with $T(z) = D_1(z)D_2(z^{-1}) - D_2(z)D_1(z^{-1})$. The sign of $Q(z)$ is the same as that of Re $\{G(z)/D_1(z)\}$ on the unit circle; an identical consideration can be made for $R(z)$ and $D_2(z)$. Therefore, $G(z)$ can be computed by finding $Q(z) = Q(z^{-1})$ and $R(z) = R(z^{-1})$ positive on the unit circle and such that

$$R(z)D_1(z) - Q(z)D_2(z) = 0 \qquad \forall z \text{ such that } T(z) = 0, |z| \geq 1.$$

The roots of $T(z)$ outside the unit circle can be readily included in $R(z)$ and $Q(z)$ (and, by symmetry, the roots inside the unit circle). However, given the positivity of $R(e^{j\omega})$ and $Q(e^{j\omega})$ for all ω, the zeros of $T(z)$ on the unit circle must be canceled out by solving the interpolation problem

$$\frac{R(z)}{Q(z)} = \frac{D_2(z)}{D_1(z)} \qquad \forall z \text{ such that } T(z) = 0, |z| = 1.$$

With $D_1(z) = (1 - \rho^M z^{-1})^N$ and $D_2(z) = (1 + \rho^M z^{-1})^N$, those roots are only two, namely, $z = 1$ and $z = -1$. The interpolation can be carried out by the algorithm presented in [8], and the degree of $C(z) = G(z^M)$ can be checked to be NM [10].

V. Simulations

In this example, the three poles of the plant are known to belong to Ω, which is given by

$$\Omega = \{z = \rho e^{j\theta}: -0.28 \leq \theta \leq 0.28, 0.1 \leq \rho \leq 0.9\}.$$

It can be proved that there does not exist $C(z)$ such that $C(z)/A_*(z)$ is SPR for all $A_*(z)$ with their roots in Ω. However, the region $\Omega^2 = \{z: \sqrt{z} \in \Omega\}$, which characterizes the uncertainty in the roots of $\mathcal{D}(z)$ as in (10), for $N = 3, M = 2$ can be enclosed in a region that admits two extreme polynomials, namely, $D_1(z) = 1$ and $D_2(z) = (1 - 0.81z^{-1})^3$, which satisfy (9). Such a region is the intersection of the circles centered at 0.5 and -0.095 and with radii 0.5 and 0.905, respectively. See [9] for more details on these types of regions. An appropriate compensator $C(z)$ is obtained following the steps shown in Section IV: $C(z) = 1 - 1.33z^{-2} + 0.36z^{-4}$. The true plant is given by $H(z) = (1/A_*(z))$ with

$$A_*(z) = 1 - 2.5799z^{-1} + 2.2804z^{-2} - 0.6885z^{-3}.$$

The input is zero-mean, unit variance white noise filtered by $S(z) = (1/A_s(z))$ with

$$A_s(z) = 1 - 2.4581z^{-1} + 3.0648z^{-2} - 1.9911z^{-3} + 0.6561z^{-4}.$$

The spectrum of $u(\cdot)$ is especially significant for those ω such that Re $(1/A_*(e^{j\omega})) < 0$. White measurement noise is introduced so that the SNR at the plant output is 26 dB. Fig. 1 shows the trajectories of the denominator coefficients in two cases. On the top, we have $M = 2$ and the robust compensator computed above. With no compensator, an expansion factor $M = 3$ was needed to achieve convergence in the bottom part of the figure.

VI. Conclusions

An analysis of hyperstability-based adaptive IIR filtering algorithms with polyphase structures has been performed. By appropriately selecting the polyphase expansion factor, the SPR condition that this class of algorithms requires for convergence can be satisfied. A method for this selection has been given, using *a priori* information about the location of the unknown plant poles in the form of uncertainty regions.

References

[1] C. R. Johnson, Jr., *Lectures on Adaptive Parameter Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[2] P. A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*. New York: Marcel Dekker, 1995.

[3] S. Shah and G. F. Franklin, "On satisfying strict positive real condition for convergence by overparameterization," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 715–716, June 1982.

[4] P. M. S. Burt and M. Gerken, "A polyphase IIR adaptive filter: Error surface analysis and application," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Apr. 1997, pp. 2285–2288.

[5] B. D. O. Anderson and C. R. Johnson, Jr., "Exponential convergence of adaptive identification and control algorithms," *Automatica*, vol. 18, pp. 1–13, May 1982.

[6] M. G. Larimore, J. R. Treichler, and C. R. Johnson, Jr., "SHARF: An algorithm for adapting IIR digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 428–440, Aug. 1980.

[7] B. D. O. Anderson, S. Dasgupta, P. Khargonekar, F. J. Kraus, and M. Mansour, "Robust strict positive realness: Characterization and construction," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 869–876, July 1990.

