VI. Conclusion

In this correspondence, a general design for allpass variable fractional delay (VFD) digital filters with minimum weighted integral squared error subject to a constraint on maximum error deviation from the desired response was formulated. Design examples show that a trade-off can be achieved between the integral squared error and the maximum error deviation for the allpass VFD filters. From the WLS solution, the maximum error deviation can be reduced while maintaining approximately the same integral squared error.

Acknowledgment

The authors would like to thank Prof. A. Cantoni for useful discussions about allpass VFD filter structure.

References

[1] C. W. Farrow, "A continuously variable digital delay element," in *Proc. IEEE Int. Symp. Circuits Syst.*, Jun. 1988, vol. 3, pp. 2641–2645.

[2] S. C. Pei and C. C. Tseng, "A comb filter design using fractional-sample delay," *IEEE Trans. Circuits Syst. II*, vol. 45, no. 6, pp. 649–653, Jun. 1998.

[3] T. B. Deng, "Noniterative WLS design of allpass variable fractional-delay digital filters," *IEEE Trans. Circuits Syst.*, vol. 53, no. 2, pp. 358–371, Feb. 2006.

[4] H. H. Dam, A. Cantoni, K. L. Teo, and S. Nordholm, "FIR variable digital filter with signed power-of-two coefficients," *IEEE Trans. Circuits Syst. I*, vol. 54, no. 6, pp. 1348–1357, Jun. 2007.

[5] H. H. Dam, A. Cantoni, K. L. Teo, and S. Nordholm, "Vatiable digital filter with least square criterion and peak gain constraints," *IEEE Trans. Circuits Syst. II*, vol. 54, no. 1, pp. 24–28, Jan. 2007.

[6] M. Makundi, T. I. Laakso, and V. Valimaki, "Efficient tunable IIR and allpass structures," *Electron. Lett.*, vol. 37, pp. 344–345, Mar. 2001.

[7] C. C. Tseng, "Design of 1-D and 2-D variable fractional delay allpass filters using weighted least square methods," *IEEE Trans. Circuits Syst. I*, vol. 49, no. 10, pp. 1413–1422, Oct. 2002.

[8] Z. Jing, "A new method for digital all-pass filter design," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, pp. 1557–1564, Nov. 1987.

[9] J. Y. Kaakinen and T. Saramaki, "An algorithm for the optimization of adjustable fractional delay all-pass filters," in *Proc. IEEE ISCAS*, Vancouver, QC, Canada, May 23–26, 2006, vol. III, pp. 153–156.

[10] K. L. Teo, V. Rehbock, and L. S. Jennings, "A new computational algorithm for functional inequality constrained optimization problems," *Automatica*, vol. 29, no. 3, pp. 780–792, 1993.

[11] Z. Y. Wu, H. W. J. Lee, L. S. Zhang, and X. M. Yang, "A novel filled function method and quasi-filled function method for global optimization," *Comput. Optim. Appl.*, vol. 34, pp. 249–272, 2005.

[12] A. T. Chotterra and G. A. Jullien, "A linear programming approach to recursive digital filter design with linear phase," *IEEE Trans. Circuits Syst.*, vol. CAS-29, no. 3, pp. 139–149, Mar. 1982.

[13] C. Z. Wu, K. L. Teo, V. Rehbock, and H. H. Dam, "Global optimum design of uniform FIR filter bank with magnitude constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5478–5486, Nov. 2008.

[14] C. K. S. Pun and S. C. Chan, "Minmax design of digital all-pass filters with prescribed pole radius constraint using semidenite programming," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2003, pp. 413–416.

Two-Channel Linear Phase FIR QMF Bank Minimax Design via Global Nonconvex Optimization Programming

Charlotte Yuk-Fan Ho, Bingo Wing-Kuen Ling, Lamia Benmesbah, Ted Chi-Wah Kok, Wan-Chi Siu, and Kok-Lay Teo

***Abstract*—In this correspondence, a two-channel linear phase finite-impulse-response (FIR) quadrature mirror filter (QMF) bank minimax design problem is formulated as a nonconvex optimization problem so that a weighted sum of the maximum amplitude distortion of the filter bank, the maximum passband ripple magnitude and the maximum stopband ripple magnitude of the prototype filter is minimized subject to specifications on these performances. A modified filled function method is proposed for finding the global minimum of the nonconvex optimization problem. Computer numerical simulations show that our proposed design method is efficient and effective.**

***Index Terms*—Filled function, global optimization, nonconvex optimization problem, two-channel linear phase FIR QMF bank minimax design.**

I. Introduction

Since transition bandwidths of the filters in two-channel filter banks are usually larger than those in multichannel filter banks, lengths of the filters in two-channel filter banks are usually shorter than those in multichannel filter banks. Moreover, as only a single prototype filter is required for the design of a quadrature mirror filter (QMF) bank and all other filters are derived from the prototype filter, the total number of filter coefficients required for the design of a QMF bank is usually smaller than those in general filter banks. Furthermore, as the linear phase property of the filters guarantees no phase distortion of the filter bank and the FIR property of the filters guarantees the bounded input bounded output stability of the filter bank, two-channel linear phase FIR QMF banks find many applications in image and video signal processing [1].

Unlike a multichannel QMF bank [2], [3], a two-channel QMF bank could not achieve the exact perfect reconstruction with the prototype filter having very good frequency selectivity [4]. Hence, it is useful to

Manuscript received September 28, 2009; accepted April 18, 2010. Date of publication April 26, 2010; date of current version July 14, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jean-Christophe Pesquet. The work obtained in this correspondence was supported by a research grant (project G-YD26) from The Hong Kong Polytechnic University, the Centre for Multimedia Signal Processing, The Hong Kong Polytechnic University, the CRGC grant (project PolyU 5105\01E) from the Research Grants Council of Hong Kong, as well as a research grant from the Australian Research Council.

