Mathematical Programming 29 (1984) 234–242
North-Holland

A NOTE ON THE COMPUTATION OF AN ORTHONORMAL BASIS FOR THE NULL SPACE OF A MATRIX*

Thomas F. COLEMAN
Computer Science Department, Cornell University, Ithaca, NY 14853, USA

Danny C. SORENSEN
Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA

Received 31 August 1982
Revised manuscript received 19 July 1983

A highly regarded method to obtain an orthonormal basis, Z, for the null space of a matrix A^T is the QR decomposition of A, where Q is the product of Householder matrices. In several optimization contexts $A(x)$ varies continuously with x and it is desirable that $Z(x)$ vary continuously also. In this note we demonstrate that the *standard implementation* of the QR decomposition does *not yield* an orthonormal basis $Z(x)$ whose elements vary continuously with x. We suggest three possible remedies.

Key words: Orthonormal Basis, QR Decomposition, Constrained Optimization.

1. Introduction

The question we are addressing in this short note is this: Let B be a ball around a point $x^* \in \mathbb{R}^n$. Suppose that $A(x)$ is an n by t matrix of rank t whose elements vary continuously with x on B. Is it possible to construct, stably and efficiently, a matrix $Z(x)$ with elements which *vary continuously* with x in B and with the additional properties

$$A(x)^T Z(x) = 0, \tag{1.1}$$

$$Z(x)^T Z(x) = I_{(n-t)} \quad ? \tag{1.2}$$

Several techniques for nonlinearly constrained optimization problems require the availability of a matrix $Z(x)$ with properties (1.1) and (1.2). (See, for example, Bartels and Conn (1982), Coleman and Conn (1982a, b), Kaufman (1975), Murray and Wright (1978), Murray and Overton (1980), Tanabe (1981), and Wright (1979)). Theoretical results given in Coleman and Conn (1982a, b) explicitly require that the elements of $Z(x)$ vary continuously in a ball around x^*, where x^* is a solution to the nonlinear programming problem. Kaufman *assumes* differentiability

* Work supported in part by the Applied Mathematical Sciences Research Program (KC-04-02) of the Office of Energy Research of the U.S. Department of Energy under Contract W-31-109-Eng-38.

of $Z(x)$. The other references are not as explicit in their dependence on continuity; however it would appear that possible future theoretical developments concerning projected quasi-Newton methods would also require that $Z(x)$ vary continuously. Surprisingly, the standard implementation of the QR factorization of $A(x)$, using Householder matrices (elementary reflectors), does not necessarily yield a matrix $Z(x)$ with continuously varying elements.

In Section 2 we support this claim in detail. We suggest three possible remedies in Section 3. The suggested techniques ensure the existence of a ball B around the solution in which $Z(x)$ is continuous (the user has little control over the radius of B). While our concern here is largely theoretical, we might point out that the remedies would be very easy to implement.

2. The standard implementation

A well-accepted procedure to obtain an orthonormal basis for the null space of A^T is given by Gill and Murray [1974]: Construct an orthogonal matrix $Q=(Q_1, Q_2)$ such that

$$Q_1^T A = R, \tag{2.1}$$

where R is t by t and upper triangular, and

$$Q_2^T A = 0. \tag{2.2}$$

We can then identify Z with Q_2. Unquestionably, the most popular method for obtaining such a Q is the formation of a product of Householder matrices. Let us consider the simple case when $t=1$ and $A=a=(a_1, a_2, \ldots, a_n)^T$. The textbook rule for constructing Q is

$$Q \leftarrow I - \frac{2uu^T}{u^T u} \quad \text{where } u = a + \operatorname{sgn}(a_1)\|a\|e^1, \text{ and}$$

$$\operatorname{sgn}(a_1) = 1 \text{ if } a_1 \geq 0,$$

$$= -1 \text{ if } a_1 < 0. \tag{2.3}$$

(The vector $(1, 0, \ldots, 0)^T$ is denoted by e^1.) Now suppose that each component $a_i(x)$ is a continuous function of x in B. We wish to examine the continuity of Q with respect to $a(x)$. To do this it is useful to partition Q in the following way:

