PROCEEDINGS OF THE YEREVAN STATE UNIVERSITY

Physical and Mathematical Sciences 2015, № 3, p. 8–16

Mathematics

MOORE–PENROSE INVERSE OF BIDIAGONAL MATRICES. II

Yu. R. HAKOPIAN[1] *, S. S. ALEKSANYAN [2]**

[1] *Chair of Numerical Analysis and Mathematical Modeling YSU, Armenia*
[2] *Chair of Mathematics and Mathematical Modeling RAU, Armenia*

The present paper is a direct continuation of the paper [1]. Here we start our study of the Moore–Penrose inversion problem for upper bidiagonal matrices with any arrangement of one or more zeros on the main diagonal. In the paper we obtain some preliminary results, which will be used in subsequent, third part of the study.

MSC2010: Primary 15A09; Secondary 65F20.

***Keywords*:** generalized inverse, Moore–Penrose inverse, bidiagonal matrix.

Introduction. In the first part [1] of this work we considered a problem of computing the Moore–Penrose inverse of upper bidiagonal matrices

$$A = \begin{bmatrix} d_1 & b_1 & & & \\ & d_2 & b_2 & & 0 \\ & & \ddots & \ddots & \\ & 0 & & d_{n-1} & b_{n-1} \\ & & & & d_n \end{bmatrix} \tag{1}$$

under the assumptions

$$b_1, b_2, \ldots, b_{n-1} \neq 0 \tag{2}$$

and

$$d_1, d_2, \ldots, d_{n-1} \neq 0\,, d_n = 0\,. \tag{3}$$

The assumption (2) does not restrict the generality of the problem, since if some of over-diagonal entries of the matrix A are zero, then the original problem is decomposed into several similar problems for bidiagonal matrices of lower order. In contrast to the assumption (3), here we will consider bidiagonal matrices of the form (1) with any arrangement of one or more zeros on the main diagonal.

* E-mail: yuri.hakopian@ysu.am ** E-mail: alexosarm@gmail.com

To compute the Moore–Perrose inverse A^+ of the matrix A from (1), we represent it in block form

$$A = \begin{bmatrix} A_1 & B_1 & & & \\ & A_2 & B_2 & & \\ & & \ddots & \ddots & \\ & & & A_{m-1} & B_{m-1} \\ & & & & A_m \end{bmatrix} \tag{4}$$

with diagonal blocks A_k, $k = 1,2,\dots,m$, of the size $n_k \times n_k$ and over-diagonal blocks B_k, $k = 1,2,\dots,m-1$, of the size $n_k \times n_{k+1}$, where $n_1 + n_2 + \dots + n_m = n$. We can carry out the partitioning (4) to get the diagonal blocks A_k having the following types:

type 1 – all diagonal entries of the block, exept the last one, are nonzero;

type 2 – all diagonal entries of the block are zero;

type 3 – all diagonal entries of the block are nonzero.

At the same time we additionally require that two blocks of type 2 are not diagonally adjacent and as a block of type 3 can be only the last block A_m. Then it is easy to see that the described partition (4) of the matrix A is unique.

In Figure we schematically show the selected diagonal blocks (the mark $\times$ stands for a nonzero entry).

$\times$	$\times$		
	$\ddots$	$\ddots$	
		$\times$	$\times$
			0

Type 1

0	$\times$		
	$\ddots$	$\ddots$	
		0	$\times$
			0

Type 2

$\times$	$\times$		
	$\ddots$	$\ddots$	
		$\times$	$\times$
			$\times$

Type 3

The types of diagonal blocks.

By virtue of the partitioning rule, the blocks B_k, $k = 1,2,\dots,m-1$, have the following structure:

$$B_k = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \\ \Delta_k & 0 & \dots & 0 \end{bmatrix}, \qquad \Delta_k \equiv b_{n_1+n_2+\dots+n_k}\,. \tag{5}$$

A Way of Computing the Moore–Penrose Invertion. Here we outline the path of finding the Moore–Penrose inverse based on the block structure of the matrix A. For this purpose we make use of well-known equality

$$A^{+} = \lim_{\varepsilon \to +0} (A^T A + \varepsilon I)^{-1} A^T, \tag{6}$$

where I is the identity matrix (see [2], for example).

