Optimizing Prototype Filters for TDAC Transforms

Ryan Wixen

June 2024

Abstract

Time domain aliasing cancellation (TDAC) transforms are ubiquitous for time-frequency mapping in audio coding. TDAC transforms are filterbanks in which each bandpass filter is created by modulating a single lowpass "prototype" filter, which we optimize to achieve the best coding gain in compression. Our objective is to minimize the prototype filter's stopband energy, a convex quadratic, but the prototype filter must satisfy quadratic equality constraints to achieve perfect reconstruction, so the optimization problem is not convex. However, it can be optimized locally with warm starts. We implemented two solutions, unconstrained quasi-Newton optimization with a reparameterization that eliminates the constraints and primal-dual gradient descent with a quasi-linearization of the constraints. Both solutions are capable of finding satisfactory filters with lengths in the thousands, and while the quasi-linearization is simpler to implement, the reparameterization finds more precise solutions and is therefore preferred.

1 Introduction

1.1 Background

Time domain aliasing cancellation (TDAC) transforms are ubiquitous in audio coding for mapping between time and frequency to enable compression via frequency domain quantization [2]. In order to provide the best coding gains in compression, TDAC transforms separate a time domain signal into frequency subbands that are as isolated from one another as possible. The TDAC transform can be viewed as a bank of bandpass filters that are all modulations of one prototype lowpass filter [5]. So, to achieve the best coding gains, we optimize the the lowpass properties of prototype filter.

The TDAC transform can also be viewed as modified discrete cosine transforms (with fewer outputs than inputs) applied to overlapping windowed blocks of input [3, 4]. The transform window is the same as prototype filter, and it must meet certain conditions in order to perfectly reconstruct the original signal from its transformed representation. Specifically, pairs of the prototype filter's polyphase components depending on the number of channels must be paraunitary[1]. In other words, the polyphase component pairs' autocorrelations must add to an impulse. This constraint is not convex, nor log-log convex. Luckily, the family of paraunitary filter pairs can be parameterized by a set of Givens rotation angles. In terms of these parameters, the problem is unconstrained [8].

Within the set of prototype filters, the literature suggests to minimize the filter's stopband energy [6, 8]. This objective is convex with respect to the filter coefficients but not with respect to the rotation parameters. The objective can still be locally optimized with proper initialization of the parameters, which can be done hierarchically, adding one parameter at a time and using the previous layer's result as the starting point for the current layer.

1.2 Problem Definition

A TDAC transform with M frequency channels and extension factor m has a prototype filter $p_0 \in \mathbb{R}^{2mM}$. To optimize the prototype filter, the objective we will minimize is a convex quadratic, the filter's stopband energy,

$$\begin{aligned}\Phi_1(p_0) &= \int_{\omega_s}^{\pi} |P_0(e^{j\omega})|^2 d\omega \\ &= p_0^T \left((\pi - \omega_s) I - \sum_{k=1}^{2mM-1} \frac{\sin(k\omega_s)}{k} (U^k + L^k) \right) p_0,\end{aligned}$$

where $U = \begin{bmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & & \ddots & 1 & \\ & & & & 0 & 1 \\ & & & & & 0 \end{bmatrix}$ is the upper shift matrix, $L = U^T$ is the lower shift matrix, $P_0(e^{j\omega})$ is the discrete time Forier transform of p_0, and ω_s is the cutoff frequency. The cutoff frequency ω_s usually ranges from $\frac{\pi}{2M}$ to about $\frac{\pi}{M}$ and controls a trade-off between the filter's stopband attenuation and main lobe width.

We place two constraints on p_0, perfect reconstruction and linear phase. The linear phase constraint is an affine equality constraint that simply requires that the prototype is equal to its time-reversal:

$$p_0[n] = p_0[2mM - 1 - n], \quad n = 0, \ldots, mM - 1$$

On the other hand, the perfect reconstruction requirement is non-convex, a quadratic equality constraint. With the downsampling matrix

$$\downarrow_{2M} = \begin{bmatrix} e_1 & & \\ & e_1 & \\ & & \ddots \end{bmatrix}^T$$

[1]See Smith's page about the paraconjugate [7].

where $e_1 = (1, 0, 0, \dots) \in \mathbb{R}^{2M}$ is the first standard basis, the perfect reconstruction requirement is,

$$\frac{1}{2} p_0^T \left((\downarrow_{2M} U^i)^T (U^k + L^k) \downarrow_{2M} U^i + (\downarrow_{2M} U^{i+M})^T (U^k + L^k) \downarrow_{2M} U^{i+M} \right) p_0 = \delta[k]$$

for $i = 0, \dots, M-1$ and $k = 0, \dots, m-1$. The perfect reconstruction constraint can be more simply expressed in terms of the prototype filter's polyphase components $g_i = \downarrow_{2M} U^i p_0$, or $g_i[n] = p_0[i + 2Mn]$, as

$$\frac{1}{2} \left(g_i^T (U^k + L^k) g_i + g_{i+M}^T (U^k + L^k) g_{i+M} \right) = \delta[k]$$

for $i = 0, \dots, M-1$, and $k = 0, \dots, m-1$. We addressed this non-convex constraint with two different optimization approaches.

