Optimal free parameters in orthonormal approximations

Optimal Free Parameters in Orthonormal Approximations

Albertus C. den Brinker and Harm J. W. Belt, *Member, IEEE*

***Abstract*— We consider orthonormal expansions where the basis functions are governed by some free parameters. If the basis functions adhere to a certain differential or difference equation, then an expression can be given for a specific enforced convergence rate criterion as well as an upper bound for the quadratic truncation error. This expression is a function of the free parameters and some simple signal measurements. Restrictions on the differential or difference equation that make this possible are given. Minimization of either the upper bound or the enforced convergence criterion as a function of the free parameters yields the same optimal parameters, which are of a simple form. This method is applied to several continuous-time and discrete-time orthonormal expansions that are all related to classical orthogonal polynomials.**

***Index Terms*— Orthogonal functions, parameter estimation, polynomials, transforms.**

I. Introduction

APPROXIMATIONS of functions by a set of orthonormal functions is an often applied technique. Examples are the use of orthogonal polynomials [1], [2]; the use of Laguerre, Hermite, and Kautz functions in system identification ([3] and references therein), and signal coding [4]–[6].

In many cases, the set of orthogonal functions depends on one or more free parameters. In that case, it is of interest to establish the optimal free parameters in the case of a limited number of expansion terms. Since the free parameters appear in the error criterion in a nonlinear way, they are usually hard to find: Many local minima may occur in the error surface. Further, one would like to establish these parameters before actually engaging the orthonormal expansion.

Therefore, the question is can we, and if we can how, can we establish these free parameters based on (preferably some simple) signal measurements only. Such an approach has been considered before [7] but there it was restricted to orthonormal functions adhering to a linear second-order differential or difference equation. In this paper, we extend these results and explicitly give the restrictions that have to be imposed on the differential or difference equation.

The specific form of the differential or difference equation results in a simple upper bound for the quadratic truncation error as well as an expression for an enforced convergence rate criterion. Both expressions are in essence identical and are a function of the free parameters and some signal measurements only. These expressions can be minimized by taking the derivative with respect to the free parameters and setting this equal to zero. This yields an expression for the free parameters directly in terms of a limited number of signal measurements.

This method is applied to orthonormal functions related to classical orthonormal polynomials where differential or difference equations and weight function adhere to the restrictions mentioned earlier. It extends results described earlier [8]–[11].

II. Considered Sets of Orthonormal Functions

Let $w(t)$ be a real-valued non-negative weight function over some finite or infinite continuous interval or a discrete interval, which are both denoted by I. The inner product of two real-valued functions $f(t), g(t)$ is denoted by

$$\langle f, g \rangle = \int_I w(t) f(t) g(t)\, dt \tag{1}$$

for continuous-time functions or

$$\langle f, g \rangle = \sum_I w(t) f(t) g(t) \tag{2}$$

for discrete-time functions. The continuous interval I is restricted to the three fundamental cases $(-1, 1), (0, \infty)$, or $(-\infty, \infty)$.

Let $\phi_n(\theta; t)$ $(n = 0, 1, \cdots)$ be a complete set of orthonormal functions in the space of real-valued functions in $L_2(I, w)$ or $\ell_2(I, w)$. The orthonormality is expressed by

$$\langle \phi_m, \phi_n \rangle = \delta_{m,n} \tag{3}$$

where $\delta_{m,n}$ denotes the Kronecker delta. It is assumed that this set of orthonormal functions depends on one or more parameters denoted by θ, where the parameter set θ consists of continuous real variables and where $\phi_n(\theta; t)$ is a continuous function of these free parameters. Last, it is required that the weight function be independent of the parameters θ.

Let the orthonormal functions ϕ_n adhere to the equation

$$\mathcal{L}_\theta \phi_n(\theta; t) = \lambda(n) \phi_n(\theta; t) \tag{4}$$

where $\mathcal{L}_\theta$ is a linear operator defined as

$$\mathcal{L}_\theta = \sum_{m=0}^{M} A_m(\theta, t) \frac{d^m}{dt^m} \tag{5}$$

Manuscript received February 28 1997; revised February 5, 1998. The associate editor coordinating the review of this paper and approving it for publication was Dr. Sawasd Tantaratana.

A. C. den Brinker is with the Faculteit der Elektrotechniek, Technische Universiteit Eindhoven, Eindhoven, The Netherlands.

