Annales Univ. Sci. Budapest., Sect. Comp. 45 (2016) 303–317

THE METHOD IRLS FOR SOME BEST ℓ_p NORM SOLUTIONS OF UNDER- OR OVERDETERMINED LINEAR SYSTEMS

Csaba J. Hegedüs (Budapest, Hungary)

Communicated by Ferenc Schipp
(Received October 1, 2016; accepted November 7, 2016)

Abstract. The iteratively reweighted least sqaures (IRLS) method is not recommended for the cases of $p = 1, \infty$ among the many approaches to solve such problems. However, some useful properties are pointed out here: It is simple to use, it offers fast convergence for polynomial best uniform approximations and it can be easily extended to multivariate polynomial (or more generally: linear) approximations. It helps finding the extremal set, moreover, it has offered invaluable help to formulate a generalized de la Vallée Poussin type characterization ($p = \infty$), when Haar condition does not apply.

1. Introduction

Let $A \in \mathbb{R}^{m\times n}$ be a rectangular matrix, $b \in \mathbb{R}^m$ and a solution is needed to the system

$$Ax = b. \tag{1.1}$$

Two cases are considered usually. The system is overdetermined in the first case, $m > n$, $b \notin \text{Range}(A)$ and the solution is sought for x such that the residual vector $r(x) = b - Ax$ have a minimal norm:

$$\text{find } x \text{ to minimize } \|r(x)\|. \tag{1.2}$$

Such solution will be called a *minimal residual* solution.

Key words and phrases: Best solutions of under- or overdetermined linear systems, pseudoinverse, Haar condition, weighted least squares, uniform approximation of functions in one or more variables, generalized de la Vallée Poussin characterization.
2010 Mathematics Subject Classification: 15A06, 65F20, 41A50.
https://doi.org/10.71352/ac.45.303

The system is underdetermined in the other case, when usually $m \leq n$ holds, there is a bundle of solutions and we look for the *minimal norm* solution:

$$\min \|x\|, \text{ subject to } Ax = b. \tag{1.3}$$

The power norm will be considered here for the choices: $p = 1, 2, \infty$.

Formulation of such problems can be traced back to the XVIII-th century. Gauss had declared to find his least squares method in 1795. In his survey paper, Watson [19] gives a deep historical account. Regarding problems in l_1 norm, Bidabad [4] also gives interesting details, going back even to Galilei (1632) and Boscovich (1760). Now these methods are common tools in statistics, robotics and also, they have strong connections to best uniform approximations.

Using the Euclidean norm ($p = 2$), the pseudoinverse A^+ gives an easy answer to both problems as A^+b is the minimal residual solution for the overdetermined system and it is the minimal norm solution in the underdetermined case. There are many books in the literature about generalized inverses, among them such as pseudoinverse, see e.g. [1].

This paper considers the pseudoinverse formulation for weighted least squares. By introducing diagonal weight matrices, one can give *iteratively reweighted least squares* (IRLS) methods to find solutions for the other norms. Such methods are shortly discussed in Watson's paper [19] for $1 < p < \infty$, $p \neq 2$. According to Watson, they have been introduced first by Beaton and Tukey [3]. The iteration process converges locally if p is close to 2, and if zero components of r are avoided, it is globally convergent (from any initial approximation) for $1 < p < \infty$ [9]. However, convergence can be slow, particularly as p nears 1 – it is linear with convergence constant $|p - 2|$, as shown by Wolfe [20].

The main results of this paper come from numerical experimentation with IRLS. It has turned out that one can achieve fast convergence for polynomial best uniform approximations and it can be easily extended to multivariate polynomial (or more generally: linear) approximations.

For $p = \infty$ the characterization of the best residual norm solution is easy if the Haar condition is satisfied. If Haar condition does not apply, our experimentation suggests a more complicated characterization by revealing groups in the extremal set with different absolute value levels of the residual vector elements.