Proceeding from (4), we have

$$A^T A + \varepsilon I = \begin{bmatrix} L_1(\varepsilon) & A_1^T B_1 & & & \\ B_1^T A_1 & L_2(\varepsilon) & A_2^T B_2 & 0 & \\ & \ddots & \ddots & \ddots & \\ & 0 & B_{m-2}^T A_{m-2} & L_{m-1}(\varepsilon) & A_{m-1}^T B_{m-1} \\ & & & B_{m-1}^T A_{m-1} & L_m(\varepsilon) \end{bmatrix},$$

where

$$L_1(\varepsilon) = A_1^T A_1 + \varepsilon I_1\,, \tag{7}$$

$$L_k(\varepsilon) = A_k^T A_k + B_{k-1}^T B_{k-1} + \varepsilon I_k\,, \quad k = 2, 3, \ldots, m \tag{8}$$

(here and below I_k stands for the identity matrix of order n_k). Since each A_k, $1 \le k \le m-1$, is a block of type 1 or type 2, we have $A_k^T B_k = 0$, $1 \le k \le m-1$. Consequently $A^T A + \varepsilon I$ is a block diagonal matrix

$$A^T A + \varepsilon I = \begin{bmatrix} L_1(\varepsilon) & & & & \\ & L_2(\varepsilon) & & 0 & \\ & & \ddots & & \\ & 0 & & L_{m-1}(\varepsilon) & \\ & & & & L_m(\varepsilon) \end{bmatrix}. \tag{9}$$

Having the block forms (4) and (9), we can write the matrix $(A^T A + \varepsilon I)^{-1} A^T$ as follows:

$$(A^T A + \varepsilon I)^{-1} A^T =$$

$$= \begin{bmatrix} L_1(\varepsilon)^{-1} A_1^T & & & & \\ L_2(\varepsilon)^{-1} B_1^T & L_2(\varepsilon)^{-1} A_2^T & & 0 & \\ & \ddots & \ddots & & \\ & 0 & L_{m-1}(\varepsilon)^{-1} B_{m-2}^T & L_{m-1}(\varepsilon)^{-1} A_{m-1}^T & \\ & & & L_m(\varepsilon)^{-1} B_{m-1}^T & L_m(\varepsilon)^{-1} A_m^T \end{bmatrix}.$$

Hence, according to the equality (6), we find

$$A^{+} = \begin{bmatrix} Z_1 & & & & \\ H_2 & Z_2 & & 0 & \\ & \ddots & \ddots & & \\ & 0 & H_{m-1} & Z_{m-1} & \\ & & & H_m & Z_m \end{bmatrix}, \tag{10}$$

where

$$Z_k = \lim_{\varepsilon\to+0} L_k(\varepsilon)^{-1} A_k^T\,, \quad k = 1,2,\dots,m, \tag{11}$$

and

$$H_k = \lim_{\varepsilon\to+0} L_k(\varepsilon)^{-1} B_{k-1}^T\,, \quad k = 2,3,\dots,m\,. \tag{12}$$

Let us consider the tasks that we will face in the process of computing the blocks Z_k and H_k in the block representation (10) of the matrix A^+.

The computation of the block Z_1 is clear. Indeed, from (7) and (11) we have

$$Z_1 = \lim_{\varepsilon\to+0} (A_1^T A_1 + \varepsilon I_1)^{-1} A_1^T = A_1^+\,. \tag{13}$$

The block A_1 may be a block of type 1 or type 2. The Moore–Penrose invertion of the block of type 1 we have completely studied in the first part [1] of this work. It remains to consider the block of type 2.

Now let us consider the blocks Z_k and H_k for the indices $2 \le k \le m$. As follows from the equalities (11) and (12), the main problem here is to invert the matrices $L_k(\varepsilon)$ defined in (8).