[8] C. Mosquera and F. Pérez, "An algorithm for interpolation with positive rational functions on the imaginary axis," *Automatica*, vol. 33, pp. 2277–2280, Dec. 1997.

[9] A. Tesi, A. Vicino, and G. Zappa, "Design criteria for robust strict positive realness in adaptive schemes," *Automatica*, vol. 30, pp. 643–654, Apr. 1994.

[10] C. Mosquera, "Relaxation of the SPR condition with applications to the convergence of adaptive recursive algorithms," Ph.D. dissertation, Univ. Vigo, Vigo, Spain, 1998.

The Continuous Wavelet Transform as a Maximum Entropy Solution of the Corresponding Inverse Problem

Laura Rebollo-Neira and Juan Fernandez-Rubio

***Abstract*—The continuous wavelet transform is obtained as a maximum entropy solution of the corresponding inverse problem. It is well known that although a signal can be reconstructed from its wavelet transform, the expansion is not unique due to the redundancy of continuous wavelets. Hence, the inverse problem has no unique solution. If we want to recognize one solution as "optimal," then an appropriate decision criterion has to be adopted. We show here that the continuous wavelet transform is an "optimal" solution in a maximum entropy sense.**

I. Introduction

The continuous wavelet transform (CWT) $\tilde{f}(a,b); (a,b) \in R^2$ of a signal $f(t) \in L^2(R)$, with respect to an admissible mother wavelet

Manuscript received January 23, 1997; revised January 24, 1998. This work was supported by CIRIT of Catalunya, CICYT of Spain (TIC96-0500-C10-01), and CICPBA of Argentina. The associate editor coordinating the review of this paper and approving it for publication was Prof. P. P. Vaidyanathan.

L. Rebollo-Neira is a Fellow of the CICPBA with the Departamento de Física, Universidad Nacional de La Plata, La Plata, Argentina and with the Departament de Teoria del Senyal i Comunicacions, Escola Tecnica Superior d'Enginyers de Telecomunicació, Polytechnic University of Catalunya (UPC), Barcelona, Spain

J. Fernandez-Rubio is with the Departament de Teoria del Senyal i Comunicacions, Escola Tecnica Superior d'Enginyers de Telecomunicació, Polytechnic University of Catalunya (UPC), Barcelona, Spain (e-mail: juan@gps.tsc.upc.es).

Publisher Item Identifier S 1053-587X(99)04641-3.

$\Psi(t) \in L^2(R)$, is defined as [1]

$$\tilde{f}(a,b) = \langle \Psi_{a,b} \mid f \rangle = \int_R f(t)\Psi^*_{a,b}(t)\,dt; \quad (a,b) \in R^2 \tag{1}$$

where $\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\Psi(\frac{t-b}{a})$; $(a,b) \in R^2$, and $\Psi^*_{a,b}(t)$ indicates the complex conjugate of $\Psi_{a,b}(t)$. Let us denote $\mathcal{F}$ as the subspace of all wavelet transforms with respect to a mother $\Psi(t) \in L^2(R)$, i.e.,

$$\mathcal{F} = \left\{ \tilde{f}(a,b) = \langle \Psi_{a,b} \mid f \rangle = \int_R f(t)\Psi^*_{a,b}(t)\,dt \right. \\ \left. \text{for some } f(t) \in L^2(R); (a,b) \in R^2 \right\}. \tag{2}$$

As a special case of the resolution of the identity property of the CWT [1], we have

$$\int_R |f(t)|^2\,dt = \frac{1}{C_\Psi}\int_{R^2} |\tilde{f}(a,b)|^2 \frac{dadb}{a^2} \tag{3}$$

where C_Ψ is the admissibility constant, $C_\Psi = \int_{-\infty}^{\infty} \frac{|\hat{\Psi}(\omega)|^2}{|\omega|}d\omega$, and $\hat{\Psi}(\omega)$ is the Fourier transform of $\Psi(t)$. Equation (3) shows that the CWT maps $L^2(R)$ isometrically into $\mathcal{L} = L^2(R^2; C_\Psi^{-1}a^{-2}dadb)$, which is the space of all the functions $h(a,b); (a,b) \in R^2$ that are square-integrable with respect to the weight function $C_\Psi^{-1}a^{-2}$. However, the isometry is only partial because its range $\mathcal{F}$ constitutes a closed subspace and not all of $\mathcal{L}$. Thus, not every $h \in \mathcal{L}$ can be the CWT of some signal $f \in L^2(R)$. A necessary and sufficient condition for a function $h \in \mathcal{L}$ to belong to $\mathcal{F}$ is that it satisfies the *reproducing kernel* equation [1], [3]

$$h(a',b') = \frac{1}{C_\Psi}\int_{R^2} h(a,b)\langle \Psi_{a',b'} \mid \Psi_{a,b} \rangle \frac{dadb}{a^2}; \quad (a',b') \in R^2. \tag{4}$$