Despite of convergence problems, such approach is attractive because of its simplicity and the well elaborated least squares programs. It is also given, how pseudoinverse theory helps to reformulate minimum norm problems to minimum residual ones.

Finally some tricks are considered to assure convergence in the $p = 1$ case.

2. Weighted least squares

The interested reader can find thorough treatment of least squares methods in the book of Björck [2], where weighted versions are also given.

In the case of weights, the norm $\|r\|_W = (r^T W r)^{1/2}$ should be minimized, where W is a symmetric positive definite (SPD) matrix of appropriate size. If A is not of full rank or we have the underdetermined case, we may look for the minimum of $\|x\|_C$, where C is another SPD matrix.

Now introduce the Cholesky decomposition for matrices W and C: $W = L^T L$ and $C = K^T K$, then one has $\|r\|_W^2 = r^T L^T L r$ and the minimum of $\|Lr\|^2$ is sought.

In a similar way, define $y = Kx$ and find the minimum of $\|y\|_2^2$ instead of $\|x\|_C^2 \to \min$.

Substituting y for x gives the minimization problem in terms of y

$$\left\|Lb - LAK^{-1}y\right\|_2^2 \to \min, \tag{2.1}$$

where practically there are no weights, their contribution is incorporated into the new residual. If this problem is solved by the pseudoinverse, then the result can be transformed back to the weighted problem. Our further considerations will be concerned with the pseudo-solutions of (2.1) as

$$y = \left(LAK^{-1}\right)^+ Lb. \tag{2.2}$$

3. Infinity norm solutions

According to [19], the minimal residual problem in this norm goes back to Laplace in 1786. Later on, the intensive work of Chebyshev led to the theory of the uniform approximation of functions, such that many times the term *infinity norm* was replaced by the *Chebyshev norm*.

Nowadays the theory behind infinity norm solutions is well elaborated, one can find abundant materials in the books of Cheney [7] or Watson [18] about the existence, uniqueness or characterization, see also [6]. The reader can find easily accessible material in the Master Thesis of Earle [11] about the existing minimal residual and minimal norm methods in infinity norm.

At first we recall some definitions and conventions. It is assumed here that the system is overdetermined: $m > n$. We shall use the sign vector notation

$s = \mathrm{sign}(r)$ with the meaning $e_i^T s = \mathrm{sign}(e_i^T r)$, e_i is the i-th Cartesian unit vector, ^{T} is the transpose. Matrix A is called of *full rank* if $\mathrm{rank}(A) = \min(m, n)$. It satisfies the Haar condition if each $n \times n$ submatrix is nonsingular. Observe that Haar condition is more demanding than full rank, for example, take the first three columns of a 10×10 unit matrix.

Many times the following situation is found in the solution of the minimal residual problem:

There exists a reordering of the rows in two blocks A_1 and A_2

$$Ax = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} x = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{3.1}$$

having the following properties:

1) A_1 has k rows, they form the *extremal set* of solution.

2) With the aid of A_1 it is possible to set up the *extremal system* of linear equations for finding the solution x^*.

3) The norm of the residual $r_1 = b_1 - A_1 x^*$ is equal to the norm of the full residual: $\|r\|_\infty = \|r_1\|_\infty$.

4) The other residual subvector $r_2 = b_2 - A_2 x^*$ has smaller absolute elements such that $\|r_2\|_\infty < \|r_1\|_\infty$ holds.

We have the following characterization theorem:

Theorem 3.1. *If A satisfies the Haar condition, all elements of r_1 have equal absolute values in the minimal residual solution* (1.2) *($p = \infty$) and $k = n + 1$. Moreover:*

(1) *There is no other vector x for which*

$$\|r(x)\|_\infty < \|r(x^*)\|_\infty . \tag{3.2}$$

(2) *There exists a vector $v^T = [\ v_1^T \quad v_2^T\ ]$ such that $v^T A = 0$, $v_2 = 0$ and the signs of the fully nonzero vector v_1 can be adjusted so that*

$$\mathrm{sign}(v_1) = \mathrm{sign}(r_1(x^*)). \tag{3.3}$$

Proof. It rests on solving the easier problem, when $m = n + 1$. In that case enlarge A_1 by s_1, $\|s_1\|_\infty = 1$ as a last column and form the system

$$[\ A_1 \quad s_1\ ] \begin{bmatrix} x^* \\ \xi \end{bmatrix} = b_1. \tag{3.4}$$

