# ADAPTIVE ITERATIVE REWEIGHTED LEAST SQUARES DESIGN OF $L_P$ FIR FILTERS

*Ricardo A. Vargas and Charles S. Burrus*

Electrical and Computer Engineering Department
Rice University
Houston, TX 77005
rickv@rice.edu, csb@ece.rice.edu

**ABSTRACT**

This paper presents an efficient adaptive algorithm for designing FIR digital filters that are efficient according to an $L_p$ error criteria. The algorithm is an extension of Burrus' iterative reweighted least-squares (IRLS) method for approximating $L_p$ filters. Such algorithm will converge for most significant cases in a few iterations. In some cases however, the transition bandwidth is such that the number of iterations increases significantly. The proposed algorithm controls such problem and drastically reduces the number of iterations required.

## 1. INTRODUCTION

In designing FIR filters, it is usually necessary to minimize an error norm. Typically, the two most commonly used norms are the Chebishev norm ($L_\infty$) and the least-squares norm ($L_2$). However, in some applications minimizing either the error energy ($L_2$) or the maximum error ($L_\infty$) is not the optimal approach to designing a filter. This article addresses the design of filters that are optimal in the $L_p$ norm (i.e. filters designed through the minimization of the $p$-th power of the error).

There is no analytical way to minimize the $p$-th power of the error; therefore an iterative approach must be used. The application of Iterative Reweighted Least Squares (IRLS) methods has been studied intensively by applied mathematicians. In 1961 Lawson [8] came up first with an IRLS algorithm to solve the Chebishev approximation problem. He proved the existence of an optimal solution and the linear convergence associated with his method. Rice and Usow [11] extended Lawson's method to a generalization of the $L_p$ problem, pointing ut that Lawson's method could occationally be required to restart. Karlovitz [7] presented in 1970 an IRLS algorithm with guaranteed linear convergence for even values of $p$. Kahng [6] came up with an extension of Lawson's algorithm to $L_p$ problems, based on Newton-Raphson's method, and proved that his algorithm always converges. Independently, Fletcher et al. [5] developed a similar algorithm. Burrus et al. [2, 4, 3] developed a robust algorithm that would converge quadratically under most conditions; however it is sensitive to certain cases where the transition bandwidth causes the algorithm to produce occasional jumps in the error. For such cases a large number of iterations is required. The method presented in this article combines the robustness of Burrus algorithm with some degree of flexibility in the use of adaptive parameters that would allow for the fast convergence of a Newton-based method and strong robustness. The basic theory of IRLS methods applied to the design of $L_p$ FIR filters is presented, and several approaches are reviewed. The need of an adaptive method that overcomes particular cases is justified and elaborated.

## 2. FIR FILTER DESIGN

Typically, FIR filters are designed by considering a sampled version of a desired frequency response $A_d(\omega)$. For linear phase filters, it is possible to express the frequency response

$$H(\omega) = \sum_{n=0}^{N-1} h(n) \exp^{-j\omega n} \tag{1}$$

(where $h(n)$ is the length-$N$ filter impulse response) in terms of a real amplitude function and a phase term. Using symmetry properties of the Fourier transform, a linear-phase FIR filter is defined by $H(\omega) = A(\omega) \exp^{-jM\omega}$, if $h(n)$ satisfies some symmetry constraints [10]. For the purposes of this paper, only odd-length filters with even symmetry will be considered. The real-valued amplitude function $A(\omega)$ is defined as follows [4]

$$A(\omega) = \sum_{n=0}^{M} a(n) \cos \omega(M-n) \tag{2}$$

where

$$a(n) = \begin{cases} 2h(n) & 0 \leq n \leq M-1 \\ h(M) & n = M \\ 0 & \text{otherwise} \end{cases}$$

and $M = (N-1)/2$. Equation (2) describes a linear system of equations whose solution is a set of coefficients that characterize the filter impulse response. In matrix form, this is expressed as $A = Ca$, where $A$ is a column vector with $L$ samples of the desired frequency response, $a$ is the vector of $N$ filter coefficients and $C$ is a cosine matrix required for the Fourier transform.

