Iterative Reweighted Least Squares*

C. Sidney Burrus

Abstract

Describes a powerful optimization algorithm which iteratively solves a weighted least squares approximation problem in order to solve an L_p approximation problem.

1 Approximation

Methods of approximating one function by another or of approximating measured data by the output of a mathematical or computer model are extraordinarily useful and ubiquitous. In this note, we present a very powerful algorithm most often called "Iterative Reweighted Least Squares" or (IRLS). Because minimizing the weighted squared error in an approximation can often be done analytically (or with a finite number of numerical calculations), it is the base of many iterative approaches. To illustrate this algorithm, we will pose the problem as finding the optimal approximate solution of a set of simultaneous linear equations

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1N} \\ a_{21} & a_{22} & a_{23} & & \\ a_{31} & a_{32} & a_{33} & & \\ \vdots & & & & \vdots \\ a_{M1} & & & \cdots & a_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_M \end{bmatrix} \tag{1}$$

or, in matrix notation

$$\mathbf{Ax} = \mathbf{b} \tag{2}$$

where we are given an M by N real matrix $\mathbf{A}$ and an M by 1 vector $\mathbf{b}$, and want to find the N by 1 vector $\mathbf{x}$. Only if $\mathbf{A}$ is non-singular (square and full rank) is there a unique, exact solution. Otherwise, an approximate solution is sought according to some criterion of approximation.

If $\mathbf{b}$ does not lie in the range space of $\mathbf{A}$ (the space spanned by the columns of $\mathbf{A}$, [8]), there is no exact solution to (2), therefore, an approximation problem is posed to be solved by minimizing the norm (or some other measure) of an equation error vector defined by

$$\mathbf{e} = \mathbf{Ax} - \mathbf{b}. \tag{3}$$

A more general discussion of the solution of simultaneous equations can be found the the Connexions module [8].

*Version 1.12: Dec 17, 2012 2:09 pm -0600

†http://creativecommons.org/licenses/by/3.0/

2 Least Squared Error Approximation

A generalized solution (an optimal approximate solution) to (2) is usually considered to be an $\mathbf{x}$ that minimizes some norm or other measure of $\mathbf{e}$. If that problem does not have a unique solution, further conditions, such as also minimizing the norm of $\mathbf{x}$, are imposed and this combined problem always has a unique solution.

The l_2 or root-mean-squared error or Euclidean norm is $\sqrt{\mathbf{e^T e}}$ and its minimization has an analytical solution. This squared error is defined as

$$||\mathbf{e}||_2^2 = \sum_i e_i^2 = \mathbf{e^T e} \tag{4}$$

which when minimized results in an exact or approximation solution of (2) if $\mathbf{A}$ has full row or column rank. The three cases are:

- If $\mathbf{A}$ has $M = N$, (square and nonsingular), then the exact solution is

$$\mathbf{x} = \mathbf{A^{-1} b} \tag{5}$$

- If $\mathbf{A}$ has $M > N$, (over specified) then the approximate solution with the least squared equation error is

$$\mathbf{x} = \left[\mathbf{A^T A}\right]^{\mathbf{-1}} \mathbf{A^T b} \tag{6}$$

- If $\mathbf{A}$ has $M < N$, (under specified) then the approximate solution with the least norm is

$$\mathbf{x} = \mathbf{A^T} \left[\mathbf{A A^T}\right]^{\mathbf{-1}} \mathbf{b} \tag{7}$$

These formulas assume $\mathbf{A}$ has full row or column rank but, if not, generalized solutions exist using the Moore-Penrose pseudoinverse [1], [6], [8].

3 Weighted Least Squared Error Approximation

In addition to these cases with "analytical" solutions, we can pose a more general problem by asking for an optimal approximation with a weighted norm [6], [7] to emphasize or de-emphasize certain components or range of equations. Here we minimize

$$||\mathbf{We}||_2^2 = \sum_i w_i^2 e_i^2 = \mathbf{e^T W^T W e} \tag{8}$$

where $\mathbf{W}$ is a diagonal matrix with the weights, w_i, along its diagonal. The approximate solutions [8] are

- If $\mathbf{A}$ has $M > N$, (over specified) then the minimum weighted equation error solution is

$$\mathbf{x} = \left[\mathbf{A^T W^T W A}\right]^{\mathbf{-1}} \mathbf{A^T W^T W b} \tag{9}$$

- If $\mathbf{A}$ has $M < N$, (under specified) then the minimum weighted norm solution is

$$\mathbf{x} = \left[\mathbf{W^T W}\right]^{\mathbf{-1}} \mathbf{A^T} \left[\mathbf{A} \left[\mathbf{W^T W}\right]^{\mathbf{-1}} \mathbf{A^T}\right]^{\mathbf{-1}} \mathbf{b} \tag{10}$$

These solutions to the weighted approximation problem are useful in their own right but also serve as the foundation for the Iterative Reweighted Least Squares (IRLS) algorithm developed next.