H. J. W. Belt is with the Philips Research Laboratories, Eindhoven, The Netherlands.

Publisher Item Identifier S 1053-587X(98)05216-7.

for continuous-time functions or

$$\mathcal{L}_\theta = \sum_{m=0}^{M} A_m(\theta, t) z^{m-\tau} \quad (6)$$

for discrete-time functions with $\tau < M$ and z^m the forward shift operator: $z^m\{f\}(t) = f(t+m)$. Further, $\lambda(n)$ is a monotonically increasing function of n independent of θ and t. Without loss of generality, we can take $\lambda(0) = 0$ since if $\lambda(0) \neq 0$, then this term can be absorbed in the operator $\mathcal{L}_\theta$. Last, we require that each coefficient $A_m(\theta, t)$ is a finite sum of separable functions in t and the free parameters, i.e.,

$$A_m(\theta, t) = \sum_{k=0}^{K_m} B_{m,k}(\theta) C_{m,k}(t). \quad (7)$$

III. Squared-Error Criterion: An Upper Bound

For a real-valued function f in $L_2(I, w)$ or $\ell_2(I, w)$, we have

$$f(t) = \sum_{n=0}^{\infty} a_n(\theta)\phi_n(\theta; t). \quad (8)$$

From the orthonormality, we have $a_n = \langle f, \phi_n \rangle$, and a_n is a function of the free parameters θ.

In practical situations, we are interested in finite approximations $f_N(t)$ of a function f according to

$$f_N(t) = \sum_{n=0}^{N-1} a_n(\theta)\phi_n(\theta; t). \quad (9)$$

The relative quadratic error in this approximation is given by

$$q_N(\theta) = \sum_{n=N}^{\infty} |a_n(\theta)|^2 / \langle f, f \rangle. \quad (10)$$

For $q_N(\theta)$, the upper bound

$$q_N(\theta) \leq F(\theta)/\lambda(N) \quad (11)$$

can be found [7], where

$$F(\theta) = \langle f, \mathcal{L}_\theta f \rangle / \|f\|^2 = \sum_{n=0}^{\infty} \lambda(n) a_n^2(\theta) / \|f\|^2. \quad (12)$$

The upper bound in (11) is attained at a certain $\theta = \theta_1$, i.e., $q_N(\theta_1) = F(\theta_1)/\lambda(N)$, only for the functions $f(t) = K_0\phi_0(\theta_1; t) + K_N\phi_N(\theta_1; t)$, where K_0 and K_N are arbitrary real numbers [7].

In view of the assumed separability of the coefficients A_m, the function $F(\theta)$ is dependent on a limited number of signal measurements, namely

$$M_{m,k} = \langle f, C_{m,k}\, df^m/dt^m \rangle / \langle f, f \rangle \quad (13)$$

or

$$M_{m,k} = \langle f, C_{m,k} z^{m-\tau}\{f\} \rangle / \langle f, f \rangle, \quad (14)$$

for $k = 0, 1, \cdots, K_m$ and $m = 0, 1, \cdots, M$. Since it was assumed that the weight function is independent of the free parameters, the signal measurements $M_{m,k}$ are independent of the parameter set θ; hence, the name signal measurements is appropriate. These signal measurements have been called "moments" because $C_{m,k}$ is often a polynomial in t.

Consider the class of functions $\mathcal{C}_F$ having equal "moments" as defined by (13) or (14). Thus, there is a single function $F(\theta)$ applicable to this whole class, i.e.,

$$F(\theta) = \sum_{m=0}^{M} \sum_{k=0}^{K_m} B_{m,k}(\theta) M_{m,k}. \quad (15)$$

Having only these moments at hand, it is clear that the best estimate for an appropriate choice of the free parameters in an orthonormal series expansion is obtained by taking the derivative of $F(\theta)$ and setting this equal to zero. This leads to

$$\frac{\partial F(\theta)}{\partial \theta_j} = \sum_{m=0}^{M} \sum_{k=0}^{K_m} \frac{\partial B_{m,k}(\theta)}{\partial \theta_j} M_{m,k} = 0 \quad (16)$$

for each of the parameters θ_j of θ. The optimal parameter set $\hat{\theta}$ is a solution of this set of equations. From a mathematical point of view, it is not even clear if there exists a solution, if there is a single global minimum of F, and if this set of equations yields an explicit expression for the optimal parameters. We refrain from studying the constraints on the functions $B_{m,k}$. Instead, we argue that from the problem formulation, we do expect that an optimal parameter set $\hat{\theta}$ exists. Furthermore, the examples show that indeed simple expressions evolve for the optimal parameters (see Section V).

Another question that arises is the following: Having minimized $F(\theta)$ and thus the upper bound for $q_N(\theta)$, is there is a function f_b within the class $\mathcal{C}_f$ for which the error criterion q_N assumes the upper bound in the minimum at $\hat{\theta}$? The answer is, in general, negative. Such a function should not only have the right "moments" but, as stated before, must also be of the form $f_b(t) = K_0\phi_0(\hat{\theta}; t) + K_N\phi_N(\hat{\theta}; t)$. This latter form has only two parameters (K_0 and K_N), which is, in general, clearly insufficient to attain the (arbitrary) "moments" of $\mathcal{C}_F$.