If $h \notin \mathcal{F}$, the right-hand side of (4) gives rise to the orthogonal projection of h onto $\mathcal{F}$ [3]. Hence, although the function f can be reconstructed (at least in the weak sense) from its CWT through the integral [1], [3]

$$f(t) = \frac{1}{C_\Psi}\int\int_{R^2} \tilde{f}(a,b)\Psi_{a,b}(t)\frac{dadb}{a^2}; \quad t \in R \tag{5}$$

the expansion is not unique. The most general solution can be expressed as $h = \tilde{f} + \tilde{f}^\perp$ for any $\tilde{f}^\perp \in \mathcal{F}^\perp$, where $\mathcal{F}^\perp$ denotes the orthogonal complement of $\mathcal{F}$ in $\mathcal{L}$ [3]. Consequently, the problem of inverting the integral equation

$$f(t) = \frac{1}{C_\Psi}\int\int_{R^2} h(a,b)\Psi_{a,b}(t)\frac{dadb}{a^2}; \quad t \in R \tag{6}$$

has an infinite number of solutions. This problem appears typically in RADAR and SONAR signal processing when the wideband model for *echo location* is considered, and only one outgoing signal can be used. In this case, the function h is the *reflectivity density*, which represents a dense-target environment, and the function f the only available *echo* [3], [10]. In the general case, (6) can be seen as the input–output relationship for a time-scale system characterized by the function h. From a single input–output operation, it is impossible to uniquely determine the system. Since infinitely many systems can produce identical output from the input, if we want to recognize a particular function h as a good estimate of the unknown system, an appropriate *decision criterion* has to be adopted. By choosing $\tilde{f}^\perp = 0$, the CWT $\tilde{f}$ appears to be "optimal" solution in a minimum norm (MN) sense. The MN requirement may be a reasonable criterion to be adopted in some applications; however, in other cases, there may be no *a priori* reason for accepting a MN solution as a suitable estimate of the unknown solution. In this correspondence, we address the problem of deciding on an appropriate estimate for the unknown system h by recourse to a postulate originally conceived for the purpose of making decisions in indeterminate situations, namely, the principle of maximum entropy (ME) [5]–[7].

The ME principle is based on the interpretation of a probability distribution as expressing the lack of the complete knowledge required to draw definite conclusions. The entropy is claimed to be a measure of such uncertainty or "ignorance," and the postulate states that the "least committal" probability distribution with regard to the missing information is the one that maximizes the entropy (uncertainty) and agrees with what is known [5]–[7]. According to the ME prescription, inferences on the basis of incomplete information should be made through predictions of expected values computed with the probability density that, while reproducing the available information, yields maximum uncertainty with respect to all other matters [5]–[7].

The problem of undertaking the selection of *one* $h \in \mathcal{L}$ as an estimate for the unknown system on the basis of the incomplete information that provides (6) is certainly well represented by the above ME formalism. Our "uncertainty" underlies the fact that we cannot be certain about which system h has generated the output f. The need for a statistical description is further supported by the wideband regime example mentioned above. Since, in such a case, the time-scale system h represents a dense-target environment, it is clear that h is most likely to change randomly. We then assign a probability density to each function h and consider the expected value of h to represent the system. The ME postulate is employed in order to reconstruct the probability density from the output f.

The ME criterion for estimating solutions of underdetermined inverse problems and integral equations has been extensively applied in several fields during the last 20 years. As examples, we can mention [2], [8], [13], [15], [16], and [18]. In particular, the ME approach with constraints on the mean solution has been used in crystallography [11] and other contexts [9]. We incorporate a restriction on the mean variance of the process in order to ensure the existence of the ME distribution over the unbounded range of definition. Once the ME probability density satisfying the given constraints is determined, we are in a position to "infer" the unknown system by predicting its expected value. The result is that the predicted expected value corresponds to the MN solution, i.e., the CWT. In other words, we provide one more reason for deciding on the MN solution as, from an ME point of view, it turns out to be the "least biased" assignment that we can make on the basis of the available data. This result has also been shown to hold as a property within the *frame theory* [14].

It should be stressed that through the proposed scheme, the ME estimate for a time-scale system is able to be improved by additional input–output operations. Provided that such information were available, it can be introduced in the proposed framework simply by adding the corresponding equations as constraints. In this sense, the result we derived here should be understood as a "first-order" ME estimate of the unknown system since such an estimate is obtained from a single input–output relationship.