$$Q = \begin{pmatrix} q_{11} & q_{*1}^T \\ q_{*1} & \bar{Q} \end{pmatrix}$$

(Note that

$$Q_1 = \begin{pmatrix} q_{11} \\ q_{*1} \end{pmatrix}$$

and the columns of

$$Q_2 = \begin{pmatrix} q_{*1}^{\mathrm{T}} \\ \bar{Q} \end{pmatrix}$$

are orthonormal bases for the range space of $a(x)$ and null space of $a(x)^{\mathrm{T}}$, respectively.) It is straightforward to show that

$$u^{\mathrm{T}}u = 2\|a\|\{\|a\| + \mathrm{sgn}(a_1)a_1\},$$

and hence

$$q_{11} = \frac{-\mathrm{sgn}(a_1)a_1}{\|a\|}, \qquad q_{j1} = \frac{-\mathrm{sgn}(a_1)a_j}{\|a\|} \quad \text{for } j > 1,$$

$$\bar{Q}_{ij} = \frac{-a_i a_j}{\|a\| \cdot \{\|a\| + \mathrm{sgn}(a_1)a_1\}} \quad \text{for } i \neq j, \text{ and } \bar{Q}_{ii} = 1 - \frac{a_i^2}{\|a\| \cdot \{\|a\| + \mathrm{sgn}(a_1)a_1\}}.$$

It is clear that q_{11} and $\bar{Q}$ are continuous with respect to $a(x)$, however q_{*1} is discontinuous at the plane $a_1 = 0$. It follows that Q_2 is discontinuous at the plane $a_1 = 0$.

Therefore we cannot, in general, assume continuity of Q_2 when Q is computed in the standard way—this is unfortunately true even for B of arbitrarily small radius. Note that when $t = 1$ the only situation that is troublesome (for B of arbitrarily small radius) is when $a_1(x^*) = 0$. This observation leads us to the first of three possible strategies described in Section 3.

We note that the elements of Q_1 do not change continuously with x. However, this is of no great concern since a continuously varying orthonormal basis for the range space of $A(x)$ is trivially available given Q_1. It is only necessary to monitor the signs of the diagonal elements of R and the corresponding columns of Q_1. Such a simple solution is not available for $Z(x)$.

3. Variations of the standard method

For simplicity of presentation, we restrict ourselves to the case $t = 1$. The extension to the general case is straightforward and we will not go into detail.

a. Row orderings

Suppose that x^* is the point of convergence and $\|a(x^*)\| \neq 0$. Hence there is an ordering of the rows of $a(x^*)$ such that $a_1(x^*) \neq 0$. Therefore, if this ordering is used for all x in B then $\mathrm{sgn}(a_1(x))$ is equal to $\mathrm{sgn}(a_1(x^*))$ for $\|x - x^*\|$ sufficiently small. Considering the formula for Q given above, it is clear that in this case Q varies continuously.

Therefore, maintaining a continuous matrix $Z(x)$ in a neighborhood B of x^* is possible (for B of sufficiently small radius) by suitably ordering the rows of $A(x)$ and applying the standard QR decomposition rules. Unfortunately, a suitable ordering is not known a priori. However, given a sequence $\{x^k\}$ that converges to x^*, it is clear that any of a number of row-interchange tests could be employed which would be independent of the given sequence and which would discover a suitable ordering and become inactive for $\|x^k - x^*\|$ sufficiently small. The following interchange tests would serve this purpose:

$$a^k \leftarrow \Pi a(x^k). \tag{3.1}$$

$$\text{if } |a_1^k| \geq \delta \|a^k\| \quad \text{then go to (3.3).} \tag{3.2}$$