We note that the statement that the procedure outlined above yields the best parameters for the whole class of functions having certain moments is rather weak. In practice, we are not dealing with a whole class of functions but rather with one specific function or a limited set of functions (not even necessarily having the same moments). Further, these moments are usually not the only available information. In fact, we did actually set out to find an approximation of the form (9); clearly, therefore, we dispose over more information. Fortunately, there is another interpretation of this method.

IV. Enforced Convergence Rate Criterion

Consider the approximation of functions by some linear combination of orthonormal functions as in (9). Further, consider the criterion

$$\zeta(\theta) = \sum_{n=0}^{\infty} \Lambda(n) a_n^2(\theta) / \|f\|^2. \quad (17)$$

Taking $\Lambda(n)$ as a monotonically increasing function of n, minimization of ζ can be considered to be an enforced convergence rate criterion.

From the previous section, it is immediately clear that by taking $\Lambda(n) = \lambda(n)$, we have

$$\zeta(\theta) = F(\theta) \tag{18}$$

and, thus, that optimization of this criterion with respect to θ yields the same result as the minimization of the upper bound $F(\theta)/\lambda(N)$ for the previously considered criterion $q_N(\theta)$.

In the examples (Section V), $\lambda(n)$ is of a very simple form, namely, $\lambda(n) = n$. This specific form has a clear physical interpretation since then, $F(\theta)$ is the center of the energy distribution in the coefficients of infinite orthonormal series, and this is what we are minimizing. In that case, $F(\theta)$ can be used as a rule of thumb for the minimum number of terms, say $N_{\min}$, that is needed to acquire some accuracy in the approximation. We could use, say, $N_{\min} = 2F(\theta)$ as a rule of thumb.

Consider the functions $f(t) = K\phi_M(\theta_0; t)$, where M is not necessarily equal to the number of terms in the approximation N. It is clear that for $F(\theta)$ at $\theta = \theta_0$, we have $F(\theta_0) = \lambda(M)$. Using a Taylor series for $F(\theta)$, it can be shown that the derivative of F with respect to θ at $\theta = \theta_0$ equals zero (see Appendix). Therefore, the parameter θ_0 that determines $f(t)$ is indeed one of the solutions of (16). Thus, if $M < N$, then the outlined procedure results in an approximation for which we have $f_N(t) = f(t)$.

This does not hold more generally. Suppose we have a function that for some $\theta = \theta_0$ can be written exactly as an N-terms orthonormal series, i.e., the function is within the model set for some value of θ. Then, the procedure will, in general, not return this value θ_0, and thus, the obtained "optimal" value of θ is not the best in terms of minimization of $q_N(\theta)$. This is not surprising since the "optimal" value of θ is not defined as a minimization of $q_N(\theta)$ but as a minimization of $\zeta(\theta)$ [or, equivalently, as a minimization of an upper bound for $q_N(\theta)$]. In practice, we do not consider this a serious problem since it is not expected that f is within the model set, and further, it is still possible that an accurate approximation is found might this situation arise.

Note that $F(\theta)$ is a linear function of the "moments" $M_{m,k}$; see (15). Therefore, an extension [10] of this method is possible to an enforced convergence rate criterion for several functions at once by introducing a relative weight to each function that is to be approximated. Let

$$\overline{M}_{m,k} = \sum_{j=1}^{J} v_j M_{m,k}^{(j)} \tag{19}$$

where $M_{m,k}^{(j)}$ is the m,kth measurement of the jth function $(j = 1, \cdots, J)$, and v_j is a positive weight that quantifies the importance of the jth function. Then, the optimal parameter set $\hat{\theta}$ is again a solution of the set of equations (16) but now with the signal measurements $M_{m,k}$ of an individual function replaced by the average signal measurements $\overline{M}_{m,k}$.

TABLE I
CLASSICAL ORTHOGONAL POLYNOMIALS; WEIGHT FUNCTION AND COEFFICIENTS IN THE DIFFERENTIAL EQUATION

	Jacobi	Laguerre	Hermite
weight	$(1-t)^{\mu}(1+t)^{\nu}$	$t^{\alpha}e^{-t}$	e^{-t^2}
β	$(1-t^2)$	t	1
γ	$\nu-\mu-(\mu+\nu)t$	$\alpha-t$	$-2t$
η_n	$-n(n+1+\mu+\nu)$	$-n$	$-2n$

V. ORTHONORMAL FUNCTIONS ASSOCIATED WITH ORTHOGONAL POLYNOMIALS

In this section, we will apply the outlined optimization procedure to transformations related to the classical orthogonal polynomials.

A. Classical Orthogonal Polynomials

The classical orthogonal polynomials are the Jacobi, the (generalized) Laguerre, and the Hermite polynomials [1], [2]. The Jacobi polynomials have as special cases the Legendre, the Chebyshev, and the Gegenbauer polynomials, and these will not be treated here. The orthogonality interval is $(-1, 1), (0, \infty)$, and $(-\infty, \infty)$ for the Jacobi, the generalized Laguerre, and the Hermite polynomials, respectively.