The residual now is $r_1 = \xi s_1$ and (3.4) can be solved for all residual vectors having infinity norm $|\xi|$. As $n + 1$ rows should be linearly dependent, one can

always find a nonzero v_1 for $v_1^T A_1 = 0$. If multiplying the above equation by this v_1^T, one gets for ξ:

$$\xi = \frac{v_1^T b_1}{v_1^T s_1}. \tag{3.5}$$

Here the numerator has a fix value. The smallest absolute ξ results for $s_1 = \text{sign}(v_1)$ in the denominator, that is, (3.3) holds. The found s_1 is unique if all elements of v_1 are nonzero, otherwise some components of s_1 can be freely chosen between $[-1, 1]$. But if v_1 has a zero element, then there exist n linearly dependent rows in A_1 and that contradicts the Haar condition.

For larger m the solution is found by checking all subsets of $n+1$ rows of A and the maximal absolute ξ identifies the extremal set. Here we have locally unique solution for each subset of rows. That explains the $<$ sign in (3.2). See also [7], [18]. ■

This theorem is essentially due to de la Vallée Poussin (1911), who identified the signs from Cramer's rule. Haar's paper [12] appeared in 1918. Later Cadzow [5] found a solution in terms of pseudoinverses for the $(n+1) \times n$ subproblem. The $v^T A = 0$ characterization is the discrete analogue of Kirchberger's 1903 result for the continuous problem [14]. The formulation and proof here can be considered an attempt to collect all observations in one to make things as simple as possible.

For non Haar systems we have found a more complicated characterization by computer experiments. This time the extremal set has g groups. In each group the residual vector elements have the same absolute values – they will be called *group levels* – and they are different between groups. The size of the extremal system is $n+g$. Let the j-th group have n_j elements. Then we have the equality

$$\sum_{j=1}^{g} n_j = n + g. \tag{3.6}$$

The sign vector s_1 is distributed into g subvectors and g new columns are added to matrix A_1. For instance, in case of 3 groups, rows of A_1 are collected in three blocks and we have the equation:

$$\begin{bmatrix} A_{11} & s_{11} & & \\ A_{21} & & s_{21} & \\ A_{31} & & & s_{31} \end{bmatrix} \begin{bmatrix} x^* \\ \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}. \tag{3.7}$$

Now there are three independent left zero space vectors such that $v_j^T A_1 = 0$, $j = 1, 2, 3$, where $A_1^T = \begin{bmatrix} A_{11}^T & A_{21}^T & A_{31}^T \end{bmatrix}$. The *group values* can be given

similarly as before:

$$\xi_j = \frac{v_j^T b_{j1}}{v_j^T s_{j1}}, \quad j = 1, 2, 3. \tag{3.8}$$

If we apply group numbering according to increasing values of group levels $|\xi_j|$, then vectors v_j are such that they have 0 elements for the preceding – or in other words: lower level – groups. This fact necessitates new groups to come in. We also have the same sign rule within one group as in the simple case and all v_j elements proved to be nonzero for their respective group and that assures uniqueness of the solution.

Despite generality of the scheme, one can still give full rank matrices, where it will not apply. Trivial example is having an invertible $n \times n$ block in A and the other rows are zero. A little more sophisticated example is the following:

$$A^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{bmatrix}$$

with b vector:

$$b^T = \begin{bmatrix} 1 & -1 & 2 & -2 & 1 & 1 & 2 & 2 \end{bmatrix}.$$

There are three groups in the solution with number of elements 4,2 and 2, so that altogether 10 rows would be needed in the extremal set, but there are only 8. If we multiply A from the right with the matrix

$$B = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

then the result is

$$B^T A^T = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

showing that the problem falls apart into independent subproblems. This shows some kind of reducibility that may be difficult to notice. However, if we apply random numbers for the right vector b, then we see an ordinary situation.