(a) $i_k = \min\{i: |a_i^k| = \max |a_j^k|\}$,

(b) $\Pi \leftarrow \Theta(1, i_k)\Pi$,

(c) $a^k \leftarrow \Theta(1, i_k)a^k$

$$\text{Compute } Q_k \text{ as in (2.3)} \tag{3.3}$$

In the program segment above δ is a constant satisfying $0 < \delta < 1/2n$, Π is a permutation matrix and $\Theta(1, i)$ is a permutation matrix that interchanges row 1 with row i when applied to a matrix from the left. It is straightforward to show that Π will be constant for k sufficiently large when $\|a^*\| \neq 0$. The extension to the QR factorization of a matrix is also straightforward. Moreover, it should be noted that while another convergent sequence may produce a different ordering, any given ordering obtained by this process will suffice for all x in a sufficiently small neighborhood of x^*.

b. Maintaining the sign bit

The source of our problems is the sign bit used in the standard rule for computing Q. Is it necessary? That is, can we compute Q as

$$Q \leftarrow I - \frac{2uu^T}{u^T u} \quad \text{where } u = a + \|a\| e^1 \quad ? \tag{3.4}$$

There are two apparent difficulties. Firstly, if $a = -\|a\| e^1$, then u is the zero vector—let us ignore this problem temporarily. Secondly, if a is 'close' to $-\|a\| e^1$, then it would appear that disastrous cancellation may occur in the computation of u and hence Q will be inaccurate. Parlett (1971) disputes the second claim and suggests that disastrous cancellation will not occur under these conditions if u is computed as follows:

$$s \leftarrow \sum_{i>1} a_i^2, \tag{3.5}$$

$$u_1 \leftarrow \frac{-s}{(a_1 - \|a\|)}, \tag{3.6}$$

$$u_j \leftarrow a_j, \quad j = 2, \ldots, n. \tag{3.7}$$

Formulas (3.5)–(3.7) do not involve the subtraction of nearly equal small quantities and thus we do not risk disastrous cancellation.

Therefore the following strategy seems appropriate: If $a_1 \geq 0$, then *compute* u_1 by

$$u_1 \leftarrow a_1 + \|a\|. \tag{3.8}$$

If $a_1 < 0$, then *compute* u_1 by (3.5)–(3.7). In either case we can obtain Q by (3.4).

Unfortunately, our problems are not over. Indeed the first difficulty, that Q is not defined at $\bar{a} = -\|a\|e^1$, is rather troublesome. The kernel of the problem is this: Q (as defined by (3.4)) does not have a limit point at $\bar{a}$. Hence it is impossible to make an appropriate *definition* of $Q(\bar{a})$. For example, consider that for $i \neq j$, $i \neq 1$, $j \neq 1$, $\|a\| \neq |a_1|$,

$$Q_{ij} = \frac{-a_i a_j}{\|a\| \cdot (\|a\| + a_1)} = \frac{-a_i a_j (\|a\| - a_1)}{\|a\|(\|a\|^2 - a_1^2)}.$$

Hence

$$\lim_{a \to -e^1} Q_{ij}(a) = \lim_{a \to -e^1} \frac{-2a_i a_j}{\|\tilde{a}\|^2}, \tag{3.9}$$

where $\tilde{a} = (a_2, \ldots, a_n)^T$. But if a approaches $-e^1$ along the line $(-1, \varepsilon, \varepsilon, \ldots, \varepsilon)$, then $Q_{ij} \to -2(n-1)$. However, if a approaches $-e^1$ along the line $(-1, \varepsilon, \ldots, \varepsilon, 0, \varepsilon, \ldots, \varepsilon, 0, \varepsilon, \ldots, \varepsilon)$, where the zeroes occur in positions i and j, then $Q_{ij} \to 0$.