4 Approximation with Other Norms and Error Measures

Most of the discussion about the approximate solutions to $\mathbf{Ax} = \mathbf{b}$ are about the result of minimizing the l_2 equation error $||Ax - b||_2$ and/or the l_2 norm of the solution $||\mathbf{x}||_2$ because in some cases that can be done by analytic formulas and also because the l_2 norm has a energy interpretation. However, both the l_1 and the l_∞ norms [18] have well known applications that are important [21], [15] and use of the more general l_p error is remarkably flexible [10], [12]. Donoho has shown [23] that l_1 optimization gives essentially the same sparsity as use of the true sparsity measure in l_0.

The definition of the l_p norm of the vector $\mathbf{x}$ for p greater than one is:

$$||\mathbf{e}||_p = \left(\sum_i |e_i|^p\right)^{1/p} \tag{11}$$

which has the same solution for the minimum as

$$||\mathbf{e}||_\mathbf{p}^\mathbf{p} = \sum_\mathbf{i} |\mathbf{e}_\mathbf{i}|^\mathbf{P} \tag{12}$$

but is often easier to use.

For the case where the rank is less than M or N, one can use one norm for the minimization of the equation error norm and another for minimization of the solution norm. And in other cases, one can simultaneously minimize a weighted error to emphasize some equations relative to others [6]. A modification allows minimizing according to one norm for one set of equations and another norm for a different set. A more general error measure than a norm can be used which uses a polynomial error [12] that does not satisfy the scaling requirement of a norm, but is convex. One could even use the so-called l_p norm for $1 > p > 0$ which is not even convex but is an interesting tool for obtaining sparse solutions or discounting outliers. Still more unusual is the use of negative p.

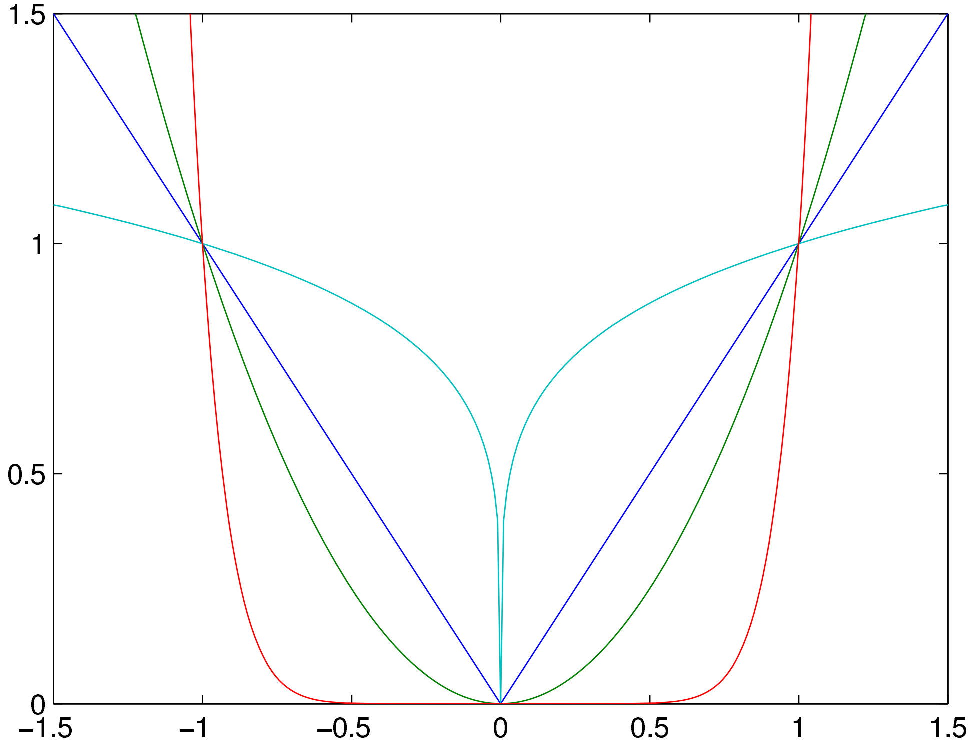

Figure 12: Different l_p norms: p = .2, 1, 2, 10.

Note from the figure how the l_{10} norm puts a large penalty on large errors. This gives a Chebyshev-like solution. The $l_{0.2}$ norm puts a large penalty on small errors making them tend to zero. This (and the l_1 norm) tends to discount "outliers" and give a sparse solution.

