A Steepest Ascent Method for the Chebyshev Problem

By Marcel Meicler*

Abstract. In this paper we present an efficient ascent method for calculating the minimax solution of an overdetermined system of linear equations $Ax = b$. The algorithm makes best use of all the information available at each cycle in order to force a very steep path to the solution.

1. Introduction. The following notation will be used: x^T is the transpose of the vector x. $\langle x, y\rangle = x^T y$ is the inner product of x and y. A^+ will denote the pseudoinverse of the matrix A; i.e., A^+ satisfies the four Penrose equations [9], $AA^+A = A$, $A^+AA^+ = A^+$, $(AA^+)^T = AA^+$, and $(A^+A)^T = A^+A$.

2. The Least-Squares and Chebyshev Residual for the $n + 1 \times n$ Case. We now consider the system of linear equations $Ax = b$ where $A(n + 1 \times n)$ is the coefficient matrix whose rows are assumed to satisfy the Haar condition and $b(n + 1 \times 1)$ is the data vector.

LEMMA 1. *The pseudoinverse of the matrix*

$$A = \begin{pmatrix} D \\ v \end{pmatrix},$$

where D has maximal column rank is given by

$$A^+ = (D^+ - Jw : J) ,$$

with

$$J = \alpha D^+ w^T$$
$$w = vD^+$$
$$\alpha = 1/(1 + ww^T) .$$

Proof. See [1], [4] or [5].

LEMMA 2. *The least-squares residual is given by*

$$d = (AA^+ - I)b .$$

Proof. See [7].

THEOREM 1. *If*

$$A = \begin{pmatrix} D \\ v \end{pmatrix} \quad \textit{and} \quad b = \begin{pmatrix} b_D \\ b_v \end{pmatrix},$$

then

Received April 22, 1968, revised January 15, 1969.

* Lockheed Electronics Company, Houston, Texas 77058, presently with UNIVAC DPD, Houston, Texas 77006.

$$d = \alpha(b_v - wb_D)\begin{pmatrix} w^T \\ -1 \end{pmatrix},$$

where

$$\alpha = 1/(1 + ww^T).$$

Proof. The theorem is obtained by direct substitution into Lemma 2.

THEOREM 2. *With the same notation as in Theorem* 1, *the amplitude* ϵ *of the Chebyshev residual* t *is given by*

$$\epsilon = \frac{|b_v - wb_D|}{1 + \sum_{i=1}^n |w_i|}. \tag{2}$$

Proof. It is known that $\epsilon = d^T d/d^T\sigma$, where σ is the sign vector of d [7]. Thus, if we let $\beta = \alpha(b_v - wb_D)$,

$$\epsilon = \frac{\beta^2[ww^T + 1]}{\beta[w\sigma' + 1]} = \frac{|\beta|}{\alpha}\left(\frac{1}{w\sigma' + 1}\right) = \frac{|b_v - wb_D|}{1 + \sum_{i=1}^n |w_i|}$$

where σ' is σ restricted to its first n components.

COROLLARY 1. *The Chebyshev residual* t *itself is given by* $t = \epsilon\,\sigma$.

Proof. See [7].

3. Discussion. The expression (2) of the Chebyshev amplitude is similar to those obtained by Moursund in [8] and Bartels and Golub in [2]. As pointed out by the referees, further investigations might result in interesting relationships between these expressions.

4. The General Case. Let $Ax = b$ be an inconsistent system of m linear equations in n unknowns ($m > n$), such that the rows of A satisfy the Haar condition. Let A_J be an arbitrary matrix consisting of $n + 1$ rows of A. If a row partition of A_J is given by

$$A_J = \begin{pmatrix} D \\ v \end{pmatrix}$$

with v a row vector, then D is nonsingular and an expression for the Chebyshev amplitude is given by (2).

We will now present several lemmas which will enable us to describe an exchange algorithm.

If a row of the matrix A_J is replaced by a vector p, the following two cases are expected.