Observe—these difficulties occur only when $a(x^*) = \pm\|a(x^*)\|e^1$. Also, if $a(x^*) = +\|a(x^*)\|e^1$ and $\|a(x^*)\| > 0$, then there is a ball around x^* for which $a(x) \neq -\|a(x^*)\|e^1$, and vice versa. Therefore, if $a(x^*) \neq -\|a(x^*)\|e^1$, then formula (3.4) can be used for all x in a ball B around x^*. The elements of Q will vary continuously on B provided the radius of B is sufficiently small. Alternatively, if $a(x^*) \neq +\|a(x^*)\|e^1$, then (3.4) can be replaced with

$$Q \leftarrow I - \frac{2uu^T}{u^T u} \quad \text{where } u = a - \|a\|e^1. \tag{3.4'}$$

If $a_1 < 0$, then we can *compute* u_1 by

$$u_1 \leftarrow a_1 - \|a\|. \tag{3.5'}$$

If $a_1 > 0$, then we can *compute* u_1 by

$$u_1 \leftarrow \frac{-s}{a_1 + \|a\|}. \tag{3.6'}$$

The elements of Q will vary continuously provided the radius of B is sufficiently small.

Unfortunately, one does not know, a priori, if $a(x^*) = \pm\|a(x^*)\|e^1$. However, it is clear that several switching rules could be employed in conjunction with (3.4) and (3.4')—if $x^k \to x^*$ the switching rule would become inactive for sufficiently large k.

For example, let $\{x^k\}$ be a sequence which converges to x^*. Denote $a(x^k)$ by a^k. A corresponding sequence of elementary reflectors can be defined by

$$\gamma_k \leftarrow \frac{\sigma_{k-1} \cdot a_1^k}{\|a^k\|}, \tag{3.10}$$

$$\text{if } \gamma_k \geq -\delta \quad \text{then } \sigma_k \leftarrow \sigma_{k-1} \quad \text{else } \sigma_k \leftarrow \operatorname{sgn}(a_1^k), \tag{3.11}$$

$$u^k \leftarrow a^k + \sigma_k \|a^k\| e^1, \tag{3.12}$$

$$Q_k \leftarrow I - \frac{2u^k (u^k)^T}{(u^k)^T u^k}. \tag{3.13}$$

To begin, choose $\sigma_0 = \operatorname{sgn}(a_1^0)$, and always *compute* u_1 at step (3.12) by formulae (3.5)—(3.8), if $\sigma_k = 1$, or by formulae (3.5′)—(3.6′) if $\sigma_k = -1$. The parameters γ and δ are introduced in an attempt to maintain the previous sign bit σ_{k-1}. This, in turn, results in the elements of $Z(x)$ (or $Q_2(x)$) behaving in a continuous manner. The parameter δ must satisfy $\delta < 1$, and should be positive in order to express a reluctance to change signs: say $\delta = 0.9$.

c. *Elementary rotation matrix*

The third strategy that we investigate shares some features with the approach described above but is based upon elementary rotation matrices rather than reflectors. If q_1, q_2 are two vectors of unit length with $q_1 \neq -q_2$ then the elementary rotation matrix sending q_1 into q_2 is

$$P = I - (q_1, q_2) D (q_1, q_2)^T \tag{3.14}$$

where

$$D = \frac{1}{(1+\gamma)} \begin{pmatrix} 1 & 1 \\ -(1+2\gamma) & 1 \end{pmatrix},$$

and $\gamma = q_1^T q_2$. Some properties of P are

$$\text{(i) } P^T P = I, \qquad \text{(ii) } P q_1 = q_2, \qquad \text{(iii) } \lim_{q_1 \to q_2} P = I.$$

Also, it can be readily verified that P rotates vectors in the plane, spanned by the vectors q_1 and q_2, through an angle of $\cos^{-1}(\gamma)$ with vectors orthogonal to this plane left untouched. Property (iii) is not shared by general elementary reflectors; it is this property which avoids the need for two definitions of the same transformation which are typically used to implement an elementary reflector stably. In fact if Q is of the form (3.4), with any nonzero vector u, then $\|Q - I\|_2 = 2$—hence Q is never close to the identity transformation.