C. Y.-F. Ho is with the School of Mathematical Sciences, Queen Mary, University of London, London, E1 4NS, U.K. (e-mail: c.ho@qmul.ac.uk).

B. W.-K. Ling is with the School of Engineering, University of Lincoln, Brayford Pool, Lincoln, Lincolnshire, LN6 7TS, U.K. (e-mail: wling@lincoln.ac.uk).

L. Benmesbah is with the Department of Electronic Engineering, Division of Engineering, King's College London, Strand, London, WC2R 2LS, U.K. (e-mail: lamia.benmesbah@kcl.ac.uk).

T. C.-W. Kok is with Canaan Microelectronics, Kowloon, Hong Kong, China (e-mail: eekok@ieee.org).

W.-C. Siu is with the Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong, China (e-mail: enwcsiu@polyu.edu.hk).

K.-L. Teo is with the Department of Mathematics and Statistics, Curtin University of Technology, Perth, CRICOS Provider Code 00301J, Australia (e-mail: K.L.Teo@curtin.edu.au).

Color versions of one or more of the figures in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2049107

design a two-channel QMF bank so that a weighted sum of the maximum amplitude distortion of the filter bank, the maximum passband ripple magnitude and the maximum stopband ripple magnitude of the prototype filter is minimized subject to specifications on these performances. Nevertheless, this QMF bank minimax design problem is a nonconvex optimization problem. As nonconvex optimization problems usually consist of many local minima [14], it is usually stuck at these local minima and very difficult to find the global minimum if conventional gradient based approaches are employed for finding the global minimum.

When nonconvex optimization problems consist of finite numbers of local minima, it is possible to find the global minimum of these nonconvex optimization problems. There are mainly two different approaches for finding the global minimum of these nonconvex optimization problems. The first type of the approaches is nongradient based approaches, such as evolutionary algorithm based approaches [5], [6]. These approaches keep generating evaluation points randomly. Those evaluation points with better performances are kept, while those evaluation points with poor performances are ignored. However, computational complexities of these nongradient based approaches are very high because most of the evaluation points are ignored. The second type of the approaches is filled function approaches [7]–[12]. The definition of filled functions and the working principle of filled function methods are discussed in Section II. Nevertheless, it is very challenging to find a filled function that satisfies the required properties. To tackle this difficulty, filled functions with several parameters are defined [7]–[12]. However, there is no general rule for the selection of these parameters. In this correspondence, extra constraints are imposed on the optimization problems so that the required properties of the filled function are guaranteed to be satisfied.

In this correspondence, a modified filled function method is proposed for finding the global minimum of a two-channel linear phase FIR QMF bank minimax design problem. The outline of this correspondence is as follows. In Section II, the definition of filled functions and the working principle of filled function methods are reviewed. In Section III, a two-channel linear phase FIR QMF bank minimax design problem is formulated as a nonconvex optimization problem and a modified filled function method is proposed for finding the global minimum of the nonconvex optimization problem. In Section IV, computer numerical simulations are illustrated. Finally, conclusions are drawn in Section V.

II. Review on Definition of Filled Functions and Working Principle of Filled Function Methods

A filled function [7]–[12] is a function satisfying the following properties: a) the current local minimum of the original cost function is the current local maximum of the filled function; b) the whole current basin of the original cost function is a part of the current hill of the filled function; c) the filled function has no stationary point in any higher basins of the original cost function; and d) there exists a local minimum of the filled function which is in a lower basin of the original cost function.

Some terminologies related to filled functions have been used above. Notably, a basin of a function is defined as the subset of the domain of the optimization variables such that any points in this subset will give the same local minimum of the function via conventional gradient based optimization methods. A hill of a function is defined as the subset of the domain of the optimization variables such that any points in this subset will give the same local maximum of the function via conventional gradient based optimization methods. A higher basin of a function is a basin of the function with the cost value of the local minimum of the basin being higher than that of the current basin of the function. A lower basin of a function is a basin of the function with the cost value of the local minimum of the basin being lower than that of the current basin of the function.

Due to property a), by evaluating the filled function at a point slightly deviated from the current local minimum of the original cost function, a lower filled function value can be obtained. Hence, the filled function could kick away from the current local minimum of the original cost function. Due to properties b)–d), the current local minimum of the filled function is neither in the current basin nor any higher basins of the original cost function. Hence, the current local minimum of the filled function is in a lower basin of the original cost function. As a result, by finding the next local minimum of the original cost function, i.e., searching the neighborhood around the current local minimum of the filled function, a better local minimum of the original cost function can be obtained. Following these procedures, if the original cost function contains a finite number of local minima [14], then the global minimum of the original cost function will be eventually reached.

III. Problem Formulation and Modified Filled Function Method

A. Problem Formulation

Denote the transpose operator, the conjugate operator and the conjugate transpose operator as the superscripts ^{T}, * and $^+$, respectively, and the modulus operator as $|\cdot|$. Let the transfer functions of the lowpass and the highpass analysis filters of a two-channel linear phase FIR QMF bank be $H_0(z)$ and $H_1(z)$, respectively, and those of the synthesis filters of the filter bank be $F_0(z)$ and $F_1(z)$, respectively. Here, $H_0(z)$ is the transfer function of the prototype filter. Denote the impulse response of the prototype filter as $h(n)$, the passband and the stopband of the prototype filter as B_p and B_s, respectively, the length of the prototype filter as N, the maximum passband ripple magnitude and the maximum stopband ripple magnitude of the prototype filter as δ_p and δ_s, respectively, the specifications on the acceptable bounds on the maximum passband ripple magnitude and the maximum stopband ripple magnitude of the prototype filter as ε_p and ε_s, respectively, and the desired magnitude response of the prototype filter as $D(\omega)$. In this correspondence, it is assumed that the prototype filter is even length and symmetric. Let the polyphase components of $H_0(z)$ be $E_0(z^2)$ and $E_1(z^2)$, that is