In the case of rank deficiency, $\operatorname{rank}(A) < \min(m, n)$, one can apply a decomposition of A, e.g. QR-decomposition. Now Q is a full rank matrix and $\operatorname{Range}(A) = \operatorname{Range}(Q)$, thus the minimal residual problem should give the same distance. Replacing A by Q may give a unique minimal residual solution with vector y and then solving $Rx = y$ gives the bundle of solutions of the starting problem.

3.1. The IRLS algorithm for the minimal residual norm problem

Let $r = b - Ax$ be the residual vector and denote by $|r|$ the vector of absolute elements of r. Introduce vector $e^T = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^T$ and the diagonal matrix

$$W = W(r) = \operatorname{diag}(|r|)/\left\|r\right\|_\infty . \tag{3.9}$$

Then it is easy to see that

$$\lim_{k\to\infty} \frac{\left\|W^k r\right\|_2}{\left\|W^k e\right\|_2} = \left\|r\right\|_\infty \tag{3.10}$$

will tend to the infinity norm of r. Hence infinity norm can be approximated by computing weighted two-norms.

If psedoinverse is used as in (2.1), the weighted residual will be minimized and by increasing the power of W, the process should tend to the minimum of the inifinity norm of r.

Now K is chosen to be the identity matrix in (2.1) and the weight matrix $L = W$ is computed iteratively. At the beginning $W = I$ is chosen and the first weight is $W_1 = W(r_1)$, where $x_1 = A^+b$ and $r_1 = b - AA^+b$. The divisor in (3.10) is omitted because it can be considered a constant multiplier and it does not influence minimization. Instead, the weight is normalized such that the largest diagonal element is equal to 1. It is observed computationally that the pseudo-solution of the next problem $W_1A = W_1b$ gives usually a smaller infinity norm of the residual. Therefore the next weight is computed to the residual of the first weighted system: $W_1r_2 = W_1(b - A(W_1A)^+W_1b)$,

$$W_2 = W(r_2)$$

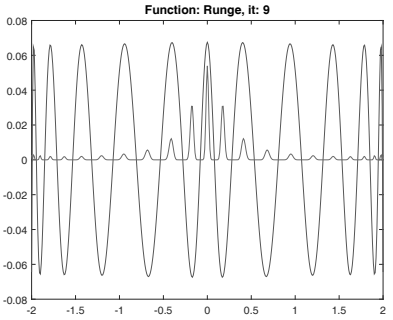
such that the updated weight will be $\gamma_2 W_2 W_1$. Here γ_2 is the normalizer.

Having the k-th iterate x_k, the belonging residual vector will be denoted by $r_k = b - Ax_k$ and the k-th weight update is :

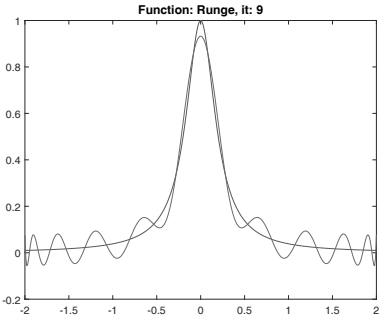
$$W_k = \gamma_k \operatorname{diag}(|r_k|)W_{k-1}. \tag{3.11}$$

This choice approximates the infinity norm of r as the larger elements are accentuated. The convergence can slow down if a weight component gets nearly zero at a place where maximum should be in the solution. To avoid such events, restarting may be applied with some higher power of the residual values.