The classical orthogonal polynomials adhere to a second-order linear differential equation

$$\beta y'' + (\beta' + \gamma)y' - \eta_n y = 0 \tag{20}$$

where $'$ denotes differentiation. The coefficients β, γ, and η_n are given in Table I. The free parameters μ, ν, and α must all be greater than -1. The polynomials are orthogonal under a weight function $w(t)$; see Table I, for which

$$w'/w = \gamma/\beta. \tag{21}$$

holds.

Functions $\psi_n(t)$, which are obtained from the nth-order orthogonal polynomials $Q_n(t)$ by

$$\psi_n(t) = \sqrt{w(t)}\, Q_n(t)/h_n, \tag{22}$$

where h_n is the normalization constant, are considered. In this way, a set of functions that are orthonormal under a unit weight function is obtained. Thus, the weight function is independent of the free parameters involved in the orthogonal polynomials.

From (20)–(22), it easily follows that ψ_n adheres to a second-order linear differential equation, namely

$$D_2 \frac{d^2\psi_n}{dt^2} + D_1 \frac{d\psi_n}{dt} + D_0\psi_n = 0 \tag{23}$$

with $D_2 = \beta, D_1 = \beta'$, and $D_0 = -\gamma^2/(4\beta) - \gamma'/2 - \eta_n$.

In the case of an infinite interval (Hermite polynomials), we can use scaling and translation of the independent variable t. Thus, we obtain what we will call the Hermite functions

$$\phi_n(\theta; t) = \sqrt{\sigma}\ \psi_n(\sigma(t - t_0)) \tag{24}$$

where θ denotes both the scale σ $(\sigma > 0)$ and the center t_0. In the case of a semi-infinite interval, we define the Laguerre

TABLE II
PARAMETERS IN THE DIFFERENTIAL EQUATION FOR THE LAGUERRE AND HERMITE FUNCTIONS

	Laguerre	Hermite
A_2	$-t/\sigma$	$-1/(2\sigma^2)$
A_1	$-1/\sigma$	0
A_0	$[(\alpha-\sigma t)^2-2\sigma t]/(4\sigma t)$	$[\sigma^2(t-t_0)^2-1]/2$
$\lambda(n)$	n	n

TABLE III
LAGUERRE AND HERMITE FUNCTIONS: OPTIMAL FREE PARAMETERS AND THE MINIMUM VALUE OF F ALL EXPRESSED IN SIGNAL MEASUREMENTS ONLY

	Laguerre	Hermite
parameters	$\hat{\alpha}=\frac{2m_0}{m_{-1}}\sqrt{\frac{m_{-1}\mu_1}{m_{-1}m_1-m_0^2}}$	$\hat{t}_0=\frac{m_1}{m_0}$
	$\hat{\sigma}=2\sqrt{\frac{m_{-1}\mu_1}{m_{-1}m_1-m_0^2}}$	$\hat{\sigma}=\sqrt[4]{\frac{m_0\mu_0}{m_0m_2-m_1^2}}$
$F(\hat{\theta})$	$\sqrt{\frac{\mu_1}{m_{-1}}}\sqrt{\frac{m_{-1}}{m_0}\frac{m_1}{m_0}-1}-\frac{1}{2}$	$\sqrt{\frac{\mu_0}{m_0}}\sqrt{\frac{m_2}{m_0}-\left(\frac{m_1}{m_0}\right)^2}-\frac{1}{2}$

functions by scaling the independent variable t in the functions ϕ_n according to

$$\phi_n(\theta;t)=\sqrt{\sigma}\,\psi_n(\sigma t). \tag{25}$$

In this case, θ stands for the scale σ ($\sigma>0$) and for the order of generalization α. For the Jacobi case, we define the Jacobi functions directly as $\phi_n(\theta;t)=\psi_n(t)$. Then, θ denotes the parameters μ and ν.

We are now able to give the differential equation for these functions $\phi_n(\theta;t)$ in the form (4). The results are given in Table II, with the exception of the Jacobi functions. These cannot be given in a form where $\lambda(n)$ is independent of both parameters μ and ν.

We define the following signal measurements or "moments"

$$m_i=\langle f,g_i\rangle, \qquad \text{with } g_i(t)=t^if(t),$$
$$\mu_i=\langle f',\tilde{g}_i\rangle, \qquad \text{with } \tilde{g}_i(t)=t^if'(t),$$

and assume that these exist for those i that are used to determine the free parameters of the orthogonal functions. Note that these signal measurements are strictly positive numbers for $t\in(0,\infty)$ (Laguerre case) and for even i for $t\in(-\infty,\infty)$ (Hermite case). The definition of the moments m_i and μ_i is a bit of an abuse of the formerly outlined procedure since there, we worked with normalized moments (i.e., relative to the energy $\langle f,f\rangle$) and used a completely different indexing. The reason for introducing these "moments" is that it results in simpler expressions.