II. ME Statistical Description

We address here the problem of estimating *one* function $h \in \mathcal{L}$ when the only information we are given is (6) and the integral is assumed to converge to f at least in the weak sense, i.e., all the inner products are equal. In order to properly expressed this restriction, we take an orthonormal basis spanning $L^2(R)$, say $y_j(t); j = 1, \ldots, \infty$, represent $f(t); t \in R$ as

$$f(t) = \sum_{j=1}^{\infty} s_j y_j(t); \quad s_j = \langle y_j \mid f \rangle \tag{7}$$

and require that

$$s_j = \frac{1}{C_\Psi}\int\int_{R^2} h(a,b)\langle y_j \mid \Psi_{a,b}\rangle \frac{dadb}{a^2}; \quad j = 1,\ldots,\infty. \tag{8}$$

Let $\alpha_n(a,b);\ n = 1,\ldots,\infty; (a,b) \in R^2$ be an orthonormal basis for $\mathcal{L}$. Thus, each function $h \in \mathcal{L}$ can be expressed as

$$h(a,b) = \sum_{n=1}^{\infty} c_n \alpha_n(a,b); \quad (a,b) \in R^2 \tag{9}$$

with

$$c_n = \frac{1}{C_\Psi}\int_{R^2} h(a,b)\alpha_n^*(a,b)\frac{dadb}{a^2}; \quad n = 1,\ldots,\infty. \tag{10}$$

The above equations provide a one-to-one correspondence between a function h and the sequence $c_1,\ c_2,\ldots$. Hence, the problem of selecting *one* solution h among the infinitely many possible is equivalent to selecting *one* sequence $c_1,\ c_2,\ldots$. Our aim is that of deciding on such a sequence through a *decision criterion* to establish how to use the available incomplete information in an optimal way. The inversion problem is then transformed into a problem of *inference*. As argued in the previous section, we place the problem within this framework by expressing our ignorance of the "true" sequence by assigning a probability density to each possible one. The statistical estimate of the sequence is represented by the expected value of the random sequence that we denote as $\overline{c_1}, \overline{c_2},\ldots$. Accordingly, the statistical estimate of h is evaluated as $\overline{h(a,b)} = \sum_{n=1}^{\infty}\overline{c_n}\alpha_n(a,b);\ (a,b) \in R^2$. Introducing the last expression in (8), the constraints of our problem adopt the form

$$s_j = \frac{1}{C_\Psi}\int_{R^2}\sum_{n=1}^{\infty}\overline{c_n}\alpha_n(a,b)\langle y_j \mid \Psi_{a,b}\rangle\frac{dadb}{a^2}; \quad j = 1,\ldots,\infty \tag{11}$$

that we split into real and imaginary parts, and they become

$$s_j^u = \sum_{n=1}^{\infty}\overline{c_n^u}w_{nj}^u - \overline{c_n^v}w_{nj}^v; \quad j = 1,\ldots,\infty \tag{12}$$

$$s_j^v = \sum_{n=1}^{\infty}\overline{c_n^u}w_{nj}^v + \overline{c_n^v}w_{nj}^u; \quad j = 1,\ldots,\infty \tag{13}$$

where $\overline{c_n^u}$ and $\overline{c_n^v}$ are the real and imaginary parts of $\overline{c_n}$, respectively, whereas s_j^u and s_j^v are the real and imaginary parts of s_j, and w_{nj}^u and w_{nj}^v are the real and imaginary parts of $w_{nj} = \frac{1}{C_\Psi}\int_{R^2}\alpha_n(a,b)\langle y_j \mid \Psi_{a,b}\rangle\frac{dadb}{a^2}$.

In what follows, we shall consider M-dimensional sequences c_1, $c_2,\ldots,c_M$ and examine the $\lim_{M\to\infty}$ at the end of the calculations. To simplify notation, let us denote $\mathbf{c^u} = c_1^u,\ldots,c_M^u$ and $\mathbf{c^v} = c_1^v,\ldots,c_M^v$. Assuming that these $2M$ random variables are distributed according to a joint probability density $P(\mathbf{c^u},\mathbf{c^v})$, the expected values $\overline{c_n^u}$, $\overline{c_n^v}$ involved in (12) and (13) are calculated using the expected value definition, i.e.,

$$\overline{c_n^u} = \int_{-\infty}^{\infty} P(\mathbf{c^u},\mathbf{c^v})c_n^u d\mathbf{c^u}\,d\mathbf{c^v}; \quad j = 1,\ldots,M \tag{14}$$

$$\overline{c_n^v} = \int_{-\infty}^{\infty} P(\mathbf{c^u},\mathbf{c^v})c_n^v d\mathbf{c^u}\,d\mathbf{c^v}; \quad j = 1,\ldots,M \tag{15}$$

where the integral sign denotes $2M$ integrals, whereas $d\mathbf{c^u} = dc_1^u,\ldots,dc_M^u$, and $d\mathbf{c^v} = dc_1^v,\ldots,dc_M^v$.