This algorithm was tested for weighted orthogonal polynomials and tabulated function values. The results were unexpectedly good. For instance, Runge's example: $1/(1+25x^2)$ was approximated in Chebyshev norm by a poly-



Error and weight function



Approximating polynomial

nomial of degree 20 and the result is shown in the figures. The oscillating error function and the positive valued weight function are shown in the left figure. These functions have different order of magnitude such that the weight function, having maximal value 1, was proportionally adjusted to the magnitude of the error function. The resulting polynomial and Runge's function are shown in the other figure. It is close to the best uniform approximating polynomial after the ninth step of iteration. The functions were tabulated between -2 and 2 with 201 equally spaced points. The speed of convergence was similar to that of the Remez algorithm such that it was not necessary to use the de la Vallée Poussin characterization by solving for the extremal set. For weighted least squares there is a hidden self-adjungate tridiagonal matrix in the background, the good behaviour can be attributed to it.

However, convergence was not so fast when applied to an overdetermined set of general linear equations. The speed of convergence was even slower in cases when having residuals nearly the same in absolute value. We have found oscillations in the weights and residuals such that it was necessary to introduce some damping. Actually $\mathrm{diag}(|r_k|)$ was replaced by

$$\mathrm{diag}(|r_k|)/\left\|r_k\right\|_\infty + 1/k \tag{3.12}$$

in (3.11). In this way we could get convergence in all randomly generated, not very large cases. Another successful damping was to apply the convex linear combination

$$W_{k-1} + \left(\mathrm{diag}(|r_k|)\left\|r_k\right\|_\infty - W_{k-1}\right)\beta \tag{3.13}$$

instead of $\mathrm{diag}(|r_k|)$ in (3.11) with $\beta = \min\{1, k/10\}$, where k is the step size. This method applies damping only in the first 9 steps. After some while of experimentation, we have found the choice (3.12) appropriate.

If convergence is slow then it may be enough to get to a stage where it is possible to identify the extremal set and the belonging sign vector. Then the exact solution may be computed by (3.4).

When looking for numbers in the iteration process, it is interesting to observe that elements of W_k are tending to zero for rows not belonging to the extremal set, thus such rows are eradicated from the approximation process. But the multipliers for rows of the extremal set are not necessarily tending to 1. Usually one of them is equal to 1 and the other values tend to a number between 0 and 1.

Thus numerical experimentation suggests a new characterization of the extremal set: There exists a diagonal submatrix $\widetilde{W}$ having diagonal elements in $(0, 1]$ such that the pseudosolution of the system

$$\widetilde{W}A_1 x = \widetilde{W}b_1 \tag{3.14}$$

gives the minimal residual solution: $x^* = (\widetilde{W}A_1)^+\widetilde{W}b_1$.

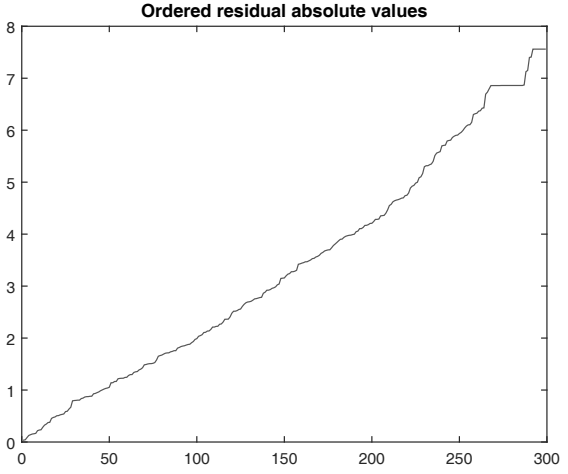
We have used Matlab's `pinv` function in the tests. It computes the pseudoinverse by solving the singular value problem that is also used to find the rank of the matrix. For computing the pseudoinverse, rank revealing QR-factorization [8] can also be used. It was shown in [13] that Gram-Schmidt orthogonalization with reorthogonalization is a competitive alternative. Then we may compute the factorization $A = QR$ or $A = Q_1 B Q_2^T$, where B is a nonsingular bidiagonal matrix. In the case of full rank, the pseudoinverse can be calculated by the formula $A^+ = R^{-1}Q^T$. Otherwise, $A^+ = Q_2 B^{-1} Q_1^T$ may be used in the rank loss case. We did our experimentation in Matlab and one step of iteration was:

```
One step of IRLS algorithm (MATLAB)
  x=pinv(diag(w)*A)*diag(w)*b; r=b-A*x;
  plot(1:m,r,1:m,0.8*max(abs(r))*w);
  w=w.*(abs(r)/norm(r,'inf')+ic^(-1));
```