5 The l_p Norm Approximation

The **IRLS** (iterative reweighted least squares) algorithm allows an iterative algorithm to be built from the analytical solutions of the weighted least squares with an iterative reweighting to converge to the optimal l_p approximation [7], [37].

5.1 The Overdetermined System with more Equations than Unknowns

If one poses the l_p approximation problem in solving an overdetermined set of equations (case 2 from Chapter 3), it comes from defining the equation error vector

$$\mathbf{e} = \mathbf{A}\mathbf{x} - \mathbf{b} \tag{13}$$

and minimizing the p-norm defined in (11) or, equivalently, (12), neither of which can we minimize directly. However, we do have formulas [6] to find the minimum of the weighted squared error

$$||\mathbf{W}\mathbf{e}||_2^2 = \sum_n \mathbf{w}_n^2 |\mathbf{e}_n|^2 \tag{14}$$

one of which is

$$\mathbf{x} = \left[\mathbf{A^T W^T W A}\right]^{-1} \mathbf{A^T W^T W b} \tag{15}$$

where $\mathbf{W}$ is a diagonal matrix of the error weights, w_n.

5.2 Iteratively Reweighted Least Squares (IRLS)

If one poses the l_p approximation problem in solving an overdetermined set of equations (case 2 [8]), it comes from defining the equation error norm as

$$||e||_p = \left(\sum_n |e(n)|^p \right)^{1/p} \tag{16}$$

and finding $\mathbf{x}$ to minimizing this p-norm of the equation error.

It has been shown this is equivalent to solving a least weighted squared error problem for the appropriate weights.

$$||e||_p = \left(\sum_n w(n)^2 |e(n)|^2 \right)^{1/2} \tag{17}$$

If one takes (16) and factors the term being summed in the following manor, comparison to (17) suggests the iterative reweighted least algorithm which is the subject of these notes.

$$||e||_p = \left(\sum_n |e(n)|^{(p-2)} |e(n)|^2 \right)^{1/p} \tag{18}$$

To find the minimum l_p approximate solution, we propose the iterative reweighted least squared (IRLS) error algorithm which starts with unity weighting, $\mathbf{W} = \mathbf{I}$, solves for an initial $\mathbf{x}$ with (15), calculates a new error from (13), which is then used to set a new weighting matrix $\mathbf{W}$ with diagonal elements of

$$w(n) = e(n)^{(p-2)/2} \tag{19}$$

to be used in the next iteration of (15). Using this, we find a new solution $\mathbf{x}$ and repeat until convergence (if it happens!). To guarantee convergence, this process should be a contraction map which converges to a fixed point that is the solution. A simple Matlab program that implements this algorithm is:

```
        % m-file IRLS0.m to find the optimal solution to Ax=b
%  minimizing the L_p norm ||Ax-b||_p, using basic IRLS.
%  csb 11/10/2012
function x = IRLS0(A,b,p,KK)
if nargin < 4, KK=10;  end;
x  = pinv(A)*b;                            % Initial L_2 solution
E = [];
for k = 1:KK                               % Iterate
   e  = A*x - b;                           % Error vector
   w  = abs(e).^((p-2)/2);                 % Error weights for IRLS
   W  = diag(w/sum(w));                    % Normalize weight matrix
   WA = W*A;                               % apply weights
   x  = (WA'*WA)\(WA'*W)*b;                % weighted L_2 sol.
   ee = norm(e,p);   E = [E ee];           % Error at each iteration
end
plot(E)
```

Listing 1: Program 1. Basic Iterative Reweighted Least Squares Algorithm

This core idea has been repeatedly proposed and developed in different application areas over the past 50 or so years with a variety of successes [7], [37], [10]. Used in this basic form, it reliably converges for $1.5 < p < 3$. In 1970 [31], a modification was made by only partially updating the solution each iteration using

$$\mathbf{x}(k) = q\ \hat{\mathbf{x}}(k) + (1-q)\ \mathbf{x}(k-1) \tag{20}$$

where $\hat{\mathbf{x}}$ is the new weighted least squares solution of (15) which is used to only partially update the previous value $\mathbf{x}(k-1)$ and k is the iteration index. The first use of this partial update optimized the value for q on each iteration to give a more robust convergence but it slowed the total algorithm considerably.

A second improvement was made by using a specific up-date factor of

$$q = \frac{1}{p-1} \tag{21}$$

to significantly improve the rate of asymptotic convergence. With this particular factor and for p>2, the algorithm becomes a form of Newton's method which has quadratic asymptotic convergence [30], [12] but, unfortunately, the initial convergence became less robust.