In the special case $q_1 = a/\|a\|$ and $q_2 = \sigma e^1$ with $\sigma = \pm 1$ the formula for P simplifies to

$$P_\sigma = \begin{pmatrix} \frac{\sigma a_1}{\|a\|} & \frac{\sigma \tilde{a}^T}{\|a\|} \\ \frac{-\sigma \tilde{a}}{\|a\|} & \bar{P}_\sigma \end{pmatrix} \tag{3.15}$$

where $\tilde{a} = (a_2, \ldots, a_n)^T$,

$$\bar{P}_\sigma = I - \left(\frac{1}{1+\gamma}\right)\frac{\tilde{a}\tilde{a}^T}{\|a\|^2} \quad \text{and} \quad \gamma = \frac{\sigma a_1}{\|a\|}.$$

This formula is briefly discussed by Parlett [1980, p. 92, Example 6.3.6). Note that P as defined in (3.15) can be stored and applied to a vector with the same efficiency as an elementary reflector. In fact only trivial modifications to existing QR codes are required to change from reflectors to rotators. Note also that orthogonal bases for the range and null space of $a(x)$ are provided by the first column and remaining $(n-1)$ columns of P_σ^T respectively.

The elementary rotator P is not defined by (3.15) at points satisfying $a = \alpha e^1$, $\sigma\alpha < 0$. However, precisely the same switching rule described in (3.10)–(3.11) may be used in this situation with P_{σ_k} computed at step (3.13) in place of Q_k using formula (3.15). As with the row interchanges, it is straightforward to show that σ_k is constant for all k sufficiently large whenever $a^k \to a^* \neq 0$. This means that for a given sequence $\{x^k\}$, convergent to x^*, the elementary orthogonal transformation will be computed by the same formula for all k sufficiently large. Moreover, while a different sequence may result in different formula in the limit, the formula selected will suffice in a neighborhood of x^*.

4. Concluding remarks

We have suggested three different strategies for maintaining a continuous orthonormal basis for the null space of a matrix A^T which varies continuously with x. The first method has the attraction that the standard QR decomposition implementation can be employed. However, it has the disadvantage that row interchanges may be necessary in order to maintain continuity of $Z(x)$. Nevertheless, if the element of maximum modulus is initially pivoted into the first row (in the case $t = 1$) it seems highly unlikely that many subsequent interchanges would be necessary.

The second procedure (b) does not require interchanges. It is based on the observation (Parlett (1971)) that disastrous cancellation need not occur when u is computed without the 'sign bit' provided the computation is done correctly. Discrete changes are necessary only in the extreme case when $a(x)$ oscillates between $+\|a\|e^1$ and $-\|a\|e^1$—a highly unlikely scenario. On the negative side, this procedure cannot use a standard black box QR decomposition routine.

Finally, the third procedure (c) has all of the advantages and some of the disadvantages attributed to method *b*. The elementary rotator (as described in c) has some additional geometric appeal, however. If the vector a is 'close' to $+\|a\|e^1(-\|a\|e^1)$, then a is rotated into $+\|a\|e^1(-\|a\|e^1)$. The opposite is true for elementary reflectors.

All three of these schemes suffer from a particular malady that requires a significant increase in work to overcome. The problem is that given any one of these strategies and a positive number ε, there is a convergent sequence $\{a^k\}$ which will require a sign switch (or row interchange) for some k with $\|a^k - a^*\| < \varepsilon$. We have not found a scheme that will overcome this difficulty that lies within the framework of the standard QR factorization.

The following is a slight modification of a suggestion by Walter Murray. If $A = QR$ is the current matrix along with its QR factorization and $A^+ = A + E$ then one may compute the QR factorization of A^+ as follows

$$A^+ = Q(R + \hat{E}) \quad \text{where } \hat{E} = Q^{\mathrm{T}}E.$$

Let $B \equiv R + \hat{E}$ and compute $B = P\hat{R}$ where $\hat{R}$ is upper triangular and $P^{\mathrm{T}} = P_{t-1}P_{t-2}\cdots P_1$ has been computed using formula (3.15) with σ chosen by the switching rules (3.10)–(3.13). Then form

$$A^+ = \hat{Q}\hat{R} \quad \text{where } \hat{Q} \equiv QP.$$

If this scheme is applied to a sequence of matrices $\{A^k\}$ that converges to a matrix A^* of full rank then $P^k \to I$. Note that the leading element in the column of B being reduced will be largest in magnitude in a ball around x^*. Therefore, the switching rule will be inactive within this ball and $Z(x)$ will change continuously. This scheme is apparently more expensive than any of the three schemes described above as it involves an additional matrix multiply as well as extra storage.