2 Methods

2.1 Approach 1: Paraunitary Reparameterization

We can eliminate the non-convex quadratic equality constraint by expressing the prototype as a set of paraunitary lattices, which completely parameterize the family of perfect reconstruction prototype filters [6, 8]. The m coefficients in the i-th paraunitary lattice, $\beta_{i,j}$ for $i = 0, \dots, \lfloor \frac{M}{2} \rfloor - 1$ and $j = 0, \dots, m-1$, determine a pair of polyphase components that necessarily satisfy the perfect reconstruction requirement. With an overall gain, $\alpha_i = \prod_{j=0}^{m-1} 1/\sqrt{1 + \beta_{i,j}^2}$, the polyphase component pairs are parameterized in terms of the lattice coefficients as

$$\begin{pmatrix} G_i(z) \\ G_{i+M}(z) \end{pmatrix} = \alpha_i \begin{pmatrix} \beta_{i,m-1} & z^{-1} \\ 1 & -\beta_{i,m-1} z^{-1} \end{pmatrix} \cdots \begin{pmatrix} \beta_{i,1} & z^{-1} \\ 1 & -\beta_{i,1} z^{-1} \end{pmatrix} \begin{pmatrix} \beta_{i,0} \\ 1 \end{pmatrix} \quad (1)$$

for $i = 0, \dots, \lfloor \frac{M}{2} \rfloor - 1$. The rest of $2M$ polyphase coefficients are determined by the linear phase requirement. For odd M, the $\lceil \frac{M}{2} \rceil$-th polyphase component is fixed as $G_{\frac{M-1}{2}} = \sqrt{0.5} z^{-\lfloor \frac{M}{2} \rfloor}$. With the paraunitary reparameterization, we can perform an unconstrained optimization of the objective over the lattice coefficients.

Objective in Terms of Lattice Parameters We will express the objective function in terms of the lattice coefficients through a series of substitutions. The objective is a quadratic of the prototype filter,

$$\Phi_1(p_0) = p_0^T P p_0,$$

with

$$P = (\pi - \omega_s) I - \sum_{k=1}^{2mM-1} \frac{\sin(k\omega_s)}{k} (U^k + L^k).$$

The prototype filter can in turn be expressed in terms of the polyphase components as

$$p_0 = \sum_{i=0}^{2M-1} L^i \uparrow_{2M} g_i,$$

with the lower shift matrix L as defined above and the upsampling matrix $\uparrow_{2M}$ defined as

$$\uparrow_{2M} = \begin{bmatrix} e_1 & & \\ & e_1 & \\ & & \ddots \end{bmatrix},$$

where $e_1 = (1, 0, 0, \dots) \in \mathbb{R}^{2M}$ is the first standard basis. Finally, the polyphase components are determined by the lattice parameters as in equation 1.

Gradient with Respect to Lattice Parameters Having expressed the objective in terms of the parameters, we will use the chain rule to compute the gradient. To begin, we write the gradient of the objective with respect to the prototype.

$$\nabla_{p_0} \Phi_1(p_0) = 2Pp_0$$

Next, we have the Jacobian of the prototype filter with respect to each polyphase component g_i.

$$\nabla_{g_i} p_0 = (L^k \uparrow_{2M})^T.$$

Finally, applying the product rule, the Jacobian of each pair of polyphase components with respect to the lattice parameters is

$$\begin{aligned} &\begin{pmatrix} \partial G_i(z)/\partial \beta_{i,j} \\ \partial G_{i+M}(z)/\partial \beta_{i,j} \end{pmatrix} = \\ &\alpha_i \begin{pmatrix} \beta_{i,m-1} & z^{-1} \\ 1 & -\beta_{i,m-1} z^{-1} \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ 0 & -z^{-1} \end{pmatrix} \cdots \begin{pmatrix} \beta_{i,1} & z^{-1} \\ 1 & -\beta_{i,1} z^{-1} \end{pmatrix} \begin{pmatrix} \beta_{i,0} \\ 1 \end{pmatrix} \\ &+ \left(-\beta_{i,j} (1 + \beta_{i,j}^2)^{-3/2} \prod_{k \neq j} 1/\sqrt{1 + \beta_{i,k}^2} \right) \begin{pmatrix} G_i(z) \\ G_{i+M}(z) \end{pmatrix} \end{aligned}$$

The dependence of the polyphase components on the lattice parameters is nonlinear, due to the gain and the delays in the formula.