$$H_0(z) \equiv E_0(z^2) + z^{-1}E_1(z^2). \tag{1}$$

Denote the transfer function of the filter bank as $T(z)$, the maximum amplitude distortion of the filter bank as δ_a, and the specification on the acceptable bound on the maximum amplitude distortion of the filter bank as ε_a. Let the vector containing these distortions and the even-time index filter coefficients be $\mathbf{x}$, that is

$$\mathbf{x} \equiv [\delta_a, \quad \delta_p, \quad \delta_s, \quad h(0), \quad h(2), \quad \cdots, \quad h(N-2)]^T. \tag{2}$$

In order to achieve both the aliasing free condition and the QMF pairs condition, the relationships among the analysis filters and the synthesis filters are governed by

$$H_1(z) = H_0(-z) \tag{3a}$$

$$F_0(z) = 2H_0(z) \tag{3b}$$

and

$$F_1(z) = -2H_0(-z). \tag{3c}$$

As the prototype filter is even length and symmetric, we have

$$H_0(z) = \sum_{n=0}^{\frac{N}{2}-1} h(2n)z^{-2n} + z^{-1}\sum_{n=0}^{\frac{N}{2}-1} h(2n)z^{-(N-2-2n)} \quad (4)$$

$$E_0(z) = \sum_{n=0}^{\frac{N}{2}-1} h(2n)z^{-n} \quad (5)$$

$$E_1(z) = \sum_{n=0}^{\frac{N}{2}-1} h(2n)z^{-(\frac{N}{2}-1-n)} = z^{-(\frac{N}{2}-1)}E_0(z^{-1}) \quad (6)$$

and

$$T(z) = 4z^{-1}E_0(z^2)E_1(z^2) = 4z^{-(N-1)}E_0(z^2)E_0(z^{-2}). \quad (7)$$

Denote

$$\boldsymbol{\eta}(\omega) \equiv \left[0, \quad 0, \quad 0, \quad 1, \quad e^{-j\omega}, \quad \cdots, \quad e^{-j(\frac{N}{2}-1)\omega}\right]^T \quad (8)$$

then

$$T(\omega) = 4e^{-j\omega(N-1)}\mathbf{x}^T\left(\boldsymbol{\eta}(2\omega)\right)^*\left(\boldsymbol{\eta}(2\omega)\right)^T\mathbf{x}. \quad (9)$$

Obviously, the filter bank does not suffer from the phase distortion and the amplitude distortion of the filter bank can be expressed as $|4\mathbf{x}^T(\boldsymbol{\eta}(2\omega))^*(\boldsymbol{\eta}(2\omega))^T\mathbf{x} - 1|$. Denote

$$\mathbf{Q}(\omega) \equiv 8\left(\boldsymbol{\eta}(2\omega)\right)^*\left(\boldsymbol{\eta}(2\omega)\right)^T \quad (10)$$

then the amplitude distortion of the filter bank can be further expressed as $|(1/2)\mathbf{x}^T\mathbf{Q}(\omega)\mathbf{x} - 1|$. Denote

$$\boldsymbol{\iota}_a \equiv [1, \quad 0, \quad \cdots, \quad 0]^T \quad (11)$$

then the constraint on the maximum amplitude distortion of the filter bank can be expressed as

$$\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\omega)\mathbf{x} - \boldsymbol{\iota}_a^T\mathbf{x} - 1 \le 0 \quad (12)$$

and

$$-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\omega)\mathbf{x} - \boldsymbol{\iota}_a^T\mathbf{x} + 1 \le 0 \quad \forall\omega \in [-\pi, \pi]. \quad (13)$$

Denote

$$\boldsymbol{\kappa}(\omega) \equiv 2\left[0, \quad 0, \quad 0, \quad \cos\left(\left(\frac{N-1}{2}\right)\omega\right), \right.$$
$$\left. \cos\left(\left(\frac{N-5}{2}\right)\omega\right), \quad \cdots, \quad \cos\left(\left(\frac{3-N}{2}\right)\omega\right)\right]^T \quad (14)$$

then

$$\begin{aligned} H_0(\omega) &= \left(\boldsymbol{\eta}(2\omega)\right)^T\mathbf{x} + e^{-j\omega(N-1)}\left(\boldsymbol{\eta}(2\omega)\right)^+\mathbf{x} \\ &= e^{-j\omega(\frac{N-1}{2})} \\ &\quad \times \left(\left[0, 0, 0, e^{j(\frac{N-1}{2})\omega}, e^{j(\frac{N-5}{2})\omega}, \cdots, e^{-j(\frac{N-3}{2})\omega}\right]\mathbf{x}\right. \\ &\quad \left. + \left[0, 0, 0, e^{-j(\frac{N-1}{2})\omega}, e^{-j(\frac{N-5}{2})\omega}, \cdots, e^{j(\frac{N-3}{2})\omega}\right]\mathbf{x}\right) \\ &= e^{-j\omega(\frac{N-1}{2})}\left(\boldsymbol{\kappa}(\omega)\right)^T\mathbf{x} \end{aligned} \quad (15)$$

and the passband ripple magnitude of the prototype filter can be expressed as $|(\boldsymbol{\kappa}(\omega))^T\mathbf{x} - D(\omega)|\forall\omega \in B_p$. Define

$$\boldsymbol{\iota}_p \equiv [0, \quad 1, \quad 0, \quad \cdots, \quad 0]^T \quad (16)$$

then the constraint on the maximum passband ripple magnitude of the prototype filter can be expressed as

$$\left|\left(\boldsymbol{\kappa}(\omega)\right)^T\mathbf{x} - D(\omega)\right| \le \boldsymbol{\iota}_p^T\mathbf{x} \quad \forall\omega \in B_p. \quad (17)$$

