DESIGN OF FIR PREDICTORS AS A TAPPED CASCADED INTERCONNECTION OF SUBFILTERS

*Olli Vainio**
Tampere University of Technology
Department of Information Technology
FIN-33101 Tampere, Finland
Tel: +358 3 365 2928; fax: +358 3 365 3095
e-mail: olli.vainio@tut.fi

ABSTRACT

The use of computationally efficient substructures for predictive FIR filters is proposed. When recursive running-sum blocks are used as subfilters, only a small number of multipliers are needed for predictor implementation. A design algorithm is described, based on using time-domain constraints to define the composition of the impulse response for least-squares optimization. Polynomial and sinusoidal predictors are given as examples, and are compared with the direct form structures.

1 INTRODUCTION

By predictors we in this paper mean Finite Impulse Response (FIR) filters which are capable of unbiased extrapolation of certain signal classes, such as polynomials and sinusoidal signals [1][2]. Such filters find applications, for example, in velocity measurements [3][4], control systems [5], data smoothing [6], and hybrid nonlinear filters [7].

The impulse response of such predictors can be analytically optimized to minimize the noise power gain of the filter. Several approaches have been described to reduce the multiplication rate required to achieve a certain noise attenuation. Recursive structures for generating the polynomial-shaped impulse response were proposed by Campbell and Neuvo [8]. The Interpolated FIR (IFIR) principle is also applicable to predictor design [9]. The use of prefilters was suggested by Laakso and Ovaska [10], making it possible to enhance noise attenuation, for example, with a simple recursive prefilter. Recursive extension of the predictor itself is also possible, introducing poles in the transfer function which is beneficial for noise suppression [11]. The poles are typically located close to the unit circle, which causes problems for finite word-length implementation.

In this paper, we describe a predictor design approach, where the unit delays of the direct-form FIR structure are replaced by multiplierless substructures. A long impulse response with favorable noise attenuation properties is then achieved with only a few actual coefficients. The structure is described in Section 2 of this paper. The design algorithm is explained in Section 3 for polynomial predictors. Sine predictor design is discussed in Section 4, and a structurally generalized approach is presented in Section 5.

* This work was supported by the Academy of Finland under grant 51849.

2 PREDICTORS WITH SUBSTRUCTURES

The idea of using a tapped cascade of identical subfilters for constructing linear-phase FIR filters is well known [12][13]. Predictors, on the other hand, are nonlinear-phase filters, and the design objectives are typically given as time-domain specifications together with noise attenuation requirements. The time-domain specification is of the form

$$x(n+p) = \sum_{k=0}^{N} h(k)x(n-k), \tag{1}$$

where p is the prediction step length and the $h(k)$'s are the terms of the impulse response. Most often the design goal is to minimize the noise gain,

$$\mathrm{NG} = \sum_{k=0}^{N} |h(k)|^2. \tag{2}$$

We propose to construct FIR predictors in the form of a tapped cascaded interconnection of identical subfilters as shown in Fig. 1(a). There are K tap coefficients h_k, $k = 0, 1, \cdots, K$. The structure of the recursive running-sum based multiplierless subfilter $R(z)$ is shown in Fig. 1(b). The structure is recursive, but when implemented with modulo arithmetic, such as two's complement arithmetic, it has the finite-length impulse response $\mathbf{r}$ consisting of a zero followed by M ones [13]. Here the subfilters do not include any scaling coefficients, instead, the necessary scaling is included in the tap coefficients h_k. The impulse response of the overall structure is of the form

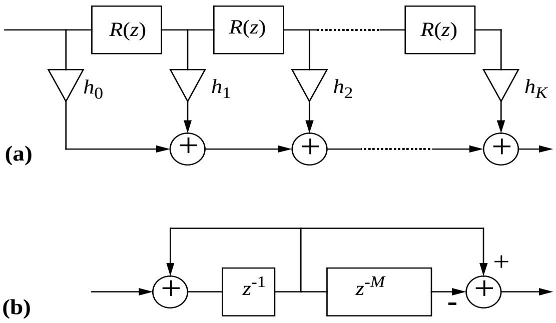

Figure 1: (a) The structure with subfilters. (b) Implementation of the subfilter $R(z)$.

$$h(k) = h_0 + h_1 r(k) + h_2 r(k) * r(k) + \cdots + h_K \overbrace{r(k) * r(k) * \cdots * r(k)}^{\text{multiplicity } K} \quad (3)$$

where $*$ denotes convolution.