Taking the derivative of F with respect to the free parameters and setting this equal to zero gives the explicit expressions given in Table III. Note that it can be easily proved that $m_{-1}m_1-m_0^2$ (Laguerre case) and $m_0m_2-m_1^2$ (Hermite case) are strictly positive numbers. Parts of these results have been established before; see [8], [10], and [11].

TABLE IV
DISCRETE CLASSICAL ORTHOGONAL POLYNOMIALS: PARAMETERS, WEIGHT FUNCTION, COEFFICIENTS IN THE DIFFERENCE EQUATION, AND NORMALIZATION CONSTANTS. FOR THE KRAWTCHOUK CASE WE HAVE $q=1-p$. FURTHER, $(b)_n$ DENOTES THE POCHHAMMER SYMBOL

	Krawtchouk	Charlier	Meixner
parameters	p	a	b, c
weight	$\binom{L}{t}p^tq^{L-t}$	$\frac{e^{-a}a^t}{t!}$	$\frac{c^t(b)_t}{t!}$
$X(t)$	$p(L+1-t)$	a	$c(t+b-1)$
$Y(t)$	qt	t	t
η_n	n	n	$n(1-c)$
h_n^2	$\binom{L}{n}p^nq^n$	1	$\frac{n!(b)_n}{c^n(1-c)^b}$

A similar procedure can be applied if one of the parameters is fixed. Then, an expression for the other (free) parameter can be obtained. This latter situation is not explicitly given here.

B. Discrete Classical Orthogonal Polynomials

The discrete classical orthogonal polynomials are the Krawtchouk, Hahn, Charlier, and Meixner polynomials. The discrete Chebyshev polynomials are a special cases of the Hahn polynomials. Similar to the Jacobi polynomials, the Hahn polynomials will, in general, not give rise to the required form of a difference equation. Therefore, we will not treat this case here. The discrete orthogonality interval I is $0,1,\cdots,L$ for the Krawtchouk polynomials and the non-negative integers for the Charlier and Meixner polynomials. The parameters appearing in the weight functions (see Table IV) are restricted in the following way. The parameters p (Krawtchouk) and c (Meixner) are restricted to the interval (0, 1). The parameters a (Charlier) and b (Meixner) are positive real numbers.

The nth-order discrete classical orthogonal polynomial $Q_n(t)$ adheres to a second-order linear difference equation

$$X(t+1)Q_n(t+1)+[\eta_n-X(t+1)-Y(t)]Q_n(t)+Y(t)Q_n(t-1)=0. \tag{26}$$

The coefficients X, Y, and η_n are given in Table IV. The polynomials are orthogonal under the weight function $w(t)$ (see Table IV), and thus

$$\sum_t w(t)Q_n(t)Q_m(t)=h_n^2\delta_{n,m} \tag{27}$$

where h_n^2 is the normalization constant (see Table IV). Further, for the weight function, we have

$$w(t)/w(t-1)=X(t)/Y(t). \tag{28}$$

Functions $\phi_n(\theta;t)$, which are obtained from the nth-order orthogonal polynomials $Q_n(t)$ by

$$\psi_n(t)=\sqrt{w(t)}\,Q_n(t)/h_n \tag{29}$$

or

$$\psi_n(t)=(-1)^t\sqrt{w(t)}\,Q_n(t)/h_n \tag{30}$$

are considered. In this way, functions that are orthonormal under a unit weight function are obtained.

From (26), (28), (29), and (30), it easily follows that ψ_n adheres to a second-order difference equation, namely

$$D_2\psi_n(t+1) + D_1\psi_n(t) + D_0\psi_n(t-1) = 0 \tag{31}$$

where $D_2 = \sqrt{X(t+1)Y(t+1)}, D_1 = \pm[\eta_n - X(t+1) - Y(t)]$, and $D_0 = \sqrt{X(t)Y(t)}$. For D_1, the $+$ sign holds for (29) and the $-$ sign for (30).

We will now introduce the Krawtchouk, Charlier, and Meixner functions in such a way that both cases (29) and (30) are incorporated and, further, that the basis functions ϕ_n are continuous functions of the free parameters. In the following definitions, it is assumed that the leading term in the orthogonal polynomial has a positive coefficient.

The *Krawtchouk functions* are defined as

$$\phi_n(\rho;t) = \binom{L}{n}^{-1/2}\binom{L}{t}^{1/2}\rho^{t-n}(1-\rho^2)^{(L-t-n)/2}\kappa_n(\rho^2;t) \tag{32}$$

where $\rho^2 = p$ is the parameter determining the Krawtchouk polynomials κ_n. Thus, the Krawtchouk functions exist for $0 < |\rho| < 1$. Although the Krawtchouk polynomials are not defined for $\rho^2 = 0$, we have $\lim_{\rho\to 0}\ \phi_n(\rho;t) = \delta(t-n)$. Therefore, we allow $\rho = 0$ as a natural extension and, consequently, the parameter ρ in the Krawtchouk functions is limited to $-1 < \rho < 1$.