Define

$$\mathbf{A}_p(\omega) \equiv \left[\boldsymbol{\kappa}(\omega) - \boldsymbol{\iota}_p, \quad -\boldsymbol{\kappa}(\omega) - \boldsymbol{\iota}_p\right]^T \quad (18)$$

and

$$\mathbf{c}_p(\omega) \equiv [D(\omega), \quad -D(\omega)]^T \quad (19)$$

then the constraint on the maximum passband ripple magnitude of the prototype filter can be further expressed as

$$\mathbf{A}_p(\omega)\mathbf{x} - \mathbf{c}_p(\omega) \le \mathbf{0} \quad \forall\omega \in B_p. \quad (20)$$

Similarly, define

$$\boldsymbol{\iota}_s \equiv [0, \quad 0, \quad 1, \quad 0, \quad \cdots, \quad 0]^T \quad (21)$$

$$\mathbf{A}_s(\omega) \equiv \left[\boldsymbol{\kappa}(\omega) - \boldsymbol{\iota}_s, \quad -\boldsymbol{\kappa}(\omega) - \boldsymbol{\iota}_s\right]^T \quad (22)$$

and

$$\mathbf{c}_s(\omega) \equiv [D(\omega), \quad -D(\omega)]^T \quad (23)$$

then the constraint on the maximum stopband ripple magnitude of the prototype filter can be expressed as

$$\mathbf{A}_s(\omega)\mathbf{x} - \mathbf{c}_s(\omega) \le \mathbf{0} \quad \forall\omega \in B_s. \quad (24)$$

Define

$$\mathbf{A}_b \equiv [\mathbf{I}, \quad \mathbf{0}] \quad (25)$$

and

$$\mathbf{c}_b \equiv [\varepsilon_a, \quad \varepsilon_p, \quad \varepsilon_s]^T \quad (26)$$

in which $\mathbf{I}$ is the 3×3 identity matrix, then the specifications on the acceptable bounds on the maximum amplitude distortion of the filter bank, the maximum passband ripple magnitude and the maximum stopband ripple magnitude of the prototype filter can be expressed as

$$\mathbf{A}_b\mathbf{x} - \mathbf{c}_b \le \mathbf{0}. \quad (27)$$

In order to minimize a weighted sum of the maximum amplitude distortion of the filter bank, the maximum passband ripple magnitude and the maximum stopband ripple magnitude of the prototype filter subject to the specifications on these performances, the filter bank design problem is formulated as the following optimization problem:

Problem (P)

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) \equiv \left(\alpha\boldsymbol{\iota}_a + \beta\boldsymbol{\iota}_p + \gamma\boldsymbol{\iota}_s\right)^T\mathbf{x}, \quad (28a)$$

$$\text{subject to} \quad g_1(\mathbf{x}, \omega) \equiv \frac{1}{2}\mathbf{x}^T\mathbf{Q}(\omega)\mathbf{x} - \boldsymbol{\iota}_a^T\mathbf{x} - 1 \le 0 \quad \forall\omega \in [-\pi, \pi], \quad (28b)$$

$$g_2(\mathbf{x}, \omega) \equiv -\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\omega)\mathbf{x} - \boldsymbol{\iota}_a^T\mathbf{x} + 1 \le 0 \quad \forall\omega \in [-\pi, \pi], \quad (28c)$$

$$g_3(\mathbf{x}, \omega) \equiv \mathbf{A}_p(\omega)\mathbf{x} - \mathbf{c}_p(\omega) \le \mathbf{0} \quad \forall\omega \in B_p, \quad (28d)$$

$$g_4(\mathbf{x}, \omega) \equiv \mathbf{A}_s(\omega)\mathbf{x} - \mathbf{c}_s(\omega) \le \mathbf{0} \quad \forall\omega \in B_s \quad (28e)$$

$$\text{and} \quad g_5(\mathbf{x}) \equiv \mathbf{A}_b\mathbf{x} - \mathbf{c}_b \le \mathbf{0}, \quad (28f)$$

where α, β and γ are the weights of different criteria for formulating the cost function, $f(\mathbf{x})$ is the cost function, and $g_1(\mathbf{x}, \omega)$, $g_2(\mathbf{x}, \omega)$,

$g_3(\mathbf{x},\omega)$, $g_4(\mathbf{x},\omega)$ and $g_5(\mathbf{x})$ are the constraint functions of the optimization problem.

As the set of the filter coefficients satisfying the constraints (28b) and (28c) is nonconvex, the optimization problem is a nonconvex optimization problem. In general, it is difficult to find the global minimum of a nonconvex optimization problem.

B. Modified Filled Function Method

To find the global minimum of a nonconvex optimization problem, the following algorithm is proposed.

Algorithm

Step 1: Initialize a minimum improvement factor ε, an accepted error ε', an initial search point $\widetilde{\mathbf{x}}_1$, a positive definite matrix $\mathbf{R}$ and an iteration index $k=1$.

Step 2: Find a local minimum of the following optimization Problem $(\mathbf{P}_f)$ using our previous proposed integration approach with the initial search point $\widetilde{\mathbf{x}}_k$ [13].