One may also require that $Z(x)$ have additional smoothness properties such as Lipschitz continuity or perhaps differentiability. It is clear that the strategies discussed in this note will allow $Z(x)$ to inherit all of the smoothness of A, in a ball around x^*, provided the rank of $A(x^*)$ is t.

Another popular way to obtain the QR decomposition of a matrix A is by using a sequence of Givens transformations. If A is sparse this may be the preferred method (see George and Heath (1980), for example), however in the dense case the Givens procedure is roughly twice as expensive as the elementary reflector approach. Continuity difficulties occur when Givens transformations are used also. To see this suppose that $t = 1$ and both $a_1(x) \to 0$ and $a_n(x) \to 0$, where we assume that elements 1 and n define the Givens transformation that introduces a zero into position n. Depending on the manner in which a_1 and a_n converge, the corresponding Givens matrices may jump around wildly—this spells trouble for the continuity of $Z(x)$. Continuity of $Z(x)$ can be achieved in conjunction with the use of Givens transformations, however, if a row interchange strategy (a.) is followed. In particular, if a is the vector to be reduced to a multiple of e^1 then it is only necessary to find

one component which is not zero at the solution (by linear independence, $a \neq 0$). This distinguished component can be used to zero the others, in a fixed order, via Givens transformations.

References

R.H. Bartels and A.R. Conn, "An approach to nonlinear l_1 data fitting", Numerical Analysis Proceedings, 3rd IIMAS Workshop (Cocoyoc, Mexico, 1982).

T.F. Coleman and A.R. Conn, "Nonlinear programming via an exact penalty function: Asymptotic analysis", *Mathematial Programming* 24 (1982a) 123–136.

T.F. Coleman and A.R. Conn, "On the local convergence of a quasi-Newton method for the nonlinear programming problem", Technical Report 82-509, Department of Computer Science, Cornell University (Ithaca, NY, 1982b).

J.A. George and M.T. Heath, "Solution of sparse linear least squares problems using Givens rotations", *Linear Algebra and Its Applications* 34 (1980) 69–83.

P. Gill and W. Murray, *Numerical methods for constrained optimization* (Academic Press, London, 1974).

P.E. Gill, W. Murray, M.A. Saunders and M.H. Wright, "On computing the null space in nonlinear optimization algorithms", Technical Report SOL 83–19, Department of Operations Research, Stanford University (1983).

L. Kaufman, "A variable projection method for solving separable nonlinear least squares problems", *BIT* 15 (1975) 49–57.

W. Murray and M. Overton, "A projected Lagrangian algorithm for nonlinear minimax optimization", *SIAM Journal on Scientific and Statistical Computing* 1 (1980) 345–370.

W. Murray and M. Wright, "Projected Lagrangian methods based on the trajectories of penalty and barrier functions", Technical Report 72-8, Stanford Optimization Laboratory (Stanford, CA, 1978).

B.N. Parlett, "Analysis of algorithms for reflections in bisectors", *SIAM Review* 13 (1971) 197–208.

B.N. Parlett, *The symmetric eigenvalue problem* (Prentice-Hall, Englewood Cliffs, NJ, 1980).

K. Tanabe, "Feasibility-improving gradient-acute-projection methods: a unified approach to nonlinear programming", *Lecture Notes in Numerical and Applied Analysis* 3 (1981) 57–76.

M. Wright, "Algorithms for nonlinearly constrained optimization", Technical Report 79-24, Stanford Optimization Laboratory (Stanford, CA, 1979).