The *Charlier functions* are defined as

$$\phi_n(\alpha;t) = [\mathrm{sgn}(\alpha)]^n\ \frac{e^{-\alpha^2/2}\alpha^t}{\sqrt{t!}}\ \mathcal{C}_n(\alpha^2;t) \tag{33}$$

where $\alpha^2 = a$ is the parameter determining the Charlier polynomials $\mathcal{C}_n$, and $\mathrm{sgn}(\cdot)$ denotes the signum function. Thus, the Charlier functions exist for $\alpha \neq 0$. Although the Charlier polynomials are not defined for $\alpha^2 = 0$, we have $\lim_{\alpha\to 0}\ \phi_n(\alpha;t) = \delta(t-n)$. Therefore, $\alpha = 0$ is allowed as a natural extension, and consequently, the parameter α in the Charlier functions can assume any real value.

The *Meixner functions* are defined as

$$\phi_n(b,\xi;t) = (-1)^n\ \frac{\xi^{n+t}(1-\xi^2)^{b/2}\sqrt{(b)_t}}{\sqrt{n!t!(b)_n}}\ \mathcal{M}_n(b,\xi^2;t) \tag{34}$$

where b and $\xi^2 = c$ are the parameters determining the Meixner polynomials $\mathcal{M}_n$. The Meixner functions exist for $0 < |\xi| < 1$. Again, although the Meixner polynomials are not defined for $\xi^2 = 0$, we have $\lim_{\xi\to 0}\ \phi_n(b,\xi;t) = \delta(t-n)$. Thus, we allow $\xi = 0$ as well, and the parameter ξ in the Meixner functions is restricted to $-1 < \xi < 1$.

The definitions of the Krawtchouk, Charlier, and Meixner functions do include the standard basis $\phi_n(t) = \delta(n-t)$. Another nice feature of the given definitions is the following. The transformation pair $f \leftrightarrow a$ has the property that

$$a_n = \langle f, \phi_n(\theta;\cdot)\rangle \text{ and } f(t) = \langle a, \phi_t(\tilde{\theta};\cdot)\rangle \tag{35}$$

where θ is ρ, α, and (b,ξ), and $\tilde{\theta}$ is $-\rho, -\alpha$, and $(b,-\xi)$ for the Krawtchouk, Charlier, and Meixner transformation, respectively. Thus, the forward and backward transformations are identical apart from a sign in the free parameter vector. This can be proved [12] from the similarity of the recurrence and difference equations for the orthonormal functions defined in (32)–(34).

TABLE V
PARAMETERS IN THE DIFFERENCE EQUATION FOR THE KRAWTCHOUK, CHARLIER, AND MEIXNER FUNCTIONS. FOR THE KRAWTCHOUK CASE WE USE $r = [1-\rho^2]^{1/2}$ AS A SHORTHAND NOTATION

	Krawtchouk	Charlier	Meixner
A_2	$-\rho r\sqrt{(t+1)(L-t)}$	$-\alpha\sqrt{t+1}$	$\dfrac{-\xi\sqrt{(t+b)(t+1)}}{(1-\xi^2)}$
A_1	$\rho^2(L-t)+r^2t$	α^2+t	$\dfrac{b\xi^2+t(1+\xi^2)}{(1-\xi^2)}$
A_0	$-\rho r\sqrt{t(L-t+1)}$	$-\alpha\sqrt{t}$	$\dfrac{-\xi\sqrt{(t+b-1)t}}{(1-\xi^2)}$
$\lambda(n)$	n	n	n

Having the definitions, we can derive the differential equation for these functions $\phi_n(\theta;t)$ in the form (4) and (6) with $\tau = 1$. The results are given in Table V. We see that only for the Meixner functions, the separability (7) of the coefficients A_m does not hold with respect to the parameter b. Therefore, the outlined method is not applicable for optimization over this parameter. In the remainder, b is assumed to be a constant.

We define the following signal measurements or "moments"

$$m_i = \langle f, g_i\rangle$$
$$\mu = \langle f, \tilde{g}\rangle$$

with $g_i(t) = t^i f(t)$ for $i = 0, 1$, and

$$\tilde{g}(t) = \begin{cases} f(t+1)\ \sqrt{(t+1)(L-t)}, \\ f(t+1)\ \sqrt{t+1}, \\ f(t+1)\ \sqrt{(t+1)(t+b)} \end{cases}$$

for the Krawtchouk, Charlier, and Meixner case, respectively.

We are now able to construct the functions $F(\theta)$ for these cases [12]. Subsequently, by taking the derivative of F with respect to the parameter, we obtain the following results for the optimal parameter $\hat{\theta}$.