$\mathbf{Problem}\ (\mathbf{P}_f)$

$$\begin{aligned}
\min_{\mathbf{x}} \quad & (28a),\\
\text{subject to} \quad & (28b)-(28f),\\
& g_6(\mathbf{x}) \equiv \boldsymbol{\iota}_a^T\,(\mathbf{x}-(1-\varepsilon)\,\widetilde{\mathbf{x}}_k) \le 0 \qquad (29a)\\
& g_7(\mathbf{x}) \equiv \boldsymbol{\iota}_p^T\,(\mathbf{x}-(1-\varepsilon)\,\widetilde{\mathbf{x}}_k) \le 0 \qquad (29b)\\
\text{and} \quad & g_8(\mathbf{x}) \equiv \boldsymbol{\iota}_s^T\,(\mathbf{x}-(1-\varepsilon)\,\widetilde{\mathbf{x}}_k) \le 0 \qquad (29c)
\end{aligned}$$

where $g_6(\mathbf{x})$, $g_7(\mathbf{x})$ and $g_8(\mathbf{x})$ are the constraint functions we imposed. Denote the obtained local minimum as $\mathbf{x}_k^*$.

Step 3: Find a local minimum of the following optimization Problem $(\mathbf{P}_H)$ using our previous proposed integration approach with the initial search point $\mathbf{x}_k^*$ [13].

$\mathbf{Problem}\ (\mathbf{P}_H)$

$$\begin{aligned}
\min_{\mathbf{x}} \quad & H(\mathbf{x}) \equiv (\alpha\boldsymbol{\iota}_a+\beta\boldsymbol{\iota}_p+\gamma\boldsymbol{\iota}_s)^T\mathbf{x}\\
& \quad + \frac{1}{(\mathbf{x}-\mathbf{x}_k^*)^T\,\mathbf{R}\,(\mathbf{x}-\mathbf{x}_k^*)} \qquad (30a)\\
\text{subject to} \quad & (28b)-(28f),\\
& g_6'(\mathbf{x}) \equiv \boldsymbol{\iota}_a^T\,(\mathbf{x}-(1-\varepsilon)\,\mathbf{x}_k^*) \le 0 \qquad (30b)\\
& g_7'(\mathbf{x}) \equiv \boldsymbol{\iota}_p^T\,(\mathbf{x}-(1-\varepsilon)\,\mathbf{x}_k^*) \le 0 \qquad (30c)\\
\text{and} \quad & g_8'(\mathbf{x}) \equiv \boldsymbol{\iota}_s^T\,(\mathbf{x}-(1-\varepsilon)\,\mathbf{x}_k^*) \le 0 \qquad (30d)
\end{aligned}$$

where $H(\mathbf{x})$ is the filled function we defined, and $g_6'(\mathbf{x})$, $g_7'(\mathbf{x})$ and $g_8'(\mathbf{x})$ are the constraint functions we imposed. Denote the obtained local minimum as $\widetilde{\mathbf{x}}_{k+1}$. Increment the value of k.

Step 4: Iterate Step 2 and Step 3 until

$$\left\|(\alpha\boldsymbol{\iota}_a+\beta\boldsymbol{\iota}_p+\gamma\boldsymbol{\iota}_s)^T\,(\mathbf{x}_k^*-\mathbf{x}_{k-1}^*)\right\| \le \varepsilon'. \qquad (31)$$

Take the final vector of $\mathbf{x}_k^*$ as the global minimum of the original optimization problem.

Step 1 is an initialization of the proposed algorithm. In order not to terminate the algorithm when the convergence of the algorithm is slow and to have a high accuracy of the solution, both ε and ε' should be chosen as small values. Also, as $\widetilde{\mathbf{x}}_1$ is an initial search point of the optimization algorithm, this initial search point should be in the feasible set. However, in general it is difficult to guarantee that $\widetilde{\mathbf{x}}_1$ is in the feasible set, it should be chosen in such a way that most of the constraints are satisfied. Moreover, as $\mathbf{R}$ is a positive definite matrix, it controls the spread of the hill of $H(\mathbf{x})$ at $\mathbf{x}_k^*$. If $\mathbf{R}$ is a diagonal matrix with all diagonal elements being the same and positive, then large values of these diagonal elements will result to a wide spread of the hill of $H(\mathbf{x})$ at $\mathbf{x}_k^*$ and vice versa. Since the local minima of nonconvex optimization problems could be located very close together [14], the spread of the hill of $H(\mathbf{x})$ at $\mathbf{x}_k^*$ should be small and the diagonal elements of $\mathbf{R}$ should be chosen as small positive numbers. Step 2 is to find a local minimum of $f(\mathbf{x})$. As the constraints (29a)–(29c) are imposed on the Problem $(\mathbf{P}_f)$, the maximum amplitude distortion of the filter bank, the maximum ripple magnitude and the maximum stopband ripple magnitude of the prototype filter corresponding to the new obtained local minimum are guaranteed to be lower than that corresponding to $\widetilde{\mathbf{x}}_k$. Similarly, Step 3 is to find a local minimum of $H(\mathbf{x})$. As the constraints (30b)–(30d) are imposed on the Problem $(\mathbf{P}_H)$, the maximum amplitude distortion of the filter bank, the maximum ripple magnitude and the maximum stopband ripple magnitude of the prototype filter are guaranteed to be lower than that corresponding to $\mathbf{x}_k^*$. Step 4 is a termination test procedure. If the difference of the weighted performance between two consecutive iterations is smaller than a certain bound ε', then the algorithm is terminated.