Putting each of these steps with the chain rule and using the fact each set of lattice parameters β_i determines the polyphase components g_i and g_{i+M} as well as g_{2M-1-i} and g_{M-1-i} for linear phase filters, we have

$$\nabla_{\beta_{i,j}} \Phi_1 = \left(\sum_{l \in S} \nabla_{\beta_{l,j}} g_l \nabla_{g_l} p_0 \right) \nabla_{p_0} r \nabla_r \Phi_1,$$

with $S = \{i, i+M, 2M-1-i, M-1-i\}$. The objective is not convex in terms of the lattice parameters, but can be locally optimized with warm starts.

2.2 Approach 2: Quasi-Linearization

Another way to address the non-convex perfect reconstruction constraint is quasi-linearization. With this approach, we approximate the quadratic equality constraint as an affine equality constraint by fixing the left side of the quadratic using the previous iterates:

$$\frac{1}{2}\left(g_i^T(U^k + L^k)g_i^+ + g_{i+M}^T(U^k + L^k)g_{i+M}^+\right) = \delta[k]$$

for $i = 0, \ldots, M-1$, and $k = 0, \ldots, m-1$, where g_i and g_{i+M} are the previous iterates and g_i^+ and g_{i+M}^+ are the variables. This approximation makes the problem a constrained sequential convex program that we can optimize.

2.3 Optimization

Since the problem is not convex in either form, we do not guarantee that the algorithm will converge to a global minimum or even a local minimum, but just a critical point.

For the paraunitary reparameterization, we used an implementation of the BFGS quasi-Newton algorithm to minimize the stopband energy of the prototype filter with respect to the lattice parameters. To improve the critical points found for this non-convex optimization, we used the solution to a simpler problem with smaller m as a warm start for problems with $m > 2$. We also implemented an unconstrained gradient descent with inverse square root and approximate Polyak step sizes.

For the quasi-linearization approach, we used a primal-dual subgradient method to simultaneously minimize an augmented Lagrangian with respect to the filter coefficients and maximize it with respect to the dual variable. This method approaches a KKT dual-optimal point, but its iterates are infeasible.

3 Results

We produced optimized filters for practical lengths and number of channels, $m \lesssim 8$ and $M \lesssim 2048$. The algorithms converged to critical points of the objective and yielded smooth filters that satisfied the perfect reconstruction constraints. Figures 1, 2, and 3 show some optimized, smooth filters and their lowpass frequency responses. These figures illustrate that greater extension factors m, which correspond to longer filters, produce improved frequency responses, offering better coding gains in the context of audio coding.

Both the reparameterization and the quasi-linearization approached the same solutions, but with different precision. While the reparameterization converged to machine precision, the quasi-linearization did not meet the KKT conditions. Specifically, at the end of optimization, the gradient of the objective of the reparameterized problem had norm around 10^{-8}, and the solution necessarily satisfied the constraints exactly. On the other hand, the quasi-linearized solution had a stationarity vector with norm around 10^{-1} and a constraint residual

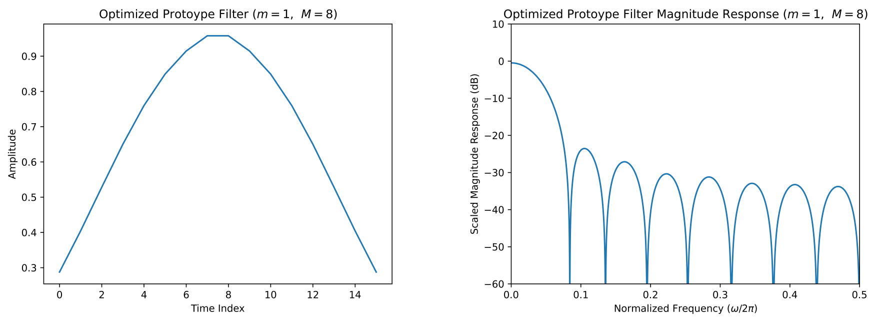

Figure 1: Optimized prototype filter for $m = 1$, with $\omega_s = \frac{\pi}{M}$