Krawtchouk Functions:

$$\hat{\rho} = \mathrm{sgn}(\mu)\sqrt{\frac{1}{2}}\sqrt{1 - \frac{c}{\sqrt{4\mu^2+c^2}}} \tag{36}$$

where $c = Lm_0 - 2m_1$. The measurement μ quantifies how alternating subsequent samples of the function f are around $t = L/2$. Therefore, the sign of μ determines the sign of $\hat{\rho}$. For the minimum of F, we have

$$F(\hat{\rho}) = L/2 - \frac{1}{m_0}\sqrt{\mu^2 + c^2/4}. \tag{37}$$

From the previous expression, it follows that for the center of the energy distribution, we have $0 \leq F(\hat{\rho}) \leq L/2$. This is achieved by allowing both positive and negative values for ρ.

Charlier Functions:

$$\hat{\alpha} = \mu / m_0 \tag{38}$$

with, at this optimum

$$F(\hat{\alpha}) = \frac{m_0 m_1 - \mu^2}{m_0^2}. \tag{39}$$

Meixner Functions (Fixed Parameter b):

$$\hat{\xi} = \beta - \operatorname{sgn}(\beta)\sqrt{\beta^2 - 1} \tag{40}$$

where $\beta = (m_1 + b m_0/2)/\mu$. For $b = 1$, the Meixner functions are called the discrete Laguerre functions. Restricting ourselves to nonnegative μ (thus, $\beta \geq 0$), the results for $b = 1$ are in accordance with those reported earlier [7], [9]. Further, for this minimum, we have

$$F(b, \hat{\xi}) = \frac{|\mu|}{m_0}\sqrt{\beta^2 - 1} - \frac{b}{2}. \tag{41}$$

VI. Discussion

We considered two criteria to establish the optimal parameters in series expansions using sets of orthogonal functions associated with classical orthogonal polynomials. It was shown that both criteria yield the same expression for the free parameters.

However, the starting point in both cases is different. In the first case, the optimality is defined over a class of functions adhering to the same "moments." In the second case, the optimality is defined by a criterion over an individual function. Since, in practice, we are dealing with a single function rather than with a class of functions (let alone a class of functions with identical "moments"), calculating the optimal parameters with the outlined method is preferably interpreted as an optimality in terms of the enforced convergence rate criterion or, equivalently, a minimum of the center of the energy distribution in the transform domain.

Instead of minimizing the center of the energy distribution in the transform domain as is done for $\lambda(n) = n$, it is also appealing to consider the minimization of the effective width of the energy distribution, especially with coding applications in mind. This could be done by considering the function

$$G(\theta) = \frac{\langle \mathcal{L}_\theta f, \mathcal{L}_\theta f \rangle}{\langle f, f \rangle} - F^2(\theta)$$

if $\lambda(n) = n$. The function G is the effective width of the energy distribution in the transform domain and can, in principle, be minimized with respect to θ. In addition, this optimization over θ can directly be given in terms of "moments" in the time domain, i.e., we can establish the optimal free parameter θ before actually performing the transformation. Thus, in a practical situation, we could restrict ourselves to the calculation of the expansion coefficients within this effective range.

Three remarks apply here [12]. First of all, the expressions for G are more involved than those for F. Second, setting the derivative of G with respect to θ equal to zero does not yield a simple explicit expression for the optimal parameters in any of the specific cases considered in this paper. Last, had we minimized the effective width by taking the appropriate θ, then the centrer of the energy distribution F might be far from 0. This is an unwanted situation since the basis functions ϕ_n are usually constructed via the recurrence relation. Thus, in order to calculate the expansion coefficients within the effective range, we have to calculate a lot of basis functions that are subsequently not used in transformation. All three remarks imply that establishing and applying the minimization of the effective width of the transform would result in a large computational load. Therefore, we consider the minimization of $F(\theta)$ to be the more practical approach.

Appendix

Consider the function $f(t) = K\phi_M(\theta_0; t)$. We can expand the center of the energy distribution $F(\theta)$ in a Taylor series around θ_0, i.e.,

$$F(\theta_0 + \Delta_\theta) = F(\theta_0) + \Delta_\theta \left.\frac{dF}{d\theta}\right|_{\theta_0} + \cdots$$

for Δ_θ sufficiently small (for convenience, we consider the case where the free parameter vector θ is a scalar). Then, we want to show that $dF/d\theta = 0$ at $\theta = \theta_0$ for the considered function f. It is clear that $F(\theta_0) = \lambda(M)$; see (12).