It has been discussed in Section I that conventional filled function methods require that a) the current local minimum of the original cost function is the current local maximum of the filled function; b) the whole current basin of the original cost function is a part of the current hill of the filled function; c) the filled function has no stationary point in any higher basins of the original cost function; and d) there exists a local minimum of the filled function which is in a lower basin of the original cost function. As $\mathbf{R}$ is a positive definite matrix and $\mathbf{x}_k^*$ is in the denominator of $H(\mathbf{x})$, $H(\mathbf{x}) \to +\infty$ as $\mathbf{x} \to \mathbf{x}_k^*$. Hence, $\mathbf{x}_k^*$ is the global maximum of $H(\mathbf{x})$ and property a) is guaranteed to be satisfied. As the constraints (30b)–(30d) are imposed on the Problem $(\mathbf{P}_H)$, when a new local minimum of $H(\mathbf{x})$ is found, this new local minimum of $H(\mathbf{x})$ will not be located at $\mathbf{x}_k^*$ and the original cost value evaluated at $\widetilde{\mathbf{x}}_{k+1}$ will guarantee to be lower than that at $\mathbf{x}_k^*$. Hence, properties b)–d) are guaranteed to be satisfied. As a result, the proposed algorithm guarantees to reach the global minimum of the nonconvex optimization problem.

As the efficiency of general nonconvex optimization algorithms would depend on the initial search points, the total number of local minima of the optimization problems and the stopping criteria of the optimization algorithms, there is always a tradeoff between the accuracy of the obtained solutions and the efficiency of the optimization algorithms. For nongradient based approaches, as most of the evaluation points are ignored, the effectiveness of these algorithms is low. On the other hand, our proposed method guarantees to obtain the local minimum in each iteration, the effectiveness of our proposed algorithm is high. Hence, for the same period of time, our proposed method would obtain a better solution than that of nongradient based approaches.

IV. Numerical Computer Simulations

In order to have a fair comparison, the performance of the QMF banks designed via our proposed method is compared to that designed via the minimax approach discussed in [4]. We choose the same passband, stopband, filter length, maximum passband ripple magnitude, maximum stopband ripple magnitude and desirable magnitude response of the prototype filter as that in [4], that is

$$B_p = [-0.4\pi, 0.4\pi] \qquad (32)$$

$$B_s = [0.6\pi, \pi] \cup [-\pi, -0.6\pi] \qquad (33)$$

$$N = 36 \qquad (34)$$

$$\varepsilon_p = -50\ \text{dB} \qquad (35)$$

$$\varepsilon_s = -50\ \text{dB} \qquad (36)$$

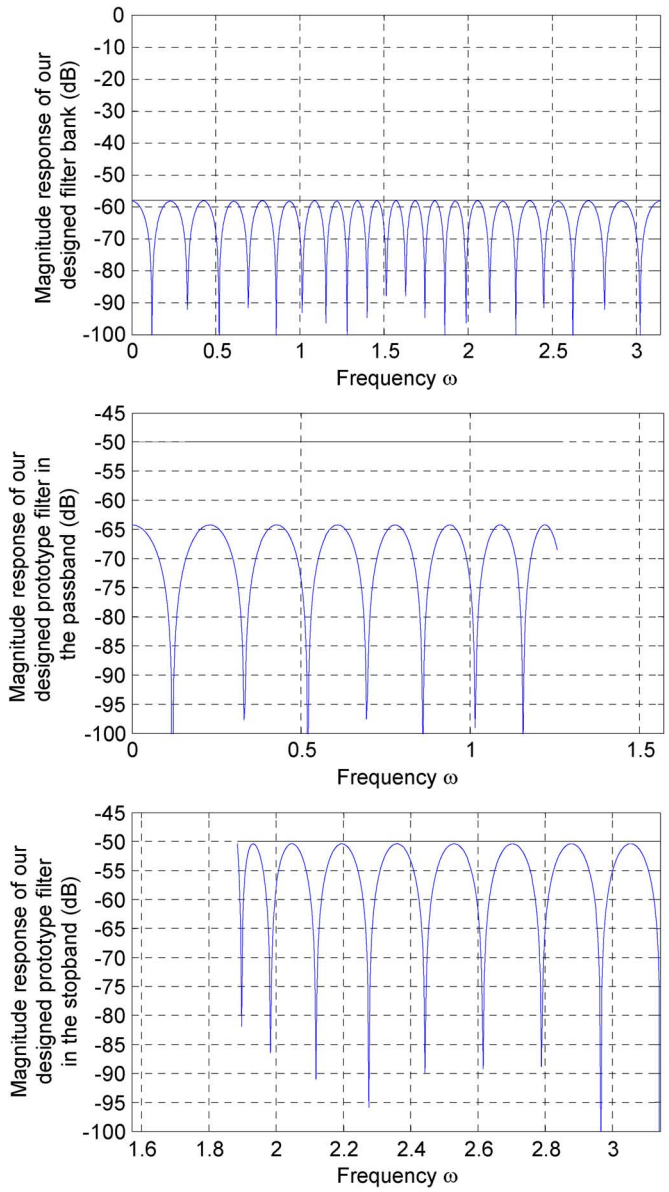

Fig. 1. (a) Magnitude response of the filter bank. (b) Magnitude response of the prototype filter in the passband. (c) Magnitude response of the prototype filter in the stopband.

and

$$D(\omega) = \begin{cases} 1 & \omega \in B_p \\ 0 & \omega \in B_s. \end{cases} \tag{37}$$

In order to guarantee that the performance of the QMF bank designed via our proposed method is better than that in [4], the specification on the maximum amplitude distortion of the filter bank is chosen as $\varepsilon_a = -58$ dB, which is better than that in [4] ($\varepsilon_a = 0.003 = -50.4576$ dB). In order not to have any bias among the maximum amplitude distortion of the filter bank, the maximum passband ripple magnitude and the maximum stopband ripple magnitude of the prototype filter, all the weights in the cost function are chosen to be the same, that is $\alpha = \beta = \gamma = 1$. In this correspondence, $\varepsilon = \varepsilon' = 10^{-6}$ are chosen which is small enough for most applications. $\widetilde{\mathbf{x}}_1$ is chosen as the filter coefficients obtained via the Remez exchange algorithm, which guarantee to satisfy the specifications on the maximum passband ripple magnitude and the maximum stopband ripple magnitude of the prototype filter. $\mathbf{R}$ is chosen as the diagonal matrix with all diagonal elements equal to 10^{-3}, which is small enough for most applications.