Let $m \ge 2$ (if $m = 1$, then the Moore–Penrose invertion problem is reduced to a problem already solved in [1]). First of all let us find out the structure of the matrices $L_k(\varepsilon)$. Write the block A_k in the form

$$A_k = \begin{bmatrix} d_1^{(k)} & b_1^{(k)} & & & \\ & d_2^{(k)} & b_2^{(k)} & 0 & \\ & & \ddots & \ddots & \\ & 0 & & d_{n_k-1}^{(k)} & b_{n_k-1}^{(k)} \\ & & & & d_{n_k}^{(k)} \end{bmatrix}, \tag{14}$$

where, according to the form of the matrix A from (1),

$$\begin{aligned} d_i^{(k)} &= d_{n_1+\dots+n_{k-1}+i}\,, \quad i = 1,2,\dots,n_k\,, \\ b_i^{(k)} &= b_{n_1+\dots+n_{k-1}+i}\,, \quad i = 1,2,\dots,n_k-1\,. \end{aligned} \tag{15}$$

Then the matrix $L_k(\varepsilon)$ can be written:

$$L_k(\varepsilon) = \\ = \begin{bmatrix} {d_1^{(k)}}^2+\Delta_{k-1}^2+\varepsilon & b_1^{(k)}d_1^{(k)} & & & \\ b_1^{(k)}d_1^{(k)} & {b_1^{(k)}}^2+{d_2^{(k)}}^2+\varepsilon & b_2^{(k)}d_2^{(k)} & 0 & \\ & \ddots & \ddots & \ddots & \\ & 0 & b_{n_k-2}^{(k)}d_{n_k-2}^{(k)} & b_{n_k-2}^{(k)^2}+d_{n_k-1}^{(k)^2}+\varepsilon & b_{n_k-1}^{(k)}d_{n_k-1}^{(k)} \\ & & & b_{n_k-1}^{(k)}d_{n_k-1}^{(k)} & {b_{n_k-1}^{(k)}}^2+{d_{n_k}^{(k)}}^2+\varepsilon \end{bmatrix}, \tag{16}$$

where $\Delta_{k-1} = b_{n_1+\dots+n_{k-1}}$ (see (5)). Thus $L_k(\varepsilon)$ is tridiagonal matrix with a special structure.

We turn now to solving the raised problems.

Block Z_1.

As it is mentioned above, the block A_1 can be one of type 1 or type 2. If A_1 is the block of type 1, then the problem of finding $Z_1 = A_1^+$ has been solved in the first part of this work (see formulae (50)–(52) in [1]). Moreover, the entries of the block Z_1 can be computed using the algorithm **2d/pinv/special** $(A_1 \Rightarrow A_1^+)$ proposed in [1]. The algorithm requires $n_1^2 + O(n_1)$ arithmetic operations.

It remains to consider the case when A_1 is a block of the type 2. If $n_1 = 1$, then $A_1 = [0]_{1\times 1}$ and by this very fact $Z_1 = [0]_{1\times 1}$ (see [2], for example). For $n_1 \geq 2$ we have

$$(A_1^T A_1 + \varepsilon I_1)^{-1} A^T = \begin{bmatrix} 0 & & & & \\ b_1(b_1^2+\varepsilon)^{-1} & 0 & & 0 & \\ & b_2(b_2^2+\varepsilon)^{-1} & 0 & & \\ & & \ddots & \ddots & \\ & 0 & & b_{n_1-1}(b_{n_1-1}^2+\varepsilon)^{-1} & 0 \end{bmatrix}.$$

Taking the limit as $\varepsilon \to +0$, according to the equality (13), we get the lower bidiagonal matrix

$$Z_1 = \begin{bmatrix} 0 & & & & \\ b_1^{-1} & 0 & & 0 & \\ & b_2^{-1} & 0 & & \\ & 0 & \ddots & \ddots & \\ & & & b_{n_1-1}^{-1} & 0 \end{bmatrix}. \tag{17}$$

Combining the above considerations, we give below a computational procedure.

Procedure/Z1 $(A_1, n_1 \Rightarrow Z_1)$.

Type 1:

– the entries of the block Z_1 are recorded by the formulae (50)–(52) derived in [1];

– the entries of the block Z_1 can be computed using the algorithm **2d/pinv/special** $(A_1 \Rightarrow A_1^+)$ constructed in [1] with an order $n_1^2 + O(n_1)$ of arithmetic operations.

Type 2:

– if $n_1 = 1$, then $Z_1 = [0]_{1\times 1}$;

– for $n_1 \geq 2$ the block Z_1 takes the form (17).

End procedure

Now let us turn to the next problem.