3 POLYNOMIAL PREDICTORS

Polynomial predictors are designed to satisfy the constraint (1) for Lth-order polynomials,

$$x(n) = a_0 + a_1 n + \cdots + a_L n^L. \quad (4)$$

Optimum designs in the least-squares sense are obtained as

$$\mathbf{h} = \mathbf{P}(\mathbf{P}^{\mathrm{T}}\mathbf{P})^{-1}\mathbf{d}_P, \quad (5)$$

where $\mathbf{P}$ is a matrix of the form

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 4 & \cdots & N^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & N & N^2 & \cdots & N^L \end{pmatrix},$$

and

$$\mathbf{d}_P = [1 \;\; -p \;\; (-p)^2 \;\; \cdots \;\; (-p)^L]^{\mathrm{T}}.$$

Such a solution has $N+1$ coefficients, and it is required that $N \geq L$. As suggested in [1], additional linear time-domain constraints can be specified on the impulse response. This possibility is exploited to specify how the impulse response $\mathbf{h} = [h(0) \; h(1) \; \cdots \; h(N)]$ is composed from the tap coefficients h_k, $k = 0, 1, \cdots, K$. However, we cannot give more than $N - L$ structural constraints, in addition to the $L+1$ prediction constraints, to keep the equation solvable. The only unconstrained term of the impulse response (3) is $h(0)$, as $h(0) = h_0$, and h_0 does not affect the rest of the terms. Furthermore, $h(1) = h_1$, but h_1 also affects the $M-1$ next terms. Thus, h_1 should be included in the structural constraints as an input parameter but it should also be a free parameter to make optimization possible (for $L > 0$). This dilemma can be solved by programming a loop where $\hat{h}_1$ is given as an stepped input parameter, and checking for the situation where the h_1 of the optimum solution coincides with the given $\hat{h}_1$. Thus both $h(0)$ and $h(1)$ are effectively unconstrained parameters in optimization, and ramp predictors ($L = 1$) can be designed. For higher L, the structure is modified as shown in Section 5.

Because of the typically small number of the tap coefficients h_k, variable step-size exhaustive grid search can be used for finding the optimum. The search space is limited by the targeted NG value. NG will be slightly higher than $\mathrm{NG}_{\mathrm{opt}}$, which is the NG of an optimum direct-form predictor with $N+1$ coefficients. As a rule of thumb, $|h(i)| < \sqrt{2\mathrm{NG}_{\mathrm{opt}}}$, $i = 0, 1, \cdots, N$. This sets the limits for the search parameters $\hat{h}_i$, $i = 1, 2, \cdots, K$, according to the composition of the impulse response.

The design algorithm can now be outlined as follows.
Step 1: Construct the matrix $\mathbf{P}$ and vector $\mathbf{d}_P$ to include both the prediction constraints and the impulse response composition constraints.
Step 2: Find the limits of the search space for the coefficients h_i, $i = 1, 2, \cdots, K$. Set $\mathrm{NG}_{\mathrm{min}}$ to a high initial value. Set the error tolerance ϵ and the coefficient step sizes to 'practical' values.
Step 3: Loop the parameters $\hat{h}_i$ through the search space. Replace the previous solution candidate by the present candidate if $\mathrm{NG} < \mathrm{NG}_{\mathrm{min}}$ and $|\hat{h}_1 - h_1| < \epsilon$. Update $\mathrm{NG}_{\mathrm{min}}$ when a better candidate is found.
Step 4: When a feasible solution has been found, refine the optimum solution to desired accuracy by using a smaller step size and error tolerance.