Among others a 299×24 matrix problem was encountered coming from a l_∞ norm spline approximation problem to human heart cardiogram data. More than 200 steps of iteration were needed to identify groups in the exremal set. The data of the extremal set can be seen in the following table:

Group	Number of elements	Group level
1	6	6.8586
2	13	6.8604
3	1	6.8702
4	1	7.1251
5	1	7.1500
6	1	7.3995
7	1	7.4065
8	8	7.5599

The extremal equation here was 32×32 and we could check all properties given in the previous paragraph. Regarding the zero space vectors v_j, they were computed as the last 8 lines of the inverse. Only the zeros of v_2 in the first six positions were not very convincing (not accurate for 15 digits), but it may be explained with the ill-conditioned character of the matrix: The first two group levels are close. Such events may give a hint that the sign vectors should be put in as first vectors of the extremal matrix. The plot of the residual vector in the solution, ordered by increasing absolute values, can be seen in the next figure.



Ordered residual absolute values in the solution

The matrix and the Matlab codes can be downloaded from:

`http://numanal.inf.elte.hu/~hegedus/bestnormpack.zip`

3.2. Minimum norm solutions

For computing minimum norm problems, the previously given method can be used. One can do it by finding the explicit form of the bundle of solutions and apply the residual norm method. Pseudoinverse theory, see [1], gives an elegant approach. The bundle of solutions can be expressed by

$$x = x_+ + (I - A^+A)t,$$

where $x_+ = A^+ b$ and $t \in \mathbb{R}^n$, otherwise arbitrary. Now rearrange this equation into the new linear system:

$$(A^+ A - I)t = x_+ \tag{3.15}$$

and find the minimal residual solution. The residual vector belonging to the here calculated solution t^* gives the solution of the starting problem:

$$x^* = x_+ + (I - A^+ A)t^*. \tag{3.16}$$

The algorithm for the problem

$$\begin{bmatrix} 5 & 3 & -1 & 3 \\ 0 & -4 & 4 & 5 \end{bmatrix} x = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

of [11] will be shown here. The weights are 1's at the beginning and the second damping method is used, see (3.13). The subsequent resulting weights are shown in the next table:

$$\begin{array}{cccc} 0.9455 & 1 & 0.9818 & 0.9636 \\ 0.8112 & 1 & 0.9379 & 0.8751 \\ 0.6256 & 1 & 0.8785 & 0.7581 \end{array}$$

It is seen, the first component is the lowest in all steps and it is diminishing the fastest. The rank is known to be 2 , therefore we seek 3 rows into the extremal set. The first row is neglected and rows 2, 3, 4 are kept for the extremal set. The residual vector in the third step is:

$$0.3675 \quad 0.3484 \quad -0.3768 \quad -0.4198 \ .$$

The belonging sign vector is chosen accordingly [1 -1 -1]. Now we delete the first row in the system (3.15), attach the column $\begin{bmatrix} 1 & -1 & -1 \end{bmatrix}^T$ to the matrix, form the augmented system (3.4) and get the pseudoinverse solution:

$$0.1057 \quad -0.0936 \quad 0.0067 \quad -0.0803 \quad 0.3846$$

The residual norm in the optimal solution is shown by the last element: 0.3846. The first four elements give the solution vector t^*. The residual to this solution comes from (3.16) and it is

$$0.3231 \quad 0.3846 \quad -0.3846 \quad -0.3846 \ ,$$

the sought solution vector to the starting problem.