In order to guarantee the existence of the ME probability distribution over the range $(-\infty,\infty)^{2M}$, we shall require the mean variance of the process to be finite. Consequently, in addition to the normalization condition

$$\int_{-\infty}^{\infty} P(\mathbf{c^u},\mathbf{c^v})d\mathbf{c^u}\,d\mathbf{c^v} = 1 \tag{16}$$

we set the constraint

$$\frac{1}{2M}\sum_{n-1}^{M}\int_{-\infty}^{\infty} P(\mathbf{c^u},\mathbf{c^v})\left(c_n^{u\,2} + c_n^{v\,2}\right)d\mathbf{c^u}\,d\mathbf{c^v} = C \tag{17}$$

where C is an *unknown* constant.

We face now the problem of determining $P(\mathbf{c^u},\mathbf{c^v})$ satisfying (12), (13), (16), and (17). Among all the $P(\mathbf{c^u},\mathbf{c^v})$ capable of fulfilling these constraints, we select one adopting the ME principle. As we are dealing with random variables of continuous-type, in the absence of *a priori* knowledge about the sought distribution, the uncertainty measure we consider is the differential entropy, which is defined as [4]

$$H(\mathbf{c^u},\mathbf{c^v}) = -\int_{-\infty}^{\infty} P(\mathbf{c^u},\mathbf{c^v})\ln P(\mathbf{c^u},\mathbf{c^v})d\mathbf{c^u}\,d\mathbf{c^v}. \tag{18}$$

(Note: The entropy maximization is tantamount to assuming a uniform prior distribution in the minimum cross entropy formalism [17].)

We look for the probability density that maximizes (18) with constraints (12), (13), (16), and (17). We incorporate each constraint (12) into the variational process through a Lagrange multiplier λ_j^u and each constraint (13) through a Lagrange multiplier λ_j^v. Constraints (16) and (17) are introduced through the Lagrange multipliers λ_0 and β, respectively. Thus, the functional S to be maximized is cast as

$$\begin{aligned} S = -\int_{-\infty}^{\infty} P(\mathbf{c^u},\mathbf{c^v})\Bigg(&\ln P(\mathbf{c^u},\mathbf{c^v}) \\ &+ \beta\sum_{n-1}^{M}\left(c_n^{u2} + c_n^{v2}\right) + \lambda_0\Bigg)d\mathbf{c^u}\,d\mathbf{c^v} \\ &- \sum_{j=1}^{\infty}\lambda_j^u\sum_{n=1}^{M}\left(\overline{c_n^u}w_{nj}^u - \overline{c_n^v}w_{nj}^v\right) \\ &- \sum_{j=1}^{\infty}\lambda_j^v\sum_{n=1}^{M}\left(\overline{c_n^u}w_{nj}^v + \overline{c_n^v}w_{nj}^u\right). \end{aligned} \tag{19}$$

From the condition $\frac{\delta S}{\delta P} = 0$, we obtain

$$\begin{aligned} P(\mathbf{c^u},\mathbf{c^v}) = &\exp -(\lambda_0 + 1) \\ &\times \exp\left(-\sum_{n=1}^{M}\left(c_n^u\gamma_n^u + c_n\gamma_n^v + \beta c_n^{u\,2} + \beta c_n^{v\,2}\right)\right) \end{aligned} \tag{20}$$

where $\gamma_n^u = \sum_{j=1}^{\infty}(\lambda_j^u w_{nj}^u + \lambda_j^v w_{nj}^v)$, and $\gamma_n^v = \sum_{j=1}^{\infty}(\lambda_j^v w_{nj}^u - \lambda_j^u w_{nj}^v)$.

The verification that the stationary H is actually a maximum, and the only one is straightforward with the aid of the well-known inequality [12]

$$\begin{aligned} -\int_{-\infty}^{\infty} P^\dagger(\mathbf{c^u},\mathbf{c^v})\ln P^\dagger(\mathbf{c^u},\mathbf{c^v})\,d\mathbf{c^u}\,d\mathbf{c^v} \\ \le -\int_{-\infty}^{\infty} P^\dagger(\mathbf{c^u},\mathbf{c^v})\ln P(\mathbf{c^u},\mathbf{c^v})\,d\mathbf{c^u}\,d\mathbf{c^v} \end{aligned} \tag{21}$$

holding for any $P^\dagger(\mathbf{c^u},\mathbf{c^v})$ and $P(\mathbf{c^u},\mathbf{c^v})$, which are both normalized. The equality is reached if and only if $P^\dagger(\mathbf{c^u},\mathbf{c^v}) = P(\mathbf{c^u},\mathbf{c^v})$.