Consider the series expansion

$$f(t) = \sum_{n=0}^{\infty} a_n(\theta)\phi_n(\theta; t)$$

with $\theta = \theta_0 + \Delta_\theta$. For the expansion coefficients, we have the Taylor series

$$\begin{aligned} a_n(\theta_0 + \Delta_\theta) &= \left\langle \phi_n(\theta_0; \cdot) + \Delta_\theta \left.\frac{d\phi_n}{d\theta}\right|_{\theta_0} + \cdots, K\phi_M(\theta_0; \cdot) \right\rangle \\ &= K\delta_{n,M} + K\Delta_\theta \left\langle \left.\frac{d\phi_n}{d\theta}\right|_{\theta_0}, \ \phi_M(\theta_0, \cdot) \right\rangle + \cdots. \end{aligned}$$

Using this Taylor series expansion in

$$F(\theta) = \left\{ \sum_{n=0}^{\infty} \lambda(n) a_n^2(\theta) \right\} \Big/ ||f||^2$$

with $||f||^2 = K^2$, we obtain

$$F(\theta_0 + \Delta_\theta) = \lambda(M) + 2\lambda(M)\Delta_\theta \left\langle \left.\frac{d\phi_M}{d\theta}\right|_{\theta_0}, \phi_M(\theta_0; \cdot) \right\rangle$$

neglecting all terms with powers of Δ_θ greater than 1.

By taking the derivative with respect to θ of the orthonormality condition $\langle \phi_n, \phi_n \rangle = 1$ and under the assumption of real-valued basis functions ϕ_n, it is found that

$$2\left\langle \frac{d\phi_n}{d\theta}, \phi_n \right\rangle = \frac{d\langle \phi_n, \phi_n \rangle}{d\theta} = 0.$$

Combining the previous two equations, we have the desired result.

References

[1] G. Szegö, *Orthogonal Polynomials*. New York: Amer. Math. Soc., 1959.

[2] P. Beckmann, *Orthogonal Polynomials for Engineers and Physicists*. Boulder, CO: Golem, 1973.

[3] B. Wahlberg and P. M. Mäkilä, "On approximation of stable linear dynamical systems using Laguerre and Kautz functions," *Automatica*, vol. 32, pp. 693–708, 1996.

[4] L. R. Lo Conte, R. Metterli, and G. V. Sandri, "Hermite expansions of compact support waveforms: Applications to myoelectric signals," *IEEE Trans. Biomed. Eng.*, vol. 41, pp. 1147–1159, 1994.

[5] J.-B. Martens, "The Hermite transform—Applications," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1607–1618, 1990.

[6] A. C. den Brinker and J. A. J. Roufs, "Evidence for a generalized Laguerre transform of temporal events by the visual system," *Biol. Cybern.*, vol. 67, pp. 395–402, 1992.

[7] N. Tanguy, P. Vilbé, and L. C. Calvez, "Optimum choice of free parameter in orthonormal approximations," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1811–1813, 1995.

[8] T. W. Parks, "Choice of time scale in Laguerre approximations using signal measurements," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 511–513, 1971.

[9] Y. Fu and G. A. Dumont, "An optimum time scale for discrete Laguerre network," *IEEE Trans. Automat. Contr.*, vol. 38, pp. 934–938, 1993.

[10] T. Oliveira e Silva and H. J. W. Belt, "On the determination of the optimal center and scale factor for truncated Hermite series," in *Signal Process. VIII: Theories Applicat., Proc. EUSIPCO, Eighth Euro. Signal Process. Conf.*, Trieste, Italy, 10–13 Sept. 1996, pp. 1563–1566.

[11] H. J. W. Belt and A. C. den Brinker, "Optimal parametrization of truncated generalized Laguerre series," in *Proc. ICASSP Int. Conf. Acoust., Speech, Signal Process.*, Munich, Germany, 21–24 Apr. 1997, pp. 3805–3808.

[12] L. Leyra Fontaíña, "Relations between moments in the time and transform domain," Res, Rep, 97-E-306, Eindhoven Univ. Technol., Eindhoven, The Netherlands, 1997.

Albertus C. den Brinker was born in Heerlen, The Netherlands, in 1957. He received the M.Sc. degree in electrical engineering from Eindhoven University of Technology (EUT), Eindhoven, The Netherlands, in 1983. In 1989, he received the Ph.D. degree from EUT for his work on dynamic models of the human visual system.

Since 1987, he has been working with the Signal Processing Group at the Faculty of Electrical Engineering at EUT. His educational activities include electrical network theory and digital filter theory. His current interests are in the fields of approximation, identification, digital signal processing, and applications of orthogonal series expansions to (local) signal analysis and adaptive filtering.

Harm J. W. Belt (M'97) was born in Hertogenbosch, The Netherlands, in 1968. He received the M.Sc. degree in electrical engineering from Eindhoven University of Technology (EUT), Eindhoven, The Netherlands, in 1993. He received the Ph.D. degree from EUT in 1997 for his work on orthonormal bases for adaptive filtering.

Since 1997, he has been working with the Digital Signal Processing Group of the Philips Research Laboratories, Eindhoven. His current interests are in the fields of approximation, identification, digital signal processing, and adaptive filtering.