A third modification applied homotopy [11], [49], [48], [45], [30], [24] by starting with a value for p equal to 2 and increasing it each iteration (or the first few iterations) until it reached the desired value, or, in the case of $p < 2$, decreasing it. This improved the initial performance of the algorithm and made a significant increase in both the range of p that allowed convergence and in the speed of calculations. Some of the history and details can be found applied to digital filter design in [10], [12].

A Matlab program that implements these ideas applied to our pseudoinverse problem with more equations than unknowns is:

```
        % m-file IRLS1.m to find the optimal solution to Ax=b
%  minimizing the L_p norm ||Ax-b||_p, using IRLS.
%  Newton iterative update of solution, x, for  M > N.
%  For 2<p<infty, use homotopy parameter K = 1.01 to 2
%  For 0<p<2, use K = approx 0.7 - 0.9
%  csb 10/20/2012
function x = IRLS1(A,b,p,K,KK)
if nargin < 5, KK=10;  end;
if nargin < 4, K = 1.5;  end;
if nargin < 3, p = 10; end;
pk = 2;                                       % Initial homotopy value
x  = pinv(A)*b;                               % Initial L_2 solution
E = [];
for k = 1:KK                                  % Iterate
   if p >= 2, pk = min([p, K*pk]);             % Homotopy change of p
      else pk = max([p, K*pk]); end
   e  = A*x - b;                              % Error vector
   w  = abs(e).^((pk-2)/2);                   % Error weights for IRLS
   W  = diag(w/sum(w));                       % Normalize weight matrix
   WA = W*A;                                  % apply weights
   x1  = (WA'*WA)\(WA'*W)*b;                  % weighted L_2 sol.
   q  = 1/(pk-1);                             % Newton's parameter
   if p > 2, x = q*x1 + (1-q)*x; nn=p;         % partial update for p>2
      else x = x1; nn=2; end                  % no partial update for p<2
   ee = norm(e,nn);   E = [E ee];             % Error at each iteration
end
plot(E)
```

Listing 2: Program 2. IRLS Algorithm with Newton's updating and Homotopy for M>N

This can be modified to use different p's in different bands of equations or to use weighting only when the error exceeds a certain threshold to achieve a constrained LS approximation [10], [12], [54].

Here, it is presented as applied to the overdetermined system (Case 2a and 2b [8]) but can also be applied to other cases. A particularly important application of this section is to the design of digital filters [12].

5.3 The Underdetermined System with more Unknowns than Equations

If one poses the l_p approximation problem in solving an underdetermined set of equations (case 3a [8]), it comes from defining the solution norm as

$$||x||_p = \left(\sum_n |x(n)|^p \right)^{1/p} \tag{22}$$

and finding $\mathbf{x}$ to minimizing this p-norm while satisfying $\mathbf{Ax} = \mathbf{b}$.

It also has been shown this is equivalent to solving a least weighted norm problem for specific weights.

$$||x||_p = \left(\sum_n w(n)^2 |x\,(n)\,|^2 \right)^{1/2} \tag{23}$$

The development follows the same arguments as in the previous section but using the formula [44], [6]

$$\mathbf{x} = \left[\mathbf{W^T W}\right]^{-1} \mathbf{A^T} \left[\mathbf{A}\left[\mathbf{W^T W}\right]^{-1} \mathbf{A^T}\right]^{-1} \mathbf{b} \tag{24}$$

with the weights, $w\,(n)$, being the diagonal of the matrix, $\mathbf{W}$, in the iterative algorithm to give the minimum weighted solution norm in the same way as (15) gives the minimum weighted equation error.

A Matlab program that implements these ideas applied to our pseudoinverse problem with more unknowns than equations is:

```
      % m-file IRLS2.m to find the optimal solution to Ax=b
%  minimizing the L_p norm ||x||_p, using IRLS.
%  Newton iterative update of solution, x, for  M < N.
%  For 2<p<infty, use homotopy parameter K = 1.01 to 2
%  For 0<p<2, use K = approx 0.7 to 0.9
%  csb 10/20/2012
function x = IRLS2(A,b,p,K,KK)
if nargin < 5, KK= 10;  end;
if nargin < 4, K = .8;  end;
if nargin < 3, p = 1.1; end;
pk = 2;                                 % Initial homotopy value
x  = pinv(A)*b;                         % Initial L_2 solution
E = [];
for k = 1:KK
   if p >= 2, pk = min([p, K*pk]);       % Homotopy update of p
      else pk = max([p, K*pk]); end
   W  = diag(abs(x).^((2-pk)/2)+0.00001);  % norm weights for IRLS
   AW = A*W;                            % applying new weights
   x1 = W*AW'*((AW*AW')\b);             % Weighted L_2 solution
   q  = 1/(pk-1);                       % Newton's parameter
   if p >= 2, x = q*x1 + (1-q)*x; nn=p; % Newton's partial update for p>2
      else x = x1; nn=1; end            % no Newton's partial update for p<2
   ee = norm(x,nn);  E = [E ee];        % norm at each iteration
end;
plot(E)
```

Listing 3: Program 3. IRLS Algorithm with Newton's updating and Homotopy for M less than N

This approach is useful in sparse signal processing and for frame representation. Our work was originally done in the context of filter design but others have done similar things in sparsity analysis [28], [21], [59].