With $P(\mathbf{c^u},\mathbf{c^v})$ given in (20), the differential entropy H is

$$H = (\lambda_0 + 1) + \sum_{j=1}^{\infty}\lambda_j^u s_j^u + \sum_{j=1}^{\infty}\lambda_j^v s_j^v + \beta 2MC \tag{22}$$

whereas if we take any other $P^\dagger(\mathbf{c^u},\mathbf{c^v})$ satisfying constraints (12), (13), (16), and (17), the corresponding differential entropy $\breve{H}$ will

be $\tilde{H} = -\int_{-\infty}^{\infty} P^{\dagger}(\mathbf{c^u}, \mathbf{c^v}) \ln P^{\dagger}(\mathbf{c^u}, \mathbf{c^v}) d\mathbf{c^u}\, d\mathbf{c^v}$, and using (21), we have

$$\begin{aligned}\tilde{H} \le \int_{-\infty}^{\infty} P^{\dagger}(\mathbf{c^u}, \mathbf{c^v}) \Bigg(& (\lambda_0 + 1) + \sum_{n-1}^{M} c_n^u \sum_{j=1}^{\infty} \left(\lambda_j^u w_{nj}^u + \lambda_j^v w_{nj}^v\right) \\ & + \sum_{n-1}^{M} c_n^v \sum_{j=1}^{\infty} \left(\lambda_j^v w_{nj}^u - \lambda_j^u w_{nj}^v\right) \\ & + \beta \sum_{n-1}^{M} \left(c_n^{u\,2} + c_n^{v\,2}\right) \Bigg) d\mathbf{c^u}\, d\mathbf{c^v} \\ = (\lambda_0 + 1) + \sum_{j=1}^{\infty} \lambda_j^u s_j^u + \sum_{j=1}^{\infty} \lambda_j^v s_j^v + \beta 2MC = H \end{aligned} \tag{23}$$

which shows that $P(\mathbf{c^u}, \mathbf{c^v})$, as in (20), corresponds to the absolute maximum of the differential entropy H subject to the given constraints.

The normalization condition (16) entails

$$\begin{aligned}& \exp(\lambda_0 + 1) \\ & \quad = \int_{-\infty}^{\infty} \exp -(\lambda_0 + 1) \\ & \qquad \times \exp\left(- \sum_{n-1}^{M} \left(c_n^u \gamma_n^u + c_n \gamma_n^v + \beta c_n^{u\,2} + \beta c_n^{v\,2}\right)\right) d\mathbf{c^u}\, d\mathbf{c^v} \\ & \quad = \left(\frac{\pi}{\beta}\right)^M \prod_{n=1}^{M} \exp\left(\frac{\left(\gamma_n^u\right)^2}{4\beta}\right) \exp\left(\frac{\left(\gamma_n^v\right)^2}{4\beta}\right) \end{aligned} \tag{24}$$

so that by substituting (20) in (14) and (15) and performing the integrals, we obtain

$$\begin{aligned}\overline{c_n} &= -\frac{1}{2\beta} \sum_{j=1}^{\infty} \lambda_j w_{nj}^* \\ &= -\frac{1}{C_\Psi 2\beta} \int_{R^2} \left\langle \Psi_{a,b} \left| \sum_{j=1}^{\infty} \lambda_j y_j \right. \right\rangle \alpha_n^*(a,b) \frac{da\,db}{a^2} \\ & \qquad\qquad n = 1, \ldots, M \end{aligned} \tag{25}$$

where $\lambda_j = \lambda_j^u + i\lambda_j^v;\ j = 1, \ldots, \infty$. On the other hand

$$\overline{c_n} = \frac{1}{C_\Psi} \int_{R^2} \overline{h(a,b)} \alpha_n^*(a,b) \frac{da\,db}{a^2}. \tag{26}$$

On comparing (25) and (26), we gather that $\overline{h(a,b)} = -(1/2\beta)\langle \Psi_{a,b} \mid \sum_{j=1}^{\infty} \lambda_j y_j \rangle; (a,b) \in R^2$, and hence, $\bar{h} \in \mathcal{F}$ [cf. (2)]. Thus, we are in a position to univocally determine $\bar{h}$. In fact, by using $\bar{h}$ in (6) and performing the inner product of both sides with $\Psi_{a',b'}$, we have

$$\begin{aligned}\langle \Psi_{a',b'} \mid f \rangle = \frac{1}{C_\Psi} \int\int_{R^2} \overline{h(a,b)} \langle \Psi_{a',b'} \mid \Psi_{a,b} \rangle \frac{da\,db}{a^2} \\ (a',b') \in R^2 \end{aligned} \tag{27}$$

and, since $\bar{h} \in \mathcal{F}$, the *reproducing kernel* equation (4) is verified. Hence, from (27), we conclude that $\overline{h(a',b')} = \langle \Psi_{a',b'} \mid f \rangle = \tilde{f}(a',b');\ (a',b') \in R^2$, which proves that the CWT is an "optimal" estimate of the inverse problem in a ME sense. This result shows that from an ME viewpoint, the MN criterion works out by averaging functions in a maximally noncommittal way and, therefore, is the "least biased" assignment we can make on the basis of the available information.

We now examine the resulting differential entropy H, which adopts the simple expression $H = M \ln \frac{\pi e}{\beta}$. Hence, the pertinent differential entropy rate, or differential entropy per degree of freedom, which is defined as [12] $\bar{H} = \lim_{M \to \infty} \frac{1}{2M} H(\mathbf{c^u}, \mathbf{c^v})$, turns out to be $\bar{H} = \frac{1}{2} \ln \frac{\pi e}{\beta}$.