Case 1. The last row of A_J is changed. To obtain the new w, we must compute $\overline{w} = p\,D^{-1}$.

Case 2. If any other row of A_J is changed, the next two lemmas will provide a method for computing $\overline{w}$.

LEMMA 3. *Let* D *be a nonsingular matrix and* $C_1, C_2, \cdots, C_n$ *the columns of its inverse. Let* $\overline{D}$ *be the matrix obtained by replacing the jth row of* D *by vector* p. *If* $\lambda = \langle p^T, C_j\rangle \neq 0$, *then* $\overline{D}$ *is nonsingular and the columns of its inverse are given by*

$$\bar{C}_j = \lambda^{-1} C_j ,$$
$$\bar{C}_i = C_i - \langle p^T, C_i \rangle \bar{C}_j , \qquad i \neq j .$$

Proof. (See [3, p. 49].)

LEMMA 4. *If the jth row of A_J is replaced by vector p, then $\bar{w}$ is obtained by*

$$\bar{w}_j = \lambda^{-1} w_j$$

and

$$\bar{w}_i = w_i - \langle p^T, C_i \rangle \bar{w}_j , \qquad i \neq j .$$

Proof. Since $\bar{w} = v\bar{D}^{-1}$ and $\bar{D}^{-1}$ is given by Lemma 3.

LEMMA 5. $\bar{w} b_{\bar{D}} = w b_D + \beta[b_p - pCb_D]$.

Proof. Let $\gamma_i = \langle p^T, C_i \rangle$, $i \neq j$ where $\beta = \bar{w}_j$

$$\bar{w} b_{\bar{D}} = \sum_{i \neq j} \bar{w}_i b_{D_i} + \bar{w}_j b_p$$
$$= \sum_{i \neq j} w_i b_{D_i} - \beta \sum_{i \neq j} \gamma_i b_{D_i} + \beta b_p .$$

Also

$$\beta[b_p - pCb_D] = \beta\left[b_p - \sum_{i \neq j} \gamma_i b_{D_i} \right] - w_j b_{D_j}$$

and

$$w b_{\bar{D}} = \sum_{i \neq j} w_i b_{D_i} + w_j b_{D_j} .$$

Thus

$$\bar{w} b_D = w b_D + \beta[b_p - pCb_D] .$$

LEMMA 6. $\sum |\bar{w}_i| = |\beta| + \sum_{i \neq j} |w_i - \beta w_j|$.

Proof. Since $\bar{w}\, b_{\bar{D}}$ is independent of b_j, we obtain the lemma by equating the like coefficients of the b's.

LEMMA 7. *The minimax solution for any subsystem A_J is given by $x = D^{-1}(b_D + \epsilon\sigma)$, where σ is the sign vector of w.*

Proof. Let t be the minimax residual. By Corollary 1 $t = \epsilon\sigma$. Since the system $A_J\, x = b_J + t$ is consistent, we can solve

$$Dx = b_D + \epsilon\sigma .$$

LEMMA 8. *The residual k_p corresponding to any row vector p of A is given by*

$$k_p = px - b_p = pD^{-1} b_D - b_p + \epsilon p D^{-1} \sigma .$$

Proof. Use Lemma 7.

LEMMA 9. *If $|k_p| > \epsilon$, then there exists an index $j \in J$ such that the exchange of p and A_j will ensure a greater minimax amplitude ϵ'.*

Proof. (See [6, Theorem 5, p. 77].)