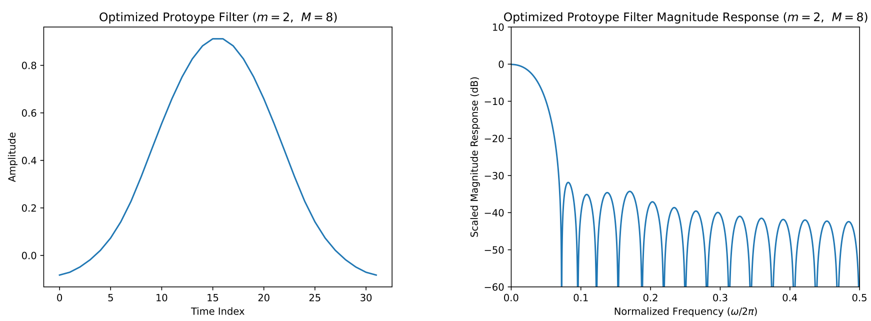

Figure 2: Optimized prototype filter for $m = 2$, with $\omega_s = \frac{\pi}{M}$

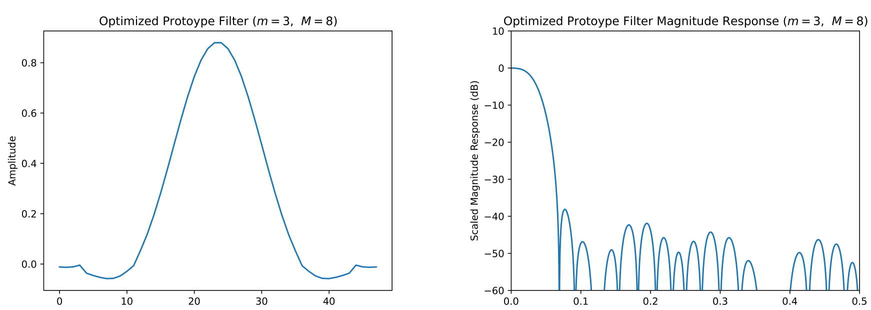

Figure 3: Optimized prototype filter for $m = 3$, with $\omega_s = \frac{\pi}{M}$

with norm around 10^{-3}. Comparing the two approaches, the paraunitary reparameterization is preferable due to its convergence to stationarity and perfect satisfaction of the constraints.

Tractability was an important issue for both approaches, which both required computationally efficient implementations that made use of the sparsity patterns in the subgradients and constraints. For the unconstrained problem, the BFGS quasi-Newton method converged with many fewer iterations than the gradient descent. For the gradient descent, an approximate Polyak step size converged more quickly but with less accuracy than the inverse square root step size.

4 Conclusion and Future Work

We implemented two routines to produce practical prototype filters for TDAC transforms for audio compression. While the quasi-linearization approach is simpler, the paraunitary reparameterization is more effective. The paraunitary implementation is publicly available online: `https://github.com/rwixen/TDAC_optimization/tree/main`

As a next step, we could implement a minimax optimization for stopband energy, using the minimum stopband energy filters as warm starts [6]. We could also optimize filters for time-varying TDAC, which have different constraints.

References

[1] Marina Bosi. Filter Banks in Perceptual Audio Coding. In *Audio Engineering Society Conference: 17th International Conference: High-Quality Audio Coding*, September 1999.

[2] Marina Bosi and Richard E. Goldberg. *Introduction to digital audio coding and standards.* Number SECS 721 in The Kluwer international series in engineering and computer science. Kluwer Academic Publishers, Boston, 2003.

[3] J. Princen and A. Bradley. Analysis/Synthesis filter bank design based on time domain aliasing cancellation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1153–1161, October 1986.

[4] J. Princen, A. Johnson, and A. Bradley. Subband/Transform coding using filter bank designs based on time domain aliasing cancellation. In *ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 12, pages 2161–2164, April 1987. Journal Abbreviation: ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing.

[5] Henrique S. Malvar. *Signal processing with lapped transforms.* The Artech House communications library. Artech House, Boston, 1992.

[6] R. D. Koilpillai and P. P. Vaidyanathan. Cosine-modulated FIR filter banks satisfying perfect reconstruction. *IEEE Transactions on Signal Processing*, 40(4):770–783, April 1992.

[7] J.O. Smith. Paraunitary Filters. `https://ccrma.stanford.edu/~jos/filters/Paraunitary_FiltersC_3.html`, 2007.

[8] P. P. Vaidyanathan. *Multirate Systems and Filter Banks.* Prentice-Hall signal processing series. Prentice Hall, Englewood Cliffs, N.J, 1993.