4. 1-norm solutions

For an overview of methods, see a recent lecture in [15], where a good account is given with many references therein. It is followed by a research paper of the authors [16] about a new suggested method.

The minimum residual problem in l_1 norm is also called as the Least Absolute Deviation (LAD) problem. It can be treated with the previous ideas, only the weighting should be changed. The here chosen weight update formula were

$$W_{k+1} = W_k \left(W_k \operatorname{diag}(|r_k|)^{-1/2} + i^{-2} \right)^{1/2},$$

where the damping i^{-2} was added to all components of $|r_k|$. It turned out in testing that some rows got too small factors in the iterations in the initial phase of computation. Finally it was hard to get into the extremal set. To overcome this difficulty, we have applied restarting sometimes by choosing $W_k = I$.

The characterization of the solution for the minimal residual problem is given in [18], [17], we cite it from [15]:

Theorem 4.1. *Let $A \in \mathbb{R}^{m\times n}$, $m > n$, be a rectangular matrix of full column rank and $b \in \mathbb{R}^m$ a given vector. Then there exists a permutation matrix $\Pi \in \mathbb{R}^{m\times m}$ such that*

$$\Pi A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \ \Pi b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

$$A_1 \in \mathbb{R}^{n\times n}, \quad A_2 \in \mathbb{R}^{(m-n)\times n}, \quad b_1 \in \mathbb{R}^n, \quad b_2 \in \mathbb{R}^{m-n},$$

whereby A_1 is a nonsingular matrix and there exists a LAD-solution $x^* \in \mathbb{R}^n$ such that $A_1 x^* = b_1$. Furthermore, if matrix $\begin{bmatrix} A_1 & b \end{bmatrix}$ satisfies the Haar condition, then the solution x^* of the system $A_1 x = b_1$ is a solution of the LAD problem if and only if vector

$$v = \left(A_1^T\right)^{-1} A_2^T s, \quad s = \operatorname{sign}(b_2 - A_2 x^*)$$

satisfies $\|v\|_\infty \le 1$. Moreover, x^* is a unique solution if and only if $\|v\|_\infty < 1$.

This theorem also refers to an extremal set that can be identified after enough number of iterations and then the accurate solution can be computed.

In case of lower rank, Watson still gives the result, that rank(A) number of rows are needed in the extremal set to find a solution that always exists.

Applying the weighted least squares method to the example found in [15], five rows of six are good in the extremal set after the second step of iteration.

Only the 14-th step will bring in the missing vector. This shows once again that some steps of iteration may serve useful information even if the convergence is not fast.

It is shown in the following, how the minimum norm problem can be treated. Now $m < n$ holds and the matrix has an LQ-factorization as $A = LQ^T$. (Transpose the $A^T = QR$ form.) $A^+ = QL^{-1}$ in the full rank case and the bundle of solution is given by

$$x_+ + \left(I - QQ^T\right) t = x_+ + PP^T t.$$

Introducing the new parameter vector $y = -P^T t$ will not effect the goal of finding the minimal residual solution for

$$Py = x_+.$$

Here P is a full column rank matrix having orthogonal columns and if it satisfies the Haar condition, the previous characterization applies. Observe that this trick can also be used in the case of Chebyshev norm for the minimal norm solution.

5. Conclusions

Although the IRLS method is not considered the first choice when solving best l_p-norm problems for linear systems, it is found that it performs beautifully for finding the best uniform polynomial approximation to functions even in multidimensional cases. The here given example was a tabulated version for functions but the applied scalar products may be replaced by integrals for more accurate results.

Despite the fact that it is not recommended for l_1 and l_∞ problems, still it may provide useful information in these cases.

Acknowledgements. Professor Ferenc Schipp is kindly appreciated for calling attention to results and literature in the field. Also, my colleague Péter Kovács is thanked for supplying the cardiogram data and his useful criticism and comments.

References

[1] **Ben Israel, A. and T.N.E. Greville,** *Generalized Inverses: Theory and Applications,* 2nd ed., Springer, New York (2002) http://www.maths.ed.ac.uk/~aar/papers/bengrev.pdf

[2] **Björck, Å.,** *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.