Invertion of a Model Matrix $L(\varepsilon)$. Applying the structure (16) of the matrices $L_k(\varepsilon)$, consider a model tridiagonal matrix

$$L(\varepsilon) = A^T A + B^T B + \varepsilon I =$$

$$= \begin{bmatrix} d_1^2+\Delta^2+\varepsilon & b_1 d_1 & & & \\ b_1 d_1 & d_2^2+b_1^2+\varepsilon & b_2 d_2 & 0 & \\ & \ddots & \ddots & \ddots & \\ & 0 & b_{n-2}d_{n-2} & d_{n-1}^2+b_{n-2}^2+\varepsilon & b_{n-1}d_{n-1} \\ & & & b_{n-1}d_{n-1} & d_n^2+b_{n-1}^2+\varepsilon \end{bmatrix}, \tag{18}$$

which is constructed by using the matrices

$$A = \begin{bmatrix} d_1 & b_1 & & & \\ & d_2 & b_2 & 0 & \\ & & \ddots & \ddots & \\ & 0 & & d_{n-1} & b_{n-1} \\ & & & & d_n \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \\ \Delta & 0 & \dots & 0 \end{bmatrix} \tag{19}$$

(we assume that in the matrix A the entries $b_1, b_2, \dots, b_{n-1}$ are nonzero; in the matrix B of the size $l \times n$ the entry Δ is also nonzero).

In a derivation of formulae for the entries of the matrix $L(\varepsilon)^{-1}$ we will consider several cases associated with the arrangement of zeros on the main diagonal of the matrix A given in (19).

<u>Case A</u>: $n \geq 1$ and $d_1, d_2, \dots, d_n \neq 0$.

<u>Case B</u>: $n \geq 2$ and $d_1, d_2, \dots, d_{n-1} \neq 0$, $d_n = 0$.

<u>Case C</u>: $n \geq 1$ $d_1 = d_2 = \dots = d_n = 0$.

Notice that the case A corresponds to the blocks of the type 3 in the block representation (4) of our primary matrix A, while the cases B and C correspond to the blocks of type 1 and 2 respectively.

Having the matrix A from (19), let us introduce the following notation:

$$r_s \equiv \frac{b_s}{d_s}, \quad s = 1, 2, \dots, n-1; \quad r_0 = r_n = 1\,. \tag{20}$$

• **Case A (for $n = 1$).** As obviously follows from (18),

$$L(\varepsilon)^{-1} = \left[\frac{1}{d_1^2 + \Delta^2 + \varepsilon} \right]_{1\times 1}. \tag{21}$$

• **Cases A and B (for $n \geq 2$).** The difference between the cases A and B just is in the value of the entry d_n. Therefore we consider these cases together.

To invert the matrix $L(\varepsilon)$ we apply the computational procedure already used in the first part [1] of this work $\big($see algorithm **3d/inv** $(C \Rightarrow C^{-1})\big)$. Comparing the records of the matrix $L(\varepsilon)$ given in (18) and the matrix C given in (6) from [1], we have

$$c_{ii} = d_i^2 + b_{i-1}^2 + \varepsilon\,,\, i = 1, 2, \dots, n \tag{22}$$

(for the purpose of unification of the records of the formulae, we set $b_0 = \Delta$), and

$$c_{i\,i+1} = b_i d_i\,,\, i = 1, 2, \dots, n-1; \quad c_{i\,i-1} = b_{i-1} d_{i-1}\,,\, i = 2, 3, \dots, n\,. \tag{23}$$

Let us figure out the dependence on ε of the quantities successively computed in the referred algorithm **3d/inv**.

Using the expressions (22) and (23) for the quantities f_i, g_i and h_i computed in the item 1 of the algorithm, we get

$$f_i = \overset{\circ}{f}_i + O(\varepsilon)\,,\, i = 2, 3, \dots, n, \quad \text{where } \overset{\circ}{f}_i = \frac{d_i^2 + b_{i-1}^2}{b_{i-1} d_{i-1}}\,; \tag{24}$$

$$g_i = \frac{b_i d_i}{b_{i-1} d_{i-1}}, \qquad i = 2,3,\ldots,n-1; \tag{25}$$

$$h_i = \overset{\circ}{h}_i + O(\varepsilon)\,,\ i = 1,2,\ldots,n-1, \quad \text{where } \overset{\circ}{h}_i = \frac{d_i^2 + b_{i-1}^2}{b_i d_i}\,. \tag{26}$$

Now turn to the quantities μ_i and ν_i recursively computed in the items 2 and 3 of the algorithm.