It is often desirable to take a large number of samples to design a small filter (in the sense that $L \gg N$, where $L$ is the number of frequency samples and $N$ is the filter order). This setting will result in an overdetermined system of equations without an exact solution and, therefore, the problem of designing an FIR filter becomes one of approximating a desired frequency response based upon a particular error norm. The weighted least-squares $(L_2)$ norm, which considers the error energy, is defined by

$$\| E(\omega) \|_2 = \left( \frac{1}{\pi} \int_0^\pi W(\omega)(A(\omega) - A_d(\omega))^2 d\omega \right)^{\frac{1}{2}}$$

where $A_d(\omega)$ and $A(\omega)$ are the desired and designed amplitude responses respectively. A more general error criteria minimizes the $p$-th power of the error with

$$\| E(\omega) \|_p = \left( \frac{1}{\pi} \int_0^{\pi} W(\omega)(A(\omega) - A_d(\omega))^p d\omega \right)^{\frac{1}{p}} \qquad (3)$$

The weighted Chebishev ($L_\infty$) criteria, which minimizes the maximum error, is given by

$$\| E(\omega) \|_\infty = \max_{\omega \in [0,\pi]} | W(\omega)(A(\omega) - A_d(\omega)) |$$

In all cases $W(\omega)$ is a nonnegative weighting function.

Considering a discretized version of (3), the objective is to find the coefficients $a(n)$ such that the scalar error

$$\varepsilon_p = \sum_{k=0}^{L-1} | A(\omega_k) - A_d(\omega_k) |^p \qquad (4)$$

is minimized over the frequencies $\omega_k$. Using the $L_2$ norm, the minimization of

$$\varepsilon = \sum_{k=0}^{L-1} | A(\omega_k) - A_d(\omega_k) |^2$$

results in

$$\varepsilon = \epsilon^T \epsilon$$

where $\epsilon$ is the residual vector

$$\epsilon = Ca - A_d$$

The resulting normal equations are given by

$$C^T Ca = C^T A_d$$

A weighted error approach results in

$$\varepsilon = \sum_{k=0}^{L-1} w_k^2 \, | A(\omega_k) - A_d(\omega_k) |^2$$

with

$$\varepsilon = \epsilon^T W^T W \epsilon$$

where $W$ is a diagonal matrix with the weights $w_k$ in its diagonal. It can be proved that the resulting normal equations have the form

$$C^T W^T W C a = C^T W^T W A_d \qquad (5)$$

The objective of IRLS algorithms consists in finding the optimal weights in (5) that minimize the scalar error of (4).

## 3. IRLS METHODS

The basic IRLS approach to solve (3) consists in finding iteratively the optimal weights $W$ for (5) using

$$a_m = [C^T W_m^T W_m C]^{-1} C^T W_m^T W_m A_d \qquad (6)$$

at the $m$-th iteration. The first guess for $a_m$ considers unit weights in the diagonal of $W$. Then, the resulting error is found using

$$\epsilon_m = Ca_m - A_d$$

Since we can express (4) as

$$\varepsilon = \sum_{k=0}^{L-1} w_m \, | A(\omega_k) - A_d(\omega_k) |^2$$

we can define a weighting vector by

$$w_{m+1} = |\epsilon_m|^{(p-2)/2}$$

For each iteration, these values are located in the diagonal of $W$. Then, (6) is used again until the method converges to the proper solution. However, it has been found that this approach has intense practical problems, since the inversion required by (6) often uses an ill-posed matrix and, in most cases, convergence is not achieved.

Rice and Usow [11] developed an algorithm based on Lawson's method that requires a multiplicative update of the weights after each iteration. They used results from Motzkin and Walsh [9] that warranteed the existence of a solution for the Chebishev approximation problem to support the use of a weighted least-squares algorithm for the $L_p$ problem. They defined

$$w_{m+1} = w_m^\alpha \, |\epsilon_m|^\beta$$

where

$$\alpha = \frac{\gamma(p-2)}{\gamma(p-2)+1}$$

and

$$\beta = \frac{\alpha}{2\gamma} = \frac{p-2}{2(\gamma(p-2)+1)}$$

The rest of the algorithm works the same way as the basic IRLS method. However, the proper selection of $\gamma$ will allow for a strong convergence algorithm. Note that for $\gamma = 0$ we obtain the basic IRLS algorithm.

