace  $N, n, \alpha$  by  $n, k, m$ , respectively. Integration of (A.6) together with  $\phi(0) = 1$  yields

$$\phi(\omega) = 1 + \frac{(-1)^n}{n!} \sum_{k=n+1}^N \frac{(N+\alpha+1)_k (-N)_k}{(n+\alpha+\frac{3}{2})_k (k-n-1)!} \frac{(\sin^2(\frac{1}{2}\omega))^k}{k}. \quad (\text{A.8})$$

Formula (A.8) coincides with formula (4.2) in Greville [24], which Greville (in his earlier form (4.1)) ascribes to Sheppard [61]. However, we have only been able to find a match of the special case  $m = 0$  of [24, (4.1)] with a formula in Sheppard's paper, namely with [61, (64)].

In the limit for  $\alpha \rightarrow \infty$  formula (A.8) becomes

$$\phi(\omega) = 1 + \frac{(-1)^n}{n!} \sum_{k=n+1}^N \frac{(-N)_k}{(k-n-1)!} \frac{(\sin^2(\frac{1}{2}\omega))^k}{k}. \quad (\text{A.9})$$

This coincides with the formula after (6.1) in Greville [24], and also with (4.33) combined with (4.35) and [1, (2.3.15)].

If  $\alpha \in \mathbb{Z}_{\geq 0}$  then the left-hand side of (A.4) can be written as a finite sum, where the number of terms is independent of  $N$ . First observe that by [50, (14.13.1), (14.3.21), (5.5.5)] we have

$$C_n^\lambda(\cos \theta) = \frac{\Gamma(2\lambda + 1)}{2^{2\lambda}\Gamma(\lambda + 1)^2} \frac{(2\lambda)_n}{(\lambda + 1)_n} (\sin \theta)^{1-2\lambda} \sum_{k=0}^{\infty} \frac{(1-\lambda)_k (n+1)_k}{(n+\lambda+1)_k k!} \sin((2k+n+1)\theta) \quad (\lambda > 0, 0 < \theta < \pi). \quad (\text{A.10})$$

For  $\lambda \in \mathbb{Z}_{>0}$  the above series terminates after the term with  $k = \lambda - 1$ . Hence, for  $\alpha \in \mathbb{Z}_{\geq 0}$  (A.4) takes the form

$$\begin{aligned} & -2i^{n+2\alpha} \binom{2N+\alpha}{\alpha} (2\sin(\frac{1}{2}\theta))^{-n-2\alpha-1} \sum_{k=0}^{n+\alpha} (-1)^k \binom{n+\alpha}{k} \frac{(-2N-2\alpha-n-1)_k}{(-2N-\alpha)_k} \\ & \times \sin(\frac{1}{2}(2N+2\alpha+n-2k+1)\theta) = \sum_{x=-N}^N w_x p_n(x) e^{ix\theta}. \quad (\text{A.11}) \end{aligned}$$

For  $\alpha = 0$  and  $n = 1$  formula (A.11) specializes to (5.6) with  $H(\omega)$  given by (5.5). Use that  $Q_1(N+x; 0, 0, 2N) = x/N$ .

## References

- [1] G. E. Andrews, R. Askey and R. Roy, *Special Functions*, Cambridge University Press, 1999.
- [2] A. Antonio, *Digital Filters*, McGraw-Hill, second ed., 1993.
- [3] T. Apostol, *Calculus, Vol. 1*, Wiley, second ed., 1967.
- [4] R. Askey and D. T. Haimo, *Similarities between Fourier and power series*, Amer. Math. Monthly 103 (1996), 297–304.
- [5] Ph. Barak, *Smoothing and differentiation by an adaptive-degree polynomial filter*, Anal. Chem. 67 (1995), 2758–2762.
- [6] J. A. Barcelo and A. Córdoba, *Band-limited functions:  $L^p$ -convergence*, Trans. Amer. Math. Soc. 313 (1989), 655–669.