To compare the efficiency of the designed method, our proposed method only takes three iterations to converge and the total time required for the computer numerical simulations is 1.6 seconds. On the other hand, the method discussed in [4] takes 68 iterations to converge and the total time required for the computer numerical simulations is 80 seconds. Hence, it can be concluded that the method discussed in [4] requires more computational efforts than our proposed method and our proposed method is more efficient than that discussed in [4]. The magnitude responses of the filter banks as well as the magnitude responses of the prototype filters in both the passband and the stopband designed via our proposed method are shown in Fig. 1. It can be seen from Fig. 1 that the prototype filter designed by our proposed method could achieve $\delta_p = -64.2416$ dB and $\delta_s = -50.3625$ dB, and the QMF bank could achieve $\delta_a = -58.1557$ dB. On the other hand, the prototype filter designed by the Remez exchange algorithm could achieve $\delta_p = -62.7693$ dB and $\delta_s = -62.6164$ dB, and the QMF bank could only achieve $\delta_a = -5.8842$ dB. It can be checked easily that the QMF bank designed via our proposed method achieves better performances on the maximum amplitude distortion of the filter bank, the maximum passband ripple magnitude and the maximum stopband ripple magnitude ripple of the prototype filter than that designed by the method discussed in [4]. This is because the QMF bank designed by the method discussed in [4] is not the global minimum, while that designed by our proposed method is the global minimum.

V. Conclusion

This correspondence proposes a modified filled function method for the design of a two-channel linear phase FIR QMF bank so that a weighted sum of the maximum amplitude distortion of the filter bank, the maximum passband ripple magnitude and the maximum stopband ripple magnitude of the prototype filter is minimized. The proposed method could find the global minimum of the nonconvex optimization problem efficiently.

References

[1] J. D. Johnston, "A filter family designed for use in quadrature mirror filter banks," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Apr. 1980, vol. 5, pp. 291–294.

[2] Y.-P. Lin and P. P. Vaidyanathan, "Linear phase cosine modulated maximally decimated filter banks with perfect reconstruction," in *Proc. Int. Symp. Circuits Systems (ISCAS)*, May 30–Jun. 2, 1994, vol. 2, pp. 17–20.

[3] T. Q. Nguyen, "Near-perfect-reconstruction pseudo-QMF banks," *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 65–76, Jan. 1994.

[4] C.-W. Kok, W.-C. Siu, and Y.-M. Law, "Peak constrained least-squares QMF banks," *Signal Process.*, vol. 88, pp. 2363–2371, 2008.

[5] P. Samadi and M. Ahmadi, "Genetic algorithm and its application for the design of QMF banks with canonical signed digit coefficients: A comparative study and new results," in *Proc. IEEE Northeast Workshop Circuits Syst.*, Aug. 5–6, 2007, pp. 357–360.

[6] H. Uppalapati, H. Rastgar, M. Ahmadi, and M. A. Sid-Ahmed, "Design of quadrature mirror filter banks with canonical signed digit coefficients using genetic algorithm," in *Proc. Int. Conf. Communications, Circuits, Systems*, May 27–30, 2005, vol. 2, pp. 682–686.

[7] Y. Zhang, Liansheng, and Y. Xu, "New filled functions for nonsmooth global optimization," *Appl. Math. Model.*, vol. 33, pp. 3114–3129, 2009.

[8] X. Liu, "Finding global minima with a computable filled function," *J. Global Optim.*, vol. 19, pp. 151–161, 2001.

[9] K. F. C. Yiu, Y. Liu, and K. L. Teo, "A hybrid descent method for global optimization," *J. Global Optim.*, vol. 28, pp. 229–238, 2004.

[10] Z. Y. Wu, H. W. J. Lee, L. S. Zhang, and X. M. Yang, "A novel filled function method and quasi-filled function method for global optimization," *Comput. Optim. Appl.*, vol. 34, pp. 249–272, 2005.

[11] R. P. Ge and Y. F. Qin, "A class of filled functions for finding global minimizers of a function of several variables," *J. Optim. Theory Appl.*, vol. 54, no. 2, pp. 241–252, 1987.

[12] R. Ge, "A filled function method for finding a global minimizer of a function of several variables," *Math. Program.*, vol. 46, pp. 191–204, 1990.
[13] C. Y.-F. Ho, B. W.-K. Ling, Z.-W. Chi, M. Shikh-Bahaei, Y.-Q. Liu, and K.-L. Teo, "Design of near-allpass strictly stable minimal-phase real-valued rational IIR filters," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 55, no. 8, pp. 781–785, 2008.
[14] K. F. C. Yiu, N. Grbić, S. Nordholm, and K.-L. Teo, "A hybrid method for the design of oversampled uniform DFT filter banks," *Signal Process.*, vol. 86, no. 7, pp. 1355–1364, 2006.