Another approach to solve (5) consists in using a temporary coefficient vector defined by

$$\hat{a}_{m+1} = [C^T W_{m+1}^T W_{m+1} C]^{-1} C^T W_{m+1}^T W_{m+1} A_d \qquad (7)$$

The filter coefficients after each iteration are then updated by

$$a_{m+1} = \lambda \hat{a}_{m+1} + (1-\lambda) a_m$$

This approach is known as the Karlovitz method [7], and it has being claimed that it converges to the global optimal solution for *even values* of $4 \leq p < \infty$. However, in practice several convergence problems have been found even under such assumptions. One drawback is that the convergence parameter $\lambda$ has to be optimized for each iteration, which requires the multiple evaluation for different values of $0 \leq \lambda \leq 1$, and a linear search for the best value must be done. Therefore the overall execution time becomes rather large.

Kahng [6] developed an algorithm based on Newton-Raphson's method that uses

$$\lambda = \frac{1}{p-1} \qquad (8)$$

to get

$$a_m = \frac{\hat{a}_m + (p-2)a_{m-1}}{p-1} \qquad (9)$$

This selection for $\lambda$ is based upon Newton's method to minimize $\epsilon$. The rest of the algorithm follows Karlovitz approach. However, since $\lambda$ is fixed, there is no need to perform the linear search for its best value. Since Kahng's method is based upon Newton's method, it converges quadratically to the optimal solution. Kahng proved that his method converges for all cases of $\lambda$ and for any problem. It can be seen that Kahng's method is a particular case of Karlovitz algorithm, with $\lambda$ as defined in (8). Newton-raphson based algorithms are not warranteed to converge to the optimal solution unless they are somewhat close to the solution since they require to know and invert the Hessian matrix of the minimized function, which must be positive definite [1]. However, their quadratic convergence makes them an appealing option.

Burrus, Barreto and Selesnick [4, 2, 3] developed a method (which from now on will be referred to as BBS) that combines the powerful quadratic convergence of Newton methods with the robust initial convergence of the basic IRLS method, thus overcoming the initial sensitivity of Newton-based algorithms and the slow linear convergence of Lawson-based methods. To accelerate initial convergence, their method for minimizing the $p$-th power of the error uses initially $p = K * 2$, where $K$ is a convergence parameter defined by $1 \leq K \leq 2$. At the next iteration, $p$ increases its value by a factor of $K$, to $p = K^2 * p$. This is done at each iteration, so to satisfy

$$p_m = \min(p_{des}, Kp_{m-1}) \tag{10}$$

The implementation of each iteration follows Karlovitz's method using the particular selection of $p$ given by (10).

## 4. ADAPTIVE ALGORITHM

Much of the performance of a method is based upon whether it can reach a global minima for a certain error measure. In the case of the methods described above, both convergence rate and stability play an important role in the method's performance. Both Karlovitz and RUL methods are supposed to converge linearly, while Kahng's and Burrus' methods converge quadratically, since they both use a Newton-based additive update of the weights.

Barreto showed in [2] that the modified version of Kahng's method (or BBS) typically converges faster than the RUL algorithm. However, this approach presents some particular problems that are dependent upon the transition bandwidth $\beta$. For some particular values of $\beta$, the BBS method will result in an ill-posed weight matrix that causes the $L_p$ error to increase dramatically after a few iterations. Two facts can be derived from Figure 1: for this particular bandwidth the error increased slightly after the fifth and eleventh iterations, and increased dramatically after the sisteenth. Also, it is worth to notice that after such increase, the error started to decrease quadratically and that, at a certain point, it flattened (thus reaching the computational limits of the computer).

The effects of different values of $K$ were studied to find out if a relationship between $K$ and the error increase could be determined. Figure 2 shows the $L_p$ error for different values of $\beta$ and for $K = 1.7$. It can be seen that some particular bandwidths cause the algorithm to produce a very large error.

The conclusions derived above suggest the possibility to use an adaptive algorithm that changes the value of $K$ so that the error always decreases. This idea was implemented by calculating temporary new weight and filter coefficient vectors that will not become the updated versions unless their resulting error is smaller than the previous one. If this is not the case, the algorithm "tries" two values of $K$, namely