As an example, consider the case $L = 1$, $K = 3$, $M = 8$, $N = 24$, and $p = 1$. In this case we have:

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 3 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 4 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 23 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 24 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

$$\mathbf{d}_P = [1 \;\; -p \;\; h_1{+}h_2 \;\; h_1{+}2h_2{+}h_3 \;\; h_1{+}3h_2{+}3h_3 \;\; \cdots \;\; h_3]^{\mathrm{T}}.$$

The amplitude responses of the optimum subfilter-based predictor and the direct-form predictor are shown in Fig. 2. The respective noise gains are 0.2343 and 0.1700. The direct-form structure requires 20 coefficients for comparable NG. The new design needs only four multipliers.

4 SINE PREDICTORS

Predictors for sinusoidal signals are designed using the single-frequency signal model

$$x(n) = A \sin(\omega_0 n + \phi), \quad (6)$$

where A and ϕ are arbitrary constants and ω_0 is the angular frequency. Subfilter-based sine predictors can be designed using a similar procedure as for polynomial

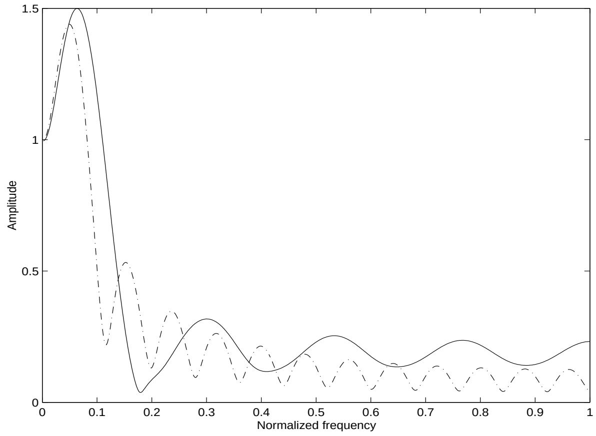

Figure 2: Amplitude responses of the subfilter-based ramp predictor (solid line) and the optimum predictor of length 25 (dash-dotted).

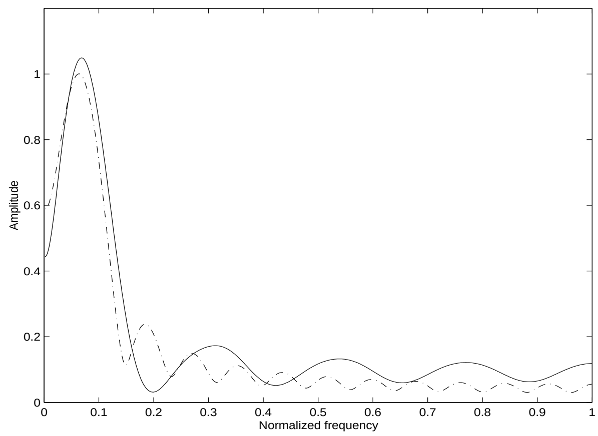

Figure 3: Amplitude responses of the subfilter-based sine predictor (solid line) and the optimum predictor of length 25 (dash-dotted).

predictors, however, the subfilter should be selected so that the frequency ω_0 is located on the 'passband' of the subfilter.

4.1 Lowpass Case

The recursive running-sum subfilter of Fig. 1(b) has a sinc function type amplitude response, with the first zero located at $2\pi/M$. The -10 dB cutoff frequency is about $3\pi/2M$, which sets a practical upper limit for ω_0.

The constraint matrix and vector are of the form [1]

$$\mathbf{S} = \begin{pmatrix} \sin(0) & \cos(0) & \cdots \\ \sin(\omega_0) & \cos(\omega_0) & \cdots \\ \sin(2\omega_0) & \cos(2\omega_0) & \cdots \\ \vdots & \vdots & \ddots \\ \sin(N\omega_0) & \cos(N\omega_0) & \cdots \end{pmatrix},$$

and

$$\mathbf{d}_S = [\sin(-p\omega_0)\ \cos(-p\omega_0)\ \cdots]^{\mathrm{T}}.$$

The composition of the impulse is specified in a similar way as for the polynomial predictors. As an example, consider the case $\omega_0 = 0.06\pi$, $K = 3$, $M = 8$, $N = 24$, and $p = 1$. The structure is thus the same as in the previous example. The amplitude responses of the optimum subfilter-based predictor and the direct-form predictor are shown in Fig. 3. The respective noise gains are 0.0939 and 0.0872. The direct-form structure requires 20 coefficients for comparable NG.