- [7] M. U. A. Bromba and H. Ziegler, *On Hilbert space design of least-weighted-squares digital filters*, Internat. J. Circuit Theory Appl. 11 (1983), 7–32.
- [8] N. Burch, P. E. Fishback and R. Gordon, *The least-squares property of the Lanczos derivative*, Math. Mag. 78 (2005), 368–378.
- [9] N. Burch and P. E. Fishback, *Orthogonal polynomials and regression-based symmetric derivatives*, Real Anal. Exchange 32 (2007), 597–607.
- [10] P. L. Chebychev, *On the interpolation of equidistant values* (in Russian), Mém. Acad. St. Pétersb. 25 (1875); translated in French in *Oeuvres de P. L. Chebychev* (A. Markoff and N. Sonin, eds., St. Petersburg, 1899/1907), Vol. 2, pp. 217–242; reprinted Chelsea, N.Y., 1962.
- [11] N. Cioranescu, *La généralisation de la première formule de la moyenne*, Enseign. Math. 37 (1938), 292–302.
- [12] A. Cohen and I. Daubechies, *A new technique to estimate the regularity of refinable functions*, Rev. Mat. Iberoamericana 12 (1996), 527–591.
- [13] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math. 41 (1988), 909–996.
- [14] I. Daubechies, *Ten lectures on wavelets*, SIAM, 1992.
- [15] E. Diekema and T. H. Koornwinder, *Generalizations of an integral for Legendre polynomials by Persson and Strang*, J. Math. Anal. Appl. 388 (2012), 125–135; [arXiv:1005.2285v2](https://arxiv.org/abs/1005.2285v2) [math.CA].
- [16] J. Dutka, *On Gauss' priority in the discovery of the method of least squares*, Arch. Hist. Exact Sci. 49 (1996), 355–370.
- [17] A. Erdélyi, *Higher transcendental functions, Vol. I*, McGraw-Hill, 1953.
- [18] A. Erdélyi, *Higher transcendental functions, Vol. II*, McGraw-Hill, 1953.
- [19] A. Fischer, *Differentiability of Peano derivatives*, Proc. Amer. Math. Soc. 136 (2008), 1779–1785.
- [20] P. Fishback, *Taylor series are limits of Legendre expansions*, Missouri J. Math. Sci. 19 (2007), 29–34.
- [21] R. A. Gordon, *Peano differentiation via integration*, Real Anal. Exchange 34 (2009), 147–156.
- [22] P. A. Gorry, *General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method*, Anal. Chem. 62 (1990), 570–573.

- [23] P. A. Gorry, *General least-squares smoothing and differentiation of nonuniformly spaced data by the convolution method*, Anal. Chem. 63 (1991), 534–536.
- [24] T. N. E. Greville, *On stability of linear smoothing formulas*, SIAM J. Numer. Anal. 3 (1966), 157–170.
- [25] C. W. Groetsch, *Lanczos’ generalized derivative*, Amer. Math. Monthly 105 (1998), 320–326.
- [26] R. W. Hamming, *Numerical methods for scientists and engineers*, second ed., Dover Publications, 1986.
- [27] R. W. Hamming, *Digital filters*, third ed., Dover Publications, 1989.
- [28] G. H. Hardy, *A course of pure mathematics*, Cambridge University Press, tenth ed., 1952.
- [29] U. S. Haslam-Jones, *On a generalized derivative*, Quart. J. Math, Oxford Ser. (2) 4 (1953), 190–197.
- [30] C. Herley and M. Vetterli, *Wavelets and recursive filter banks*, IEEE Trans. Signal Process. 41 (1993), 2536–2556.
- [31] O. Herrmann, *On the approximation problem in nonrecursive digital filter design*, IEEE Trans. Circuit Theory 18 (1971), 411–413.
- [32] D. L. Hicks and L. M. Liebrock, *Lanczos’ generalized derivative: insights and applications*, Appl. Math. Comput. 112 (2000), 63–73
- [33] F. B. Hildebrand, *Introduction to numerical analysis*, Dover Publications, 1974.
- [34] A. Ignjatović, *Local approximations based on orthogonal differential operators*, J. Fourier Anal. Appl. 13 (2007), 309–330.
- [35] D. E. Johnson, *Introduction to filter theory*, Prentice-Hall, 1976.
- [36] R. Koekoek, P. A. Lesky and R. F. Swarttouw, *Hypergeometric orthogonal polynomials and their  $q$ -analogues*, Springer-Verlag, 2010.
- [37] T. H. Koornwinder, *Orthogonal polynomials with weight function  $(1-x)^\alpha(1+x)^\beta + M\delta(x+1) + N\delta(x-1)$* , Canad. Math. Bull. 27 (1984), 205–214.
- [38] T. H. Koornwinder, *The continuous wavelet transform*, in *Wavelets: an elementary treatment of theory and applications*, World Scientific, 1993, pp. 27–48.
- [39] T. H. Koornwinder and M. J. Schlosser, *On an identity by Chaundy and Bullard. I*, Indag. Math. (N.S.) 19 (2008), 239–261.
- [40] T. H. Koornwinder and M. J. Schlosser, *On an identity by Chaundy and Bullard. II*, in preparation.