[3] **Beaton, A.E. and J.W. Tukey,** The fitting of power series, meaning polynomials, illustrated on band-spectrographic data, *Technometrics*, **16** (1974), 147–185.

[4] **Bidabad, Bijan,** L_1 *Norm Based Computational Algorithms*, Homepage, http://www.bidabad.com/doc/l1-article6.pdf

[5] **Cadzow, J.A.,** An efficient algorithmic procedure for obtaining a minimum l_∞-norm solution to a system of consistent linear equations, *SIAM Journal on Numerical Analysis*, **11(6)** (1974), 1151–1165.

[6] **Cadzow, J.A.,** Minimum l_1, l_2 and l_∞ norm approximate solutions to an overdetermined system of linear equations, *Digital Signal Processing*, **12(4)** (2002), 524–560.

[7] **Cheney, E.W.,** *Introduction to Approximation Theory, 2nd ed.*, AMS Chelsea Publishing, Providence, RI, 1982.

[8] **Chan, T.F.,** Rank revealing QR factorizations, *Lin. Alg, and its Applic.*, **88/89** (1987), 67–82.

[9] **Dutter, R.,** Robust regression: different approaches to numerical solutions and algorithms, *Research Report No 6*, Technische Hochschule, Zurich, 1975.

[10] **de la Vallée Poussin, C.J.,** Sur la methode de l'approximation minimum, *Soc. Scient. de Bruxelles, Ann.* 2^{me} *partie, Memoires*, **35** (1911), 1–16.

[11] **Earle, A.C.,** *Minimum* l_∞ *Norm Solutions To Finite Dimensional Algebraic Underdetermined Linear Systems*, Master of Science Dissertation, Univ. of the Witwatersrand, 2014, 134 pages. http://wiredspace.wits.ac.za/bitstream/handle/10539/16858/20140910_Infinity20Norm%20Minimization_MSc%20Thesis_EarleAC.pdf?sequence=1

[12] **Haar, A.,** Die Minkowskische Geometrie und die Annäherung an stetige Funktionen, *Math. Ann.*, **78** (1918), 294–311.

[13] **Hegedüs, Cs.J.,** Reorthogonalization methods revisited, *Acta Polytechnica Hungarica*, **12(8)** (2015), 7–28. http://dx.doi.org/10.12700/APH.12.8.2015.8.1

[14] **Kirchberger, P.,** Über Tchebychefsche Annäherungsmethoden, *Math. Ann.*, **57** (1903), 509–540.

[15] **Kušec, G., I. Kuzmanović, K. Sabo and R. Scitovski,** A new method for searching an L_1 solution of an overdetermined sytem of linear equations and applications, *Lecture slides. Univ. Osijek*, Croatia. http://www.mathos.unios.hr/~scitowsk/Seminari/KOI08_System-prezentacija.pdf

[16] **Sabo, K., R. Scitovski and I. Vazler,** Searching for a best LAD-solution of an overdetermined system of linear equations motivated by searching for a best LAD-hyperplane on the basis of given data, *Journal of Optimization Theory and Applications*, **149** (2011), 293–314.

[17] **Pinkus, A.,** *On L_1-approximation*, Cambridge University Press, New York, 1993.

[18] **Watson, G.A.,** *Approximation Theory and Numerical Methods*, Wiley, Chichester, New York, 1980.

[19] **Watson, G.A.,** Approximation in normed linear spaces, *Journal of Computational and Applied Mathematics*, **121** (2000), 1–36.

[20] **Wolfe, J.M.,** On the convergence of an algorithm for discrete L_p approximation, *Numer. Math.*, **32** (1979), 439–459.

Cs. J. Hegedüs
Department of Numerical Analysis
Faculty of Informatics
Eötvös Loránd University
H-1117 Budapest
Pázmány Péter sétány 1/C
Hungary
hegedus@numanal.inf.elte.hu