4.2 Bandpass Case

For higher frequencies, i.e., $\omega_0 > 3\pi/2M$, the subfilter needs to be modified so that it passes ω_0 without too much attenuation. Multiplierless subfilter implementations are still possible in many cases. As an example, we design a predictor for $\omega_0 = 0.4\pi$, using a subfilter with the impulse response $\mathbf{r} = [0\ 1\ 1\ -1\ -1\ 1\ 1\ -1\ -1]$. The amplitude response of the resulting predictor is shown in Fig. 4, together with the amplitude responses of the corresponding optimum predictor and the subfilter.

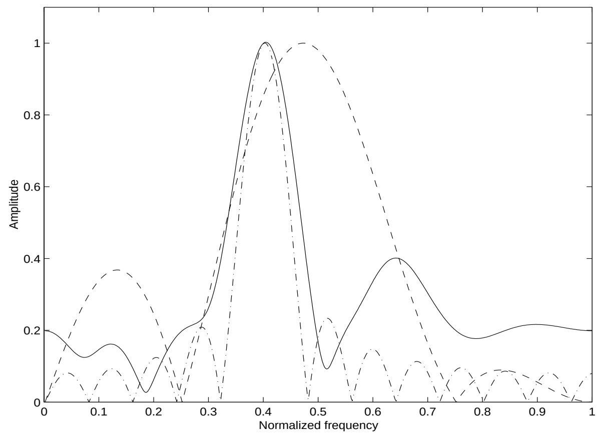

Figure 4: Amplitude responses of the subfilter-based sine predictor (solid line) and the optimum predictor of length 25 (dash-dotted). The respective noise gains are 0.1397 and 0.0800. The amplitude response of the subfilter $R(z)$ is also shown (dashed), scaled to unity peak gain.

5 A GENERALIZED STRUCTURE

From the predictor design point of view, the structure of Fig. 1 is rather restricted, since only two prediction constraints can be specified when using the proposed semi-analytical design procedure. Prediction is therefore limited to ramp signals ($L = 1$) or single-frequency sinusoids. A more flexible structure is needed for higher order polynomials, multiple sinusoids, and combinations thereof. Such a structure can be constructed as a hybrid of the direct-form structure and the subfilter-based approach, as shown in Fig. 5. The first J subfilters have been replaced by unit delays, which means that there are no structural constraints on the coefficients h_0 to h_J, and the corresponding degrees of freedom can be designated for prediction specifications.

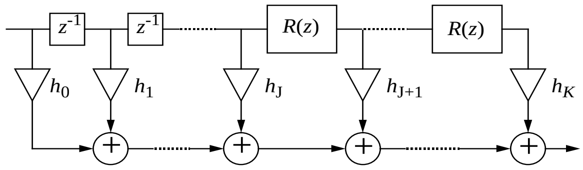

Figure 5: The hybrid structure with subfilters.

Consider the following example. We wish to design a one step-ahead predictor for $\omega_0 = 0.06\pi$, while fully suppressing the third harmonic frequency at $\omega_1 = 0.18\pi$. The design is carried out using $J = 3$, $K = 5$, and $M = 8$. The two subfilters are of the basic running-sum type of Fig. 1(b). Four degrees of freedom are needed for the prediction constraints, as we need to specify one step-ahead prediction for ω_0 and zero output prediction for ω_1, i.e., the predictor frequency response must have a notch at ω_1. 16 additional constraints are given for the terms of impulse response, specifying how the terms $h(4)$ to $h(19)$ are composed of the coefficients h_4 and h_5. In this case, it is not necessary to test the candidate designs for $|\hat{h}_4 - h_4| < \epsilon$, since there are enough degrees of freedom to simply set the constraint $h(4) = h_4$. No constraints are given for $h(0)$ to $h(3)$ since obviously $h(0) = h_0$, $h(1) = h_1$, $h(2) = h_2$, and $h(3) = h_3$. Thus, only h_4 and h_5 are searched for numerically.

The amplitude responses of the resulting optimized predictor and the running-sum subfilter are shown in Fig. 6. The noise gain of the predictor is 0.1048.