- [41] D. Kopel and M. Schramm, *A new extension of the derivative*, Amer. Math. Monthly 97 (1990), 230–233.
- [42] C. Lanczos, *Applied analysis*, Prentice-Hall, 1956.
- [43] L. Lorch and P. Szego, *A Bessel function inequality connected with stability of least square smoothing. II*, Glasgow Math. J. 9 (1968), 119–122.
- [44] J. Luo, K. Ying. P. He and J. Bai, *Properties of Savitzky-Golay digital differentiators*, Digital Signal Processing 15 (2005), 122–136.
- [45] G. Mantica, *Generalized Bessel functions: theoretical relevance, and computational techniques*, in *Self-similar systems*, Joint Inst. Nuclear Res., Dubna, 1999, pp. 306–315.
- [46] G. Mantica, *Fourier-Bessel functions of singular continuous measures and their many asymptotics*, Electron. Trans. Numer. Anal. 25 (2006), 409–430.
- [47] P. Meer and I. Weiss, *Smoothed differentiation filters for images*, J. Visual Comm. Image Repr. 3 (1992), 58–72.
- [48] M. Moncayo and R. J. Yáñez, *Continuous wavelet transforms based on classical orthogonal polynomials and functions of the second kind*, J. Comput. Anal. Appl. 9 (2007), 207–220.
- [49] A. F. Nikiforov, S. K. Suslov and V. B. Uvarov, *Classical orthogonal polynomials of a discrete variable*, Springer-Verlag, 1991.
- [50] *NIST Handbook of Mathematical Functions*, Cambridge University Press, 2010; <http://dlmf.nist.gov>.
- [51] A. V. Oppenheim and R. W. Schaffer, *Digital signal processing*, Prentice Hall, 1975.
- [52] G. Peano, *Sulla formula di Taylor*, Torino Atti 27 (1891), 40–46.
- [53] P.-E. Persson and G. Strang, *Smoothing by Savitzky-Golay and Legendre filters*, in *Mathematical systems theory in biology, communications, computation, and finance*, IMA Vol. Math. Appl. 134, 2003, pp. 301–316.
- [54] S. K. Rangarajan and S. P. Purushothaman, *Lanczos’ generalized derivative for higher orders*, J. Comput. Appl. Math. 177 (2005), 461–465.
- [55] A. Rieder, *The high frequency behaviour of continuous wavelet transforms*, Appl. Anal. 52 (1994), 125–141.
- [56] S. Samadi and A. Nishihara, *The world of flatness*, IEEE Circuits Systems Magazine 7 (2007), 38–44.
- [57] A. Savitzky and M. J. E. Golay, *Smoothing and differentiation of data by simplified least squares procedures*, Anal. Chem. 36, 1964, 1627–1639.

- [58] I. J. Schoenberg, *Some analytical aspects of the problem of smoothing*, in *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, Interscience, 1948. pp. 351–370.
- [59] I. J. Schoenberg, *On smoothing operations and their generating functions*, Bull. Amer. Math. Soc. 59 (1953), 199–230.
- [60] J. Shen, *On the generalized “Lanczos’ generalized derivative”*, Amer. Math. Monthly 106 (1999), 766–768.
- [61] W. F. Sheppard, *Reduction of errors by means of negligible differences*, Proceedings International Congress of Mathematicians, Cambridge, Vol. 2, 1912, 348–384.
- [62] J. Steinier, Y. Termonia and J. Deltour, *Smoothing and differentiation of data by simplified least square procedure*, Anal. Chem. 44 (1972), 1906–1909.
- [63] F. Strik, *Ophthalmodynamographie und Ophthalmodynamometrie in der neurologischen Praxis*, Dissertation, Erasmus Universiteit Rotterdam, 1977.
- [64] G. Szegő, *Orthogonal polynomials*, Amer. Math. Soc. Colloquium Publications 23, Amer. Math. Soc., Fourth ed., 1975.
- [65] W. F. Trench, *On the stability of midpoint smoothing with Legendre polynomials*, Proc. Amer. Math. Soc. 18 (1967), 191–199.
- [66] L. Washburn, *The Lanczos’ derivative*, Senior project, Whitman College, 2006;  
<http://www.whitman.edu/mathematics/SeniorProjectArchive/2006/washbuea.pdf>.

E. Diekema, P.O. Box 2439, 7302 EP Apeldoorn, The Netherlands;  
 email: [e.diekema@gmail.com](mailto:e.diekema@gmail.com)

T. H. Koornwinder, Korteweg-de Vries Institute, University of Amsterdam,  
 P.O. Box 94248, 1090 GE Amsterdam, The Netherlands;  
 email: [T.H.Koornwinder@uva.nl](mailto:T.H.Koornwinder@uva.nl)