The methods developed here [10], [11] are based on the (IRLS) algorithm [43], [42], [16] and they can solve certain FIR filter design problems that neither the Remez exchange algorithm nor analytical L_2 methods can.

6 History

The idea of using an IRLS algorithm to achieve a Chebyshev or L_∞ approximation was first developed by Lawson in 1961 [32] and extended to L_p by Rice and Usow in 1968 [41], [40]. The basic IRLS method presented here for L_p was given by Karlovitz [31] and extended by Chalmers, et. al. [17], Bani and Chalmers [3], and Watson [58]. Independently, Fletcher, Grant and Hebden [27] developed a similar form of IRLS but based on Newton's method and Kahng [30] did likewise as an extension of Lawson's algorithm. Others analyzed and extended this work [24], [36], [16], [58]. Special analysis has been made for $1 \le p < 2$ by [53], [57], [42], [33], [36], [43], [60] and for $p = \infty$ by [27], [3], [42], [35], [2], [34]. Relations to the Remez exchange algorithm [18], [39] were suggested by [3], to homotopy [45], and to Karmarkar's linear programming algorithm [50] by [42], [47]. Applications of IRLS algorithms to complex Chebyshev approximation in FIR filter design have been made in [25], [20], [22], [51], [4] and to 2-D filter design in [19], [5]. Application to array design can be found in [52] and to statistics in [16].

The papers [12], [54] unify and extend the IRLS techniques and applies them to the design of FIR digital filters. They develops a framework that relates all of the above referenced work and shows them to be variations of a basic IRLS method modified so as to control convergence. In particular, [12] generalizes the work of Rice and Usow on Lawson's algorithm and explain why its asymptotic convergence is slow.

The main contribution of these recent papers is a new robust IRLS method for filter design [10], [11], [54] that combines an improved convergence acceleration scheme and a Newton based method. This gives a very efficient and versatile filter design algorithm that performs significantly better than the Rice-Usow-Lawson algorithm or any of the other IRLS schemes. Both the initial and asymptotic convergence behavior of the new algorithm is examined [54] and the reason for occasional slow convergence of this and all other IRLS methods is presented.

It is then shown that the new IRLS method allows the use of p as a function of frequency to achieve different error criteria in the pass and stopbands of a filter. Therefore, this algorithm can be applied to solve the constrained L_p approximation problem. Initial results of applications to the complex and two-dimensional filter design problem are also presented.

When used for the range $1 < p < 2$, the new IRLS algorithm works well in outlier suppression in data and in finding sparsity for data compression.

Although the traditional IRLS methods were sometimes slower than competing approaches, the results of this paper and the availability of fast modern desktop computers make them practical now and allow exploitation of their greater flexibility and generality.

7 Convergence

Both theory and experience indicate there are different convergence problems connected with several different ranges and values of p. In the range $1.5 \le p < 3$, virtually all methods converge [26], [16]. In the range $3 \le p < \infty$, the basic algorithm diverges and the various methods discussed in this paper must be used. As p becomes large compared to 2, the weights carry a larger contribution to the total minimization than the underlying least squared error minimization, the improvement at each iteration becomes smaller, and the likelihood of divergence becomes larger. For $p = \infty$ the optimal approximation solution to (2) is unique but the weights in (6) that give that solution are not. In other words, different matrices W give the same solution to (2) but will have different convergence properties. This allows certain alteration to the weights to improve convergence without harming the optimality of the results [34].

In the range $1 < p < 2$, both convergence and numerical problems exist as, in contrast to $p > 2$, the IRLS iterations are undoing what the underlying least squares is doing. In particular, the weights near frequencies with small errors become very large. Indeed, if the error happens to be zero, the weight becomes infinite because of the negative exponent in (6). A small term can be added to the error to overcome this problem. For $p = 1$ the solution to the optimization problem is not even unique. The various algorithms that are presented are based on schemes to address these problems.