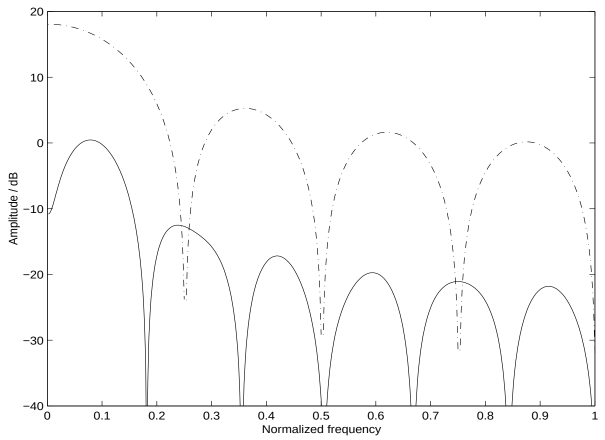

Figure 6: Amplitude responses of the predictor (solid line) and the running-sum subfilter (dash-dotted).

6 CONCLUSIONS

The subfilter-based approach is well suitable for designing computationally efficient predictors, since good noise attenuation is achieved with a small number of multipliers. The hybrid structure makes it possible to include an arbitrary number of prediction specifications in the design. The proposed semi-analytical design procedure works reasonably well in typical cases. Better design methods are needed for cases where the number of nontrivial subfilters is large.

References

[1] K. Koppinen, J. Yli-Hietanen, and J. Astola, "Optimization of generalized predictors," in *Proc. IEEE Instrumentation and Measurement Technology Conf.*, Ottawa, Canada, May 1997, pp. 54-59.

[2] S. Väliviita, S. J. Ovaska, and O. Vainio, "Polynomial predictive filtering in control instrumentation: a review," *IEEE Transactions on Industrial Electronics*, vol. 46, pp. 876-888, no. 5, Oct. 1999.

[3] J. Pasanen, O. Vainio, and S. J. Ovaska, "Predictive synchronization and restoration of corrupted velocity samples," *Measurement*, vol. 13, pp. 315-324, July 1994.

[4] O. Vainio and S. J. Ovaska, "Tachometer signal smoothing with analog discrete-time polynomial estimators," *IEEE Transactions on Industrial Electronics*, vol. 41, pp. 147-154, April 1994.

[5] S. J. Ovaska and O. Vainio, "Predictive compensation of time-varying computing delay on real-time control systems," *IEEE Transactions on Control Systems Technology*, vol. 5, no. 5, pp. 523-526, Sept. 1997.

[6] P. Steffen, "On digital smoothing filters: a brief review of closed form solutions and two new approaches," *Circuits, Syst., Signal Process.*, vol. 5, pp. 187-210, 1986.

[7] P. Heinonen and Y. Neuvo, "FIR-median hybrid filters with predictive FIR substructures," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 892-899, June 1988.

[8] T. G. Campbell and Y. Neuvo, "Predictive FIR filters with low computational complexity," *IEEE Trans. Circuits. Syst.*, vol. 38, pp. 1067-1071, Sept. 1991.

[9] O. Vainio, "Predictors based on interpolated FIR filters," in *Proc. Seventh European Signal Processing Conference*, Edinburgh, Scotland, Sept. 1994, pp. 1744-1747.

[10] T. I. Laakso and S. J. Ovaska, "Prefiltering approach for optimal polynomial prediction," *IEEE Transactions on Signal Processing*, vol. 44, pp. 701-705, March 1996.

[11] S. J. Ovaska, O. Vainio, and T. I. Laakso, "Design of predictive IIR filters via feedback extension of FIR forward predictors," *IEEE Transactions on Instrumentation and Measurement*, vol. 46, pp. 1196-1201, Oct. 1997.

[12] J. F. Kaiser and R. W. Hamming, "Sharpening the response of a symmetric nonrecursive filter by multiple use of the same filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-25, pp. 415-422, Oct. 1977.

[13] T. Saramäki, "Design of FIR filters as a tapped cascaded interconnection of identical subfilters," *IEEE Trans. Circuits. Syst.*, vol. CAS-34, pp. 1011-1029, Sept. 1987.