As a final remark, we wish to stress that the proposed framework can be used to improve on the given ME estimate for the unknown system h if additional input-out operations are available. Indeed, if we input another waveform $\Psi' \neq \Psi$ and obtain an output $f' \neq f$, this information can be incorporated into the above scheme simply by adding the corresponding equation to the previous constraints. The ME estimate we have derived here should be understood as a "first-order" ME estimate of the unknown system in the sense that it was obtained from a *single* input–output relationship.

III. Conclusion

The CWT has been derived as a ME estimate of the concomitant inverse problem. This was achieved by

i) transforming each possible solution into a sequence of coefficients;
ii) assigning a probability density to each sequence;
iii) representing the unknown solution through the expected value of such random sequence;
iv) undertaking the reconstruction of the probability density on the basis of the given constraints and the principle of maximum entropy;
v) deciding on the unknown solution through the expected value sequence inferred by the ME probability density.

The so-estimated solution was shown to be identical to the CWT. Thus, we are led to conclude that the CWT, or MN solution, is an "optimal" estimate of the concomitant inverse problem from a ME point of view. This result supports the use of the MN solution as being the "least biased" assignment we can make on the basis of the incomplete information that provides (6). In addition, the proposed ME treatment for estimating a time-scale system represented by the function h allows for the inclusion of additional information to improve upon the given estimate.

Acknowledgment

The authors wish to thank the reviewers for their valuable comments, A. Maclean for corrections of the manuscript, and L. Janer for useful discussions.

References

[1] I. Daubechies, *Ten Lectures on Wavelets.* Philadelphia, PA: SIAM, 1992.
[2] P. P. B. Eggermont, "Maximum entropy regularization for fredholm integral equations of the first kind," *SIAM J. Math. Anal.*, vol. 24, no. 6, pp. 1556–1576, 1993.
[3] G. Kaiser, *A Friendly Guide to Wavelets.* Berlin, Germany: Birkhäuser, 1994.
[4] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.
[5] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, pp. 620–630, 1957.
[6] ——, "Where do we stand maximum entropy?," in *The Maximum Entropy Formalism*, R. Levine and M. Tribus, Eds. Cambridge, MA: MIT Press, 1979.
[7] ——, "Prior information and ambiguity in inverse problems," in *Proc. SIAM-AMS*, 1984, vol. 14, pp. 151–166.
[8] A. Mohammad-Djafari and G. Demoment, Eds., *Maximum Entropy and Bayesian Methods*. Boston, MA: Kluwer, 1993.
[9] A. Mohammad-Djafari, "Maximum entropy and linear inverse problems. A short review," in *Maximum Entropy and Bayesian Methods*, A. Mohammad-Djafari and G. Demoment, Eds. Boston, MA: Kluwer, 1993, pp. 253–264.
[10] H. Naparst, "Dense target signal processing," *IEEE Trans. Inform. Theory*, vol. 37, pp. 317–327, 1991.
[11] J. Navaza, "On the maximum-entropy estimate of the electron density function," *Acta Cryst. A*, vol. 41, pp. 232–244, 1985.

[12] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1991.
[13] C. R. Smith and W. T. Grandy Jr., Eds., *Maximum Entropy and Bayesian Methods in Inverse Problems*. Dordrecht, The Netherlands: Reidel, 1985.
[14] L. Rebollo-Neira, J. F. Rubio, and A. Plastino, "Frames: A maximum entropy statistical estimate of the inverse problem," *J. Math. Phys.*, vol. 38, no. 9, pp. 4863–4871, 1997.
[15] L. Rebollo-Neira and A. G. Constantinides, "Power spectrum estimation from noisy and limited autocorrelation values: A maximum entropy approach," *Signal Process.*, vol. 56, no. 2, pp. 135–147, 1997.
[16] J. Skilling, Ed., *Maximum Entropy and Bayesian Methods*. Boston, MA: Kluwer, 1988.
[17] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 26–37, 1980.
[18] C. R. Smith and W. T. Grandy Jr., Ed., *Maximum Entropy and Bayesian Methods in Inverse Problems*. Dordrecht, The Netherlands: Reidel, 1985.