For some applications, the convergence is not robust. In these cases, the use of an adaptive step size can considerably improve convergence but at the expense of speed [54][55]. For large and small p, the weight functions w_i are not unique even though the solution is unique [54]. This allows flexibility that prevents the occurrence of very large or very small weights [34].

One of the most versatile modifications of the basic IRLS algorithm is the use of different powers p for different equations in (2). For filters, this allows an l_2 approximation in the passband and, simultaneously, an l_∞ approximation in the stopband [12]. Another powerful modification is to use an polynomial error measure such as

$$q = \sum_i |e_i|^2 + K|e_i|^{100} \tag{25}$$

in the IRLS algorithm. The first term dominates for $e_i < 1$ and the second term dominates for $e_i > 1$ which gives a constrained least squares approximation which is a very useful result [12], [14], [13], [54]. There are other modifications which can be made for specific applications where no alternative methods exist such as with 2D approximation [5] and cases where the error measure is not a norm nor even convex. Obviously, there are huge convergence problems but it is a good starting point. Still another application is to the design of recursive or IIR filters. It is possible to formulate these filters as a matrix multiplication and pose an iterative algorithm to find an optimal solution [56], [54], [38], [9].

A particular convergence problem that many iterative algorithms have is to have the error measure decrease for several iterations, then increase. This happens with Newton type algorithms when the function and its derivative are near zero giving a correction term approximately equal to zero over zero, or, indeterminate. This happens when finding a high order zero and that is similar to IRLS and $||\mathbf{e}||_p$ with a large p. One method used by Yagle [59] is to not iterate, but simply make one or two corrections for each of a fixed number of homotopy steps. This needs more investigation.

8 Summary

The individual ideas used in an IRLS program are

- The basic IRLS algorithm of (12),(14),(15),(19) and Program 1. [32]
- Adding a partial update to the iterations as in (20) with optimization of the amount of update [31].
- Choosing an update to give a Newton's method as in (21) and Program 2. [27], [30]
- Applying homotopy by increasing p each iteration as in Program 2. [30], [24], [12]
- Adaptive step size on partial updating and/or homotopy [55], [54]
- Allowing p to vary with different equations allowing a mixture of approximation criteria [4], [12]
- Using p less than one, even zero. [59]
- Using a polynomial error measure as in (25) to give a near constrained least squares approximation. [12]
- Using weighting envelops to stablize convergence for very large p[54]
- Applying IRLS to complex and 2D approximation, to Chebyshev and l_1 approximation, to filter design, outlier supression in regression, and to creating sparse solutions [4], [5].

Iteratively Reweighted Least Squares (IRLS) approximation is a powerful and flexible tool for many engineering and applied problems. By combining several modifications to the basic IRLS algorithm, one can have a fast and robust approximation tool. But, while proofs of convergence can be given for individual parts of the combined algorithm, no formal proof seems to be possible for the combination.

9 Iterative Least Squares

The algorithm for iterative reweighted least squares approximation can be generalized to allow updates in the iterations other than on the weights. This is called iterative least squares and it includes IRLS as a special case where the iteration is on the weights. In all cases, the power and easy implementation of LS is the driving engine of the algorithm. In most, the algorithm is posed as a successive approximation in the form of a contraction map. An example of that is the design of a digital filter using optimal squared magnitude approximation. The ILS algorithm sets a desired magnitude frequency response and finds an optimal approximation to it by iteratively solving an optimal complex approximation (which we can easily do).

Start with a specified desired magnitude frequency response and no specified phase response. Applied a guess as to the phase response and construct the desired complex response. Find the optimal least squares complex approximation to that using formulas. Take the phase from that and apply it to the original desired magnitude. Then, do another complex approximation. In other words, we keep the ideal magnitude response and update the phase each iteration until convergence occurs (if it does!). These ideas have been used by Soewito [46], Jackson [29], Kay and probably others.

References

[1] Arthur Albert. *Regression and the Moore-Penrose Pseudoinverse*. Academic Press, New York, 1972.

[2] V. Ralph Algazi and Minsoo Suk. On the frequency weighted least-square design of finite duration filters. *IEEE Transactions on Circuits and Systems*, 22(12):943–953, December 1975.

[3] Mohamed S. Bani and Bruce L. Chalmers. Best approximation in via iterative hilbert space procedures. *Journal of Approximation Theory*, 42:173–180, 1984.

[4] J. A. Barreto and C. S. Burrus. Complex approximation using iterative reweighted least squares for fir digital filters. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages III:545–548, IEEE ICASSP-94, Adelaide, Australia, April 19-22 1994.

[5] J. A. Barreto and C. S. Burrus. Iterative reweighted least squares and the design of two-dimensional fir digital filters. In *Proceedings of the IEEE International Conference on Image Processing*, volume I, pages I:775–779, IEEE ICIP-94, Austin, Texas, November 13-16 1994.