Prediction-Based Incremental Refinement for Binomially-Factorized Discrete Wavelet Transforms

Yiannis Andreopoulos, *Member, IEEE*, Dai Jiang, *Member, IEEE*, and Andreas Demosthenous, *Senior Member, IEEE*

***Abstract*—It was proposed recently that quantized representations of the input source (e.g., images, video) can be used for the computation of the two-dimensional discrete wavelet transform (2D DWT) *incrementally*. The coarsely quantized input source is used for the initial computation of the forward or inverse DWT, and the result is successively refined with each new refinement of the source description via an embedded quantizer. This computation is based on the direct two-dimensional factorization of the DWT using the generalized spatial combinative lifting algorithm. In this correspondence, we investigate the use of prediction for the computation of the results, i.e., exploiting the correlation of neighboring input samples (or transform coefficients) in order to reduce the dynamic range of the required computations, and thereby reduce the circuit activity required for the arithmetic operations of the forward or inverse transform. We focus on binomial factorizations of DWTs that include (amongst others) the popular 9/7 filter pair. Based on an FPGA arithmetic co-processor testbed, we present energy-consumption results for the arithmetic operations of incremental refinement and prediction-based incremental refinement in comparison to the conventional (nonrefinable) computation. Our tests with combinations of intra and error frames of video sequences show that the former can be 70% more energy efficient than the latter for computing to half precision and remains 15% more efficient for full-precision computation.**

***Index Terms*—Approximate signal processing, discrete wavelet transform, energy consumption, incremental refinement of computation, lifting scheme.**

I. Introduction

The two-dimensional discrete wavelet transform (2D DWT) has been established as one of the main tools for image compression [1], image denoising and other popular image processing operations [2], [18]. In the vast majority of applications, the transform coefficients are produced to the maximum degree of precision and then they are quantized and processed as appropriate [1]. However, it has been recognized that this wastes system resources for the cases where severe quantization would render the majority of the coefficients not being used at all, or used at very low precision [3]. For example, this is commonly the case for low-bitrate image and video coding applications [3] and resource-constrained image and video processing operations [4]. For this reason, previous work proposed schemes for approximate computation of transforms and signal processing operations [3]. A property that has been recognized to be of great importance is incremental refinement of computation [4]–[6], where the transform representation of a signal (image, video) is produced incrementally with the use of embedded (bitplane-based) quantization. In our recent work [4], this design has been theoretically analyzed both for the forward and the inverse two-dimensional multilevel DWT using the generalization of the spatial combinative lifting algorithm (SCLA) of Meng and Wang [7]. The overall framework is depicted in Fig. 1. There, the multilevel DWT decomposition of the input source (video frame) occurs independently for each quantization threshold (bitplane), starting from the most-significant bitplane (MSB) and going down to the least-significant bitplane (LSB). The results are accumulated after each multilevel SCLA computation to form an incrementally-refined output. Similarly, for the DWT reconstruction, the MSBs of the transform-coefficients are inserted first and the multilevel inverse DWT reconstructs the image incrementally. Each additional processing step requires additional energy consumption. If the processing resources are terminated, one receives the decomposed or reconstructed image with the best-possible quality (controlled by the number of bitplanes processed).

Although the framework of Fig. 1 receives individual bits (per pixel or per wavelet coefficient), the dynamic range of computations performed is increasing according to i) the lifting coefficients of each lifting step; ii) the input-source statistics; and iii) the number of decomposition levels. In order to adapt the circuit activity according to varying input statistics, it is crucial to have arithmetic units that perform variable dynamic-range computation. Xanthopoulos [8] proposed a suitable framework for this purpose: for all arithmetic units, very low-cost "MSB-rejection" circuits are utilized, which identify the exact number of active bits within each element (adder or multiplier). In this correspondence, we use a "zero-detection" circuit to avoid performing parts of multiplications with zero inputs and demonstrate its effectiveness in conjunction with incremental computation on an FPGA arithmetic co-processor testbed.

The contribution of this correspondence is twofold: firstly, we propose incremental computation of the DWT with the use of prediction within each refinement layer (bitplane) of the input (Sections II and III); in addition, via the utilized FPGA co-processor (introduced Section IV), we demonstrate the energy-distortion scalability offered by incremental computation and the proposed prediction-based incremental computation in comparison to the conventional (nonrefinable) computation (Section V). Our results are relevant to DWT architectures localizing memory accesses to on-chip memory [9], [10], or to cases when the entire image can be stored on-chip, since energy consumption stems predominantly from arithmetic operations and not memory accesses in such cases [9], [10].

II. Overview of SCLA-Based DWT Under Incremental Refinement of Computation

The 2-D DWT of an $R \times C$ input matrix $\mathbf{X}$ consisting of image intensity values is expressed in the spatial domain by[1] $\mathbf{S} = \mathbf{E} \cdot \mathbf{X} \cdot \mathbf{E}^T$, where $\mathbf{E}$ is the analysis polyphase matrix consisting of alternating rows of low- and high-pass filters shifted by two (in order to apply the DWT downsampling), and $\mathbf{S}$ the 2-D matrix of output wavelet coefficients.

Manuscript received October 17, 2009; accepted March 26, 2010. Date of publication April 22, 2010; date of current version July 14, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Cédric Richard. This work was supported by the EPSRC Grant EP/F020015/1.

The authors are with the University College London, Department of Electronic and Electrical Engineering, WC1E 7JE, London, U.K. (e-mail: iandreop@ee.ucl.ac.uk; djiang@ee.ucl.ac.uk; a.demosthenous@ee.ucl.ac.uk).

Digital Object Identifier 10.1109/TSP.2010.2048707

[1]We are not concerned with the scaling performed after the lifting analysis and before the lifting synthesis [11] because all scaling factors can be incorporated into the subsequent encoding or processing stage [12]. In addition, for notational simplicity, we assume that the image dimensions are integer multiples of $2^{L_{\max}}$, with $L_{\max}$ the number of wavelet decomposition levels.