$$K_L = K * (1 - \Delta) \text{ and } K_H = K * (1 + \Delta) \tag{11}$$

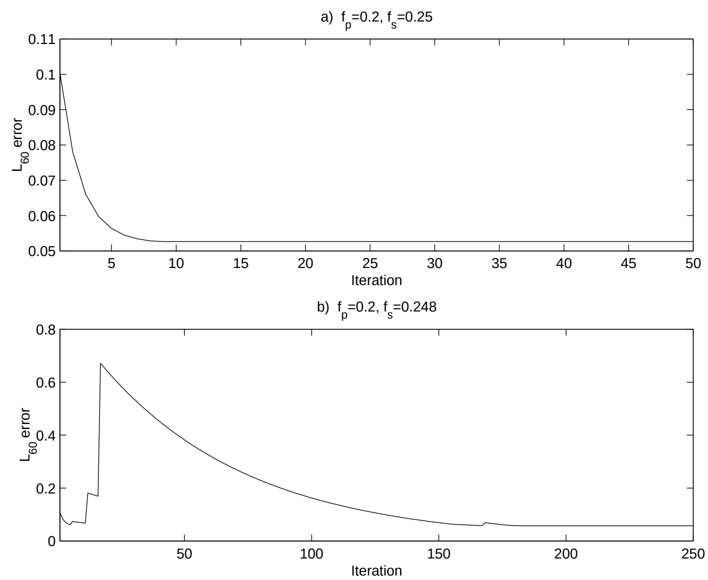


Figure 1: a) BBS method, normal behavior; b) a problematic bandwith.

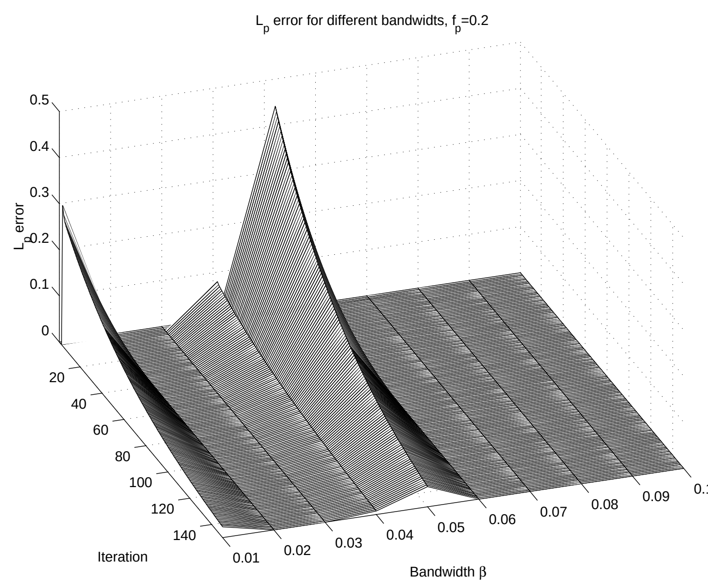


Figure 2: BBS results for different bandwidths.

(where $\Delta$ is an updating variable). The resulting errors for each attempt are calculated, and $K$ is updated according to the value that produced the smallest error. The error of this new $K$ is compared to the error of the nonupdated weights and coefficients, and if the new $K$ produces a smaller error, then such vectors are updated; otherwise another update of $K$ is performed. The algorithm can be summarized as follows,

1. Find the unweighted approximation $a = C^{-1}A_d$ and use $p_0 = K \cdot 2$, where $1 \leq K \leq 2$.
2. Iteratively solve (7) and (9) using

$$\lambda_k = \frac{1}{p_k - 1}$$

   and find the resulting error $\varepsilon_k$ for the $k$-th iteration.
3. If $\varepsilon_k \gg \varepsilon_{k-1}$,
   - Calculate (11).
   - Select the smallest of $\varepsilon_{K_L}$ and $\varepsilon_{K_H}$ to compare it with $\varepsilon_k$ until a value is found that results in a decreasing error.

The algorithm described above changes the value of $K$ that causes the algorithm to produce a large error. $K$ is updated as many times as necessary without changing the values of the weights, the filter coefficients or the variable error power $p$. If an optimal value of $K$ exists, the algorithm will find it and continue with this new value until another update in $K$ becomes necessary. This algorithm was implemented for several combinations of $K$ and $\beta$; for all cases the new algorithm converged faster than the BBS algorithm (obviously unless the values of $K$ and $\beta$ are such that the error never increases; in this case the new algorithm works exactly as such method). The results are shown in Figure 3.a for the specifications from Figure 1. Whereas using the BBS method for this particular case results in a large error after the sixteenth iteration, the adaptive method converged before ten iterations.