[6] Adi Ben-Israel and T. N. E. Greville. *Generalized Inverses: Theory and Applications*. Wiley and Sons, New York, 1974. Second edition, Springer, 2003.

[7] Ake Bjorck. *Numerical Methods for Least Squares Problems*. Blaisdell, Dover, SIAM, Philadelphia, 1996.

[8] C. S. Burrus. Vector space methods in signal and system theory. Unpublished notes, ECE Dept., Rice University, Houston, TX, 1994, 2012. also in Connexions: http://cnx.org/content/col10636/latest/.

[9] C. S. Burrus. Prony, pade, and linear prediction for the time and frequency domain design of iir digital filters and parameter identification. *Connexions: http://cnx.org/content/m45121/latest/*, 2012.

[10] C. S. Burrus and J. A. Barreto. Least p-power error design of fir filters. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, volume 2, pages 545–548, ISCAS-92, San Diego, CA, May 1992.

[11] C. S. Burrus, J. A. Barreto, and I. W. Selesnick. Reweighted least squares design of fir filters. In *Paper Summaries for the IEEE Signal Processing Society's Fifth DSP Workshop*, page 3.1.1, Starved Rock Lodge, Utica, IL, September 13-16 1992.

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

[13] C. Sidney Burrus. Constrained least squares design of fir filters using iterative reweighted least squares. In *Proceedings of EUSIPCO-98*, pages 281–282, Rhodes, Greece, September 8-11 1998.

[14] C. Sidney Burrus. Convergence of constrained optimal design of fir filters using iterative reweighted least squares. In *Proceedings of the IEEE DSP Workshop - 1998*, Bryce Canyon, Utah, August 9-12 1998. paper #113.

[15] C. Sidney Burrus. *Digital Signal Processing and Digital Filter Design*. Connexions, cnx.org, 2008. http://cnx.org/content/col10598/latest/.

[16] Richard H. Byrd and David A. Pyne. Convergence of the iteratively reweighted least squares algorithm for robust regression. Technical report 313, Dept. of Mathematical Sciences, The Johns Hopkins University, June 1979.

[17] B. L. Chalmers, A. G. Egger, and G. D. Taylor. Convex approximation. *Journal of Approximation Theory*, 37:326–334, 1983.

[18] E. W. Cheney. *Introduction to Approximation Theory*. McGraw-Hill, New York, 1966. Second edition, AMS Chelsea, 2000.

[19] Chong-Yung Chi and Shyu-Li Chiou. A new self-initiated wls approximation method for the design of two-dimensional equiripple fir digital filters. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, pages 1436–1439, San Diego, CA, May 1992.

[20] Nassim N. Chit and John S. Mason. Complex chebyshev approximation for fir digital filters. *IEEE Transactions on Signal Processing*, 39(1):49–54, January 1991.

[21] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C. Sinan Gunturk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63(1):1–38, January 2010.

[22] Daniel Tavora de Queiroz Cobra. Fir digital filter design by the method of successive projections. Masters thesis, MIT, January 1986.

[23] David L. Donoho. For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829, June 2006. http://stats.stanford.edu/~donoho/Reports/2004/L1L0EquivCorrected.pdf.

[24] Hakan Ekblom. Calculation of linear best l_p-approximations. *BIT*, 13(3):292–300, 1973.

[25] John D. Fisher. *Design of Finite Impulse Response Digital Filters*. Ph. d. thesis, Rice University, 1973.

[26] R. Fletcher. *Practical Methods of Optimization*. John Wiley & Sons, New York, second edition, 1987.

[27] R. Fletcher, J. A. Grant, and M. D. Hebden. The calculation of linear best approximations. *Computer Journal*, 14:276–279, 1971.

[28] Irina F. Gorodnitsky and Bhaskar D. Rao. Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 45(3), March 1997.

[29] Leland B. Jackson. A frequency-domain steiglitz-mcbride method for least-squares iir filter design, arma modeling, and periodogram smoothing. *IEEE Signal Processing Letters*, 15(8):49–52, 2008.

[30] S. W. Kahng. Best approximation. *Mathematics of Computation*, 26(118):505–508, April 1972.

[31] 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.

[32] C. L. Lawson. *Contribution to the Theory of Linear Least Maximum Approximations*. Ph. d. thesis, University of California at Los Angeles, 1961.

[33] Yuying Li. A globally convergent method for problems. Technical report TR 91-1212, Computer Science Department, Cornell University, Ithaca, NY 14853-7501, June 1991.