The following assertion can be easily obtained using the relations (8) from [1] and the formulae (24), (25). Moreover, it can also be considered as a simple consequence of similar Lemma 1 from [1].

Lemma 1. The quantities μ_i are represented by

$$\mu_i = \overset{\circ}{\mu}_i + O(\varepsilon)\,, \quad i = 1,2,\ldots,n, \tag{27}$$

where the quantities $\overset{\circ}{\mu}_i$ satisfy the following recurrence relations:

$$\begin{aligned} &\overset{\circ}{\mu}_n = 1\,,\ \overset{\circ}{\mu}_{n-1} = -\overset{\circ}{f}_n\,, \\ &\overset{\circ}{\mu}_i = -\overset{\circ}{f}_{i+1}\overset{\circ}{\mu}_{i+1} - g_{i+1}\overset{\circ}{\mu}_{i+2}\,,\ i = n-2, n-3, \ldots, 1\,. \end{aligned} \tag{28}$$

At the same time the quantities $\overset{\circ}{\mu}_i$ computed by the recursion (28) may be given in a closed form. The following assertion can be established by a strightforward calculation.

Lemma 2. The quantities $\overset{\circ}{\mu}_i$ can be written in the form

$$\overset{\circ}{\mu}_i = (-1)^{n-i}\left[\prod_{s=i}^{n-1} r_s + d_n^2 \sum_{k=i}^{n-1} \frac{1}{d_k^2}\left(\prod_{s=i}^{k-1} r_s\right)\left(\prod_{s=k}^{n-1} \frac{1}{r_s}\right)\right], \quad i = 1,2,\ldots,n. \tag{29}$$

Notice that similar Lemma 2 from [1] is a particular case of just formulated statement. The next assertion is a consequence of the formula (29).

Corollary 1. The relation

$$\overset{\circ}{\mu}_i = -r_i\,\overset{\circ}{\mu}_{i+1} + \frac{d_n^2}{d_i^2}\,\frac{1}{\alpha_i}\,, \quad i = 1,2,\ldots,n-1\,, \tag{30}$$

where

$$\alpha_i \equiv (-1)^{n-i}\prod_{s=i}^{n-1} r_s\,, \quad i = 1,2,\ldots,n-1\,, \tag{31}$$

holds.

It can be readily seen that the quantities α_i defined in (31) can be computed recursively:

$$\alpha_{n-1} = -r_{n-1}; \quad \alpha_i = -r_i\alpha_{i+1}\,,\ i = n-2, n-3, \ldots, 1\,. \tag{32}$$

A representation similar to (27) takes place also for the quantities $\overset{\circ}{\nu}_i$. Using the relations (9) from [1] and the formulae (25), (26), we get the following statement.

Lemma 3. The quantities v_i are represented by

$$v_i = \overset{\circ}{v}_i + O(\varepsilon), \quad i = 1,2,\dots,n, \tag{33}$$

where the quantities $\overset{\circ}{v}_i$ satisfy the following recurrence relations:

$$\begin{aligned} &\overset{\circ}{v}_1 = 1\,,\ \overset{\circ}{v}_2 = -\overset{\circ}{h}_1\,, \\ &\overset{\circ}{v}_i = -\overset{\circ}{h}_{i-1}\overset{\circ}{v}_{i-1} - \frac{1}{g_{i-1}}\overset{\circ}{v}_{i-2}\,, \quad i = 3,4,\dots,n\,. \end{aligned} \tag{34}$$

Notice that the last Lemma can also be considered as a consequence of similar Lemma 3 from [1].

The quantities $\overset{\circ}{v}_i$ may be given in a closed form as well.