Multidimensional Polynomial Transform Algorithm for Multidimensional Discrete W Transform

Yonghong Zeng and Xiaomei Li

***Abstract*—The multidimensional (MD) polynomial transform is used to convert the MD W transform (MDDWT) into a series of one-dimensional (1-D) W transforms (DWT's). Thus, a new polynomial transform algorithm for the MDDWT is obtained. The algorithm needs no operations on complex data. The number of multiplications for computing an r-dimensioanl DWT is only $\frac{1}{r}$ times that of the commonly used row-column method. The number of additions is also reduced considerablely.**

***Index Terms*— Discrete transform, fast algorithm, multidimensional signal processing.**

I. Introduction

As the generalization of the multidimensional (MD) discrete Hartley transform [1]–[3], the MD discrete W transform (MDDWT) is a tool for MD problems. MDDWT is symmetrical and real. It is simpler than the MD discrete Fourier transform (MDDFT) since it needs no operations on complex data. Therefore, it is useful for processing MD signals. In applications such as image processing and optical engineering, DWT (DHT) has been effectively used [3], [4]. In areas of computer vision, high definition television (HDTV), and vision telephony, where we must process or analysis motion images [which is also called multiframe detection (MFD)] MD discrete transforms are often used. Currently, the discrete Fourier transform (DFT) and the discrete cosine transform (DCT) are the two most commonly used transforms in these areas [5], [6]. If we process the motion of three-dimensional (3-D) images, the four-dimensional (4-D) DWT may be used. Furthermore, the MD DWT can be used to compute the MD DFT in order to reduce computational complexity [7].

Since the computational complexity for the MDDWT is extraordinary high, designing a good fast algorithm is very important. Although the kernel of the MDDWT is not separable, we can turn it into a transform with its kernel separable [3], [8]. The resulting transform can be computed by the well-known row-column method. The row-column method is commonly used, but it is far from the best. The polynomial transform is considered to be a useful tool for handling MD problems. It has been successfully used for computing MD DFT, MD convolution, and MD DCT, and so on [9], [11]. In this correspondence, we use the polynomial transform to derive a new algorithm for the MDDWT. The number of operations needed by the proposed algorithm is reduced considerably compared with the row-column method; the number of multiplications is only $\frac{1}{r}$, and the number of additions is also less, where r represents the dimension of the transform. Furthermore, the algorithm is simple in structure and easy for programming.

Manuscript received December 20, 1996; revised December 2, 1998. This work was supported by the National Natural Science Foundation under Grant 19601012 and the China 863 Foundation under Grant 863-306-ZD-03-3. The associate editor coordinating the review of this paper and approving it for publication was Dr. Ali H. Sayed.

Y. Zeng is with Department 7, National University of Defense Technology, Changsha, P.R. China (e-mail: yhzhen@nudt.edu.cn).

X. Li is with Department 6, National University of Defense Technology, Changsha, P.R. China.

Publisher Item Identifier S 1053-587X(99)04663-2.

II. MD Polynomial Transform Algorithm for r–Dimensional DWT-II

In the following, assume that $N_r = 2^t$, $N_r/N_i = 2^{l_i}$, $i = 1, 2, \ldots, r-1$, where $l_i \geq 0$ is an integer. An r-dimensional DWT-II with size $N_1 \times N_2 \times \cdots \times N_r$ is

$$X(k_1, k_2, \ldots, k_r) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \cdots \sum_{n_r=0}^{N_r-1} x(n_1, n_2, \ldots, n_r) \cdot \operatorname{cas}\left[\frac{\pi(2n_1+1)k_1}{N_1} + \cdots + \frac{\pi(2n_r+1)k_r}{N_r}\right] \quad (1)$$

$$k_i = 0, 1, \ldots, N_i - 1; \quad i = 1, 2, \ldots, r. \quad (2)$$

This is a multiple sum. In the following, we try to change the order of the sum and eliminate the redundant operations in the sum.

Lemma 1: Let $p_i(n_r)$ be the least non-negative remainder of the $(2p_i + 1)n_r + p_i$ module N_i, and let $A = \{(n_1, n_2, \ldots, n_r) \mid 0 \leq n_i \leq N_i - 1;\ 1 \leq i \leq r\}$, $B = \{(p_1(n_r), p_2(n_r), \ldots, p_{r-1}(n_r), n_r) \mid 0 \leq p_i \leq N_i - 1; 1 \leq i \leq r-1; 0 \leq n_r \leq N_r - 1\}$. Then, $A = B$.

Proof: It is sufficient to prove that the elements in B are different from each other. Let $(p_1(n_r), p_2(n_r), \ldots, p_{r-1}(n_r), n_r)$ and $(p_1'(n_r'), \ldots, p_{r-1}'(n_r'), n_r')$ be two elements in B. If they are equal, then

$$p_i(n_r) = p_i'(n_r'), \quad i = 1, 2, \ldots, r-1; \quad n_r = n_r'.$$

From the definition of $p_i(n_r)$, we see that

$$(2p_i + 1)n_r + p_i \equiv (2p_i' + 1)n_r + p_i' \bmod N_i.$$

Hence

$$(2n_r + 1)(p_i - p_i') \equiv 0 \bmod N_i, \quad i = 1, 2, \ldots, r-1.$$

Since $2n_r + 1$ is relatively prime with N_i, we get $p_i \equiv p_i' \bmod N_i$, that is, $p_i = p_i'$.