[34] Y. C. Lim, J. H. Lee, C. K. Chen, and R. H. Yang. A weighted least squares algorithm for quasi-equripple fir and iir digital filter design. *IEEE Transactions on Signal Processing*, 40(3):551–558, March 1992.

[35] J. S. Mason and N. N. Chit. New approach to the design of fir digital filters. *IEE Proceedings, Part G*, 134(4):167–180, 1987.

[36] G. Merle and H. Spaeth. Computational experiences with discrete l_p-approximation. *Computing*, 12:315–321, 1974.

[37] M. R. Osborne. *Finite Algorithms in Optimization and Data Analysis*. Wiley, New York, 1985.

[38] T. W. Parks and C. S. Burrus. *Digital Filter Design*. John Wiley & Sons, New York, 1987. most is in Connexions: http://cnx.org/content/col10598/latest/.

[39] M. J. D. Powell. *Approximation Theory and Methods*. Cambridge University Press, Cambridge, England, 1981.

[40] J. R. Rice. *The Approximation of Functions*, volume 2. Addison-Wesley, Reading, MA, 1969.

[41] John R. Rice and Karl H. Usow. The lawson algorithm and extensions. *Math. Comp.*, 22:118–127, 1968.

[42] S. A. Ruzinsky and E. T. Olsen. L_1 and l_infty minimization via a variant of karmarkar's algorithm. *IEEE Transactions on ASSP*, 37(2):245–253, February 1989.

[43] Jim Schroeder, Rao Yarlagadda, and John Hershey. Normed minimization with applications to linear predictive modeling for sinusoidal frequency estimation. *Signal Processing*, 24(2):193–216, August 1991.

[44] Ivan W. Selesnick. Least squares solutions to linear system of equations, 2009. to be published in Connexions.

[45] Allan J. Sieradski. *An Introduction to Topology and Homotopy*. PWS-Kent, Boston, 1992.

[46] Atmadji W. Soewito. *Least Squares Digital Filter Design in the Frequency Domain*. Ph. d. thesis, Rice University, 1990. At: http://scholarship.rice.edu/handle/1911/16483.

[47] Richard E. Stone and Craig A. Tovey. The simplex and projective scaling algorithms as iteratively reweighted least squares methods. *SIAM Review*, 33(2):220–237, June 1991.

[48] Virginia L. Stonick and S. T. Alexander. A relationship between the recursive least squares update and homotopy continuation methods. *IEEE Transactions on Signal Processing*, 39(2):530–532, Feburary 1991.

[49] Virginia L. Stonick and S. T. Alexander. Globally optimal rational approximation using homotopy continuation methods. *IEEE Transactions on Signal Processing*, 40(9):2358–2361, September 1992.

[50] Gilbert Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley, MA, 1986.

[51] Ching-Yih Tseng. A numerical algorithm for complex chebyshev fir filter design. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, volume 3, pages 549–552, San Diego, CA, May 1992.

[52] Ching-Yih Tseng and Lloyd J. Griffiths. A simple algorithm to achieve desired patterns for arbitrary arrays. *IEEE Transaction on Signal Processing*, 40(11):2737–2746, November 1992.

[53] Karl H. Usow. On approximation 1: Computation for continuous functions and continuous dependence. *SIAM Journal on Numerical Analysis*, 4(1):70–88, 1967.

[54] Ricardo Vargas and C. Sidney Burrus. Iterative design of digital filters. *arXiv*, 2012. arXiv:1207.4526v1 [cs.IT] July 19, 2012.

[55] Ricardo A. Vargas and C. Sidney Burrus. Adaptive iterataive reweighted least squares design of filters. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume III, IEEE ICASSP-99, Phoenix, March 14-19 1999.

[56] Ricardo Arturo Vargas. *Iterative Design of Digital Filters*. Ph. d. thesis, Rice University, ECE Department, Houston, TX 77251, May 2008. in Connexions: http://cnx.org/content/col11383/latest/.

[57] G. A. Watson. An algorithm for linear approximation of continuous functions. *IMA Journal of Numerical Analysis*, 1:157–167, 1981.

[58] G. A. Watson. Convex approximation. *Journal of Approximation Theory*, 55(1):1–11, October 1988.

[59] Andrew E. Yagle. Non-iterative reweighted-norm least-squares local minimization for sparse solutions to underdetermined linear systems of equations, 2008. preprint at http://web.eecs.umich.edu/~aey/sparse/sparse11.pdf.

[60] Wen-Shyong Yu, I-Kong Fong, and Kuang-Chiung Chang. An l_1 approximation based method for synthesizing {FIR} filters. *IEEE Transactions on Circuits and Systems II*, 39(8):578–561, August 1992.