Lemma 4. The quantities $\overset{\circ}{v}_i$ can be written in the form

$$\overset{\circ}{v}_i = (-1)^{i+1}\left[\prod_{s=1}^{i-1}\frac{1}{r_s} + \Delta^2\sum_{k=1}^{i-1}\frac{1}{b_k^2}\left(\prod_{s=1}^{k} r_s\right)\left(\prod_{s=k+1}^{i-1}\frac{1}{r_s}\right)\right], \quad i = 1,2,\dots,n. \tag{35}$$

The next assertion can be readily obtained from the formula (35).

Corollary 2. The relation

$$\overset{\circ}{v}_{i+1} = -\frac{1}{r_i}\overset{\circ}{v}_i + \frac{\Delta^2}{b_i^2}\beta_i\,, \quad i = 1,2,\dots,n-1\,, \tag{36}$$

where

$$\beta_i \equiv (-1)^i\prod_{s=1}^{i} r_s\,, \quad i = 1,2,\dots,n-1\,, \tag{37}$$

holds.

The quantities β_i defined in (37) can be computed recursively:

$$\beta_1 = -r_1; \quad \beta_{i+1} = -r_{i+1}\beta_i\,, i = 1,2,\dots,n-2\,. \tag{38}$$

Further, in the item 4 of the algorithm **3d/inv** from [1] the quantity

$$t = (c_{11}\mu_1 + c_{12}\mu_2)^{-1}$$

is computed. Using the representation (27) of the quantities μ_i and the formulae (29), (35) by strightforward calculation, we get the following assertion.

Lemma 5. We have

$$t = (\overset{\circ}{t} + O(\varepsilon))^{-1}, \tag{39}$$

where

$$\overset{\circ}{t} = d_n^2\,\overset{\circ}{v}_n + \Delta^2\alpha_1\,. \tag{40}$$

Remark. Having formulae (31) and (35), one can easily show that $\overset{\circ}{t} \neq 0$.

Finally, the entries of the inverse matrix

$$L(\varepsilon)^{-1} = [x_{ij}]_{n\times n} \tag{41}$$

are computed in the items 5 and 6 of the algorithm **3d/inv** from [1]. First we find the entries of the lower triangular part of the matrix, including the main diagonal:

$$x_{ij} = \mu_i v_j t\,, \quad j = 1,2,\dots,n\,,\ i = j, j+1,\dots,n\,. \tag{42}$$

Then we get the entries of the upper triangular part:

$$x_{ij} = \mu_j \nu_i t\,, \quad j = 2,3,\ldots,n\,, \;\; i = 1,2,\ldots,j-1\,. \tag{43}$$

At the same time we note that due to the symmetry of the matrix $L(\varepsilon)^{-1}$ practically it is not necessary to perform computations by the formula (43) , we can simply set $x_{ij} = x_{ji}$.

• **Case C.** As follows from (18), the matrix $L(\varepsilon)$ in this case is diagonal. Hence

$$L(\varepsilon)^{-1} = \begin{bmatrix} (\Delta^2+\varepsilon)^{-1} & & & & \\ & (b_1^2+\varepsilon)^{-1} & & 0 & \\ & & \ddots & & \\ & 0 & & (b_{n-2}^2+\varepsilon)^{-1} & \\ & & & & (b_{n-1}^2+\varepsilon)^{-1} \end{bmatrix}. \tag{44}$$

Thus the inversion process of the matrix $L(\varepsilon)$ is fully described.

Forthcoming Studies. As has been said above, our main objective is to derive a formulae for the entries of the blocks Z_k and H_k involved in the block representation (10) of the matrix A^+. These blocks are defined by means of the equalities (11) and (12) respectively. In a subsequent, third part of the present work we will consider more general problem of computing the model matrices

$$Z = \lim_{\varepsilon \to +0} L(\varepsilon)^{-1} A^T \tag{45}$$

and

$$H = \lim_{\varepsilon \to +0} L(\varepsilon)^{-1} B^T\,, \tag{46}$$

where the matrices $L(\varepsilon)$, A, B are given in (18) and (19).

Received 25.09.2015

REFERENCES

1. **Hakopian Yu.R., Aleksanyan S.S.** Moore–Penrose Inverse of Bidiagonal Matrices. I // Proceedings of the Yerevan State University. Physical and Mathematical Sciences, № 2, 2015, p. 11–20.
2. **Ben-Israel A., Greville T.N.E.** Generalized Inverses. Theory and Applications (2nd ed.). NY: Springer, 2003.