Figure 3.b illustrates the change of $K$ per iteration in the adaptive method, using an update factor of $\Delta = 0.1$. The $L_p$ error stops decreasing after the fifth iteration (where the BBS method introduces the large error); however, the adaptive algorithm adjusts the value of $K$ so that the $L_p$ error continues decreasing. The algorithm decreased the initial value of $K$ from 1.75 to its final value of 1.4175 (at the expense of only one additional iteration with $K = 1.575$).

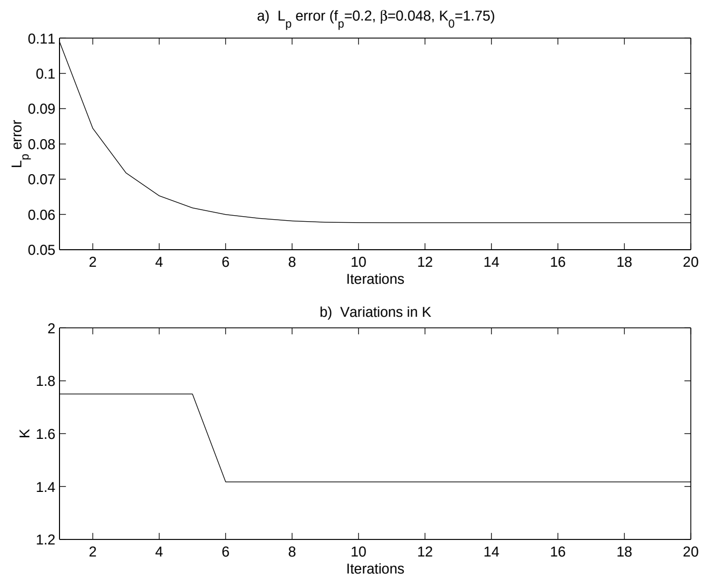


Figure 3: a) $L_p$ error obtained with the adaptive method; b) Change of K.

## 5. CONCLUSIONS

A description of the $L_p$ FIR filter design problem was presented. The use of iterative reweighted least squares methods to solve this problem was illustrated and an overview of different approaches was covered. Both RUL and Kahng's methods improve the basic IRLS performance with certain acceleration techniques; however these methods are not robust enough since they are either sensitive to the starting conditions (Kahng) or they have slow linear convergence (RUL).

The BBS method is a fast-convergent algorithm. However, its sensitiveness to the transition bandwidth requires a more robust method. An adaptive algorithm was introduced as a solution to this problem. The new method exhibits the same convergence properties as the BBS algorithm for most filters, but is able to adjust its parameters to avoid the occasional dramatic error increases, thus making it a robust alternative for designing $L_p$ linear phase filters.

## 6. REFERENCES

[1] Masanao Aoki. *Introduction to Optimization Techniques*. The Macmillan Company, 1971.

[2] Jose A. Barreto. $l_p$ approximation by the iterative reweighted least squares method and the design of digital FIR filters in one dimension. Master's thesis, Rice University, 1992.

[3] C. S. Burrus and J. A. Barreto. Least $p-$power error design of fir filters. In *Proc. IEEE Int. Symp. Circuits, Syst. ISCAS-92*, pages 545–548, San Diego, CA, May 1992.

[4] C. S. Burrus, J. A. Barreto, and I. W. Selesnick. Iterative reweighted design of FIR filters. *IEEE Transactions on Signal Processing*, 42(11):2926–2936, November 1994.

[5] R. Fletcher, J. A. Grant, and M. D. Hebden. The calculation of linear best $l_p$ approximations. *The Computer Journal*, 14(118):276–279, Apr 1972.

[6] S. W. Kahng. Best $l_p$ approximations. *Mathematics of Computation*, 26(118):505–508, April 1972.

[7] L. A. Karlovitz. Construction of nearest points in the $l^p$, $p$ even and $l^\infty$ norms, i. *Journal of Approximation Theory*, 3:123–127, 1970.

[8] C. L. Lawson. *Contributions to the Theory of Linear Least Maximum Approximations*. PhD thesis, UCLA, 1961.

[9] T. S. Motzkin and J. L. Walsh. Polynomials of best approximation on an interval. *Nat. Acad. Sci. U.S.A.*, 45:1523–1528, 1959.

[10] T. W. Parks and C. S. Burrus. *Digital Filter Design*. John Wiley and Sons, 1987.

[11] John R. Rice and Karl H. Usow. The Lawson algorithm and extensions. *Mathematics of Computation*, 22:118–127, 1968.