LEMMA 10. *Let N_v and N_p represent $b_v - wb_D$ and $b_p - pCb_D$, respectively. Then*

$$\epsilon' = \frac{|N_v - N_p|}{1 + \sum |\bar{w}_i|} .$$

Proof.

$$\epsilon' = \frac{|b_v - \overline{w}b_{\overline{D}}|}{1 + \sum |\overline{w}_i|} = \frac{|b_v - wb_D - N_p|}{1 + \sum |\overline{w}_i|} = \frac{|N_v - N_p|}{1 + \sum |\overline{w}_i|}.$$

5. Algorithm. We are now ready to describe the algorithm. We choose any n rows of A to form a basis matrix D and express all other rows of A in terms of that basis. Since any vector p, not in D, together with the n elements of the basis will generate an ϵ, we obtain with the use of (2), $(m - n)$ different values of ϵ. The largest one of these will be the current value, ϵ', of the minimax amplitude. Lemmas 7 and 8 enable us to compute the residual vector k. Lemma 9 guarantees at least one possible exchange for the largest residual (in absolute value) and with the aid of Lemma 10 we can predict the value of any interchange. We will follow the maximum value. The outgoing row will be part of the basis since v, the last row of A_J, is a member of the set which generates ϵ'. The process is terminated when the maximum residual (in absolute value) equals the amplitude.

6. Example.**

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 1 & 2 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad b = \begin{bmatrix} 6.9 \\ 7.2 \\ 7 \\ 3 \\ 1 \end{bmatrix}.$$

We have the following sequence:

2	1	6.9		1	0	0
3	1	7.2		3/2	−1/2	−3.15
1	2	7	→	1/2	3/2	3.55
1	1	3		1/2	1/2	− .45
1	−1	1		1/2	−3/2	−2.45

			ϵ	k		
1	0	0	*	*		
0	1	0	*	*	ϵ_1'	ϵ_2'
5	−3	−5.9	.655	1.3	.775	.98
2	−1	−3.6	.9	.9		
−4	3	7.0	.875	− .7		

Thus the pivot is (3, 2).

			ϵ	k		
1	0	0	*	*		
5/3	−1/3	−1.966	.655	.66		
0	1	0	*	*		
1/3	1/3	−1.634	.98	.98	ϵ_1'	ϵ_2'
1	−1	1.1	.366	−1.1	1	.775

** This example is a slight modification of the one found in [3, p. 44].

The pivot is (5, 1).

			ϵ	k
1	1	-1.1	.366	$-$ .9
5/3	4/3	-3.8	.95	.8
0	1	0	*	*
1/3	2/3	-2	1	1
1	0	0	*	*

Thus $J = \{3, 4, 5\}$, $\epsilon = 1$, $x = (2, 2)$.

Acknowledgment. The author wants to express his gratitude to the referees for their careful consideration and detailed reading of this paper.

Sperry Rand Corporation
Houston, Texas 77006

1. T. L. Boullion & P. L. Odell, *An Introduction to the Theory of Generalized Matrix Invertibility*, Texas Center for Research, 1966, p. 120.

2. R. H. Bartels & G. H. Golub, "Stable numerical methods for obtaining the Chebyshev solution to an overdetermined system of equations," *Comm. ACM*, v. 11, 1968, pp. 401–406.

3. E. W. Cheney, *Introduction to Approximation Theory*, McGraw-Hill, New York, 1966. MR **36** #5568.

R. E. Cline, "Representations for the generalized inverse of a partitioned matrix," *J. Soc. Indust. Appl. Math.*, v. 12, 1964, pp. 588–600. MR **30** #3106.

5. T. N. E. Greville, "Some applications of the pseudo-inverse of a matrix," *SIAM Rev.*, v. 2, 1960, pp. 15–22. MR **22** #1067.

6. D. C. Handscomb, *Methods of Numerical Approximation*, Pergamon Press, London, 1965.

7. M. Meicler, "Chebyshev solution of an inconsistent system of $n + 1$ linear equations in n unknowns in terms of its least-squares solution," *SIAM Rev.*, v. 10, 1968, pp. 373–375.

8. D. Moursund, "Chebyshev solution of $n + 1$ linear equations in n unknowns," *J. Assoc. Comput. Mach.*, v. 12, 1965, pp. 383–387. MR **31** #6362.

9. R. Penrose, "A generalized inverse for matrices," *Proc. Cambridge Philos. Soc.*, v. 51, 1955, pp. 406–413. MR **16**, 1082.