

BAND GENERALIZATION OF THE GOLUB–KAHAN BIDIAGONALIZATION, GENERALIZED JACOBI MATRICES, AND THE CORE PROBLEM*

IVETA HNĚTYNKOVÁ[†], MARTIN PLEŠINGER[‡], AND ZDENĚK STRAKOŠ[§]

Abstract. Definition of the core problem in total least squares (TLS) problems was extended to the multiple right-hand sides case $AX \approx B$ in [I. Hnětynková, M. Plešinger, and Z. Strakoš, *SIAM J. Matrix Anal. Appl.*, 34, 2013, pp. 917–931]. In this paper we prove that the band generalization of the Golub–Kahan bidiagonalization proposed in this context by Å. Björck yields the core problem. We introduce generalized Jacobi matrices and investigate their properties. They prove useful in further analysis of the core problem concept. This paper assumes exact arithmetic.

Key words. total least squares problem, multiple right-hand sides, core problem, Golub–Kahan bidiagonalization, generalized Jacobi matrices.

AMS subject classifications. 15A06, 15A18, 15A21, 15A24, 65F20, 65F25.

1. Introduction. This paper further elaborates on extending the core problem concept to total least squares problems with multiple right-hand sides; see [12]. We will use the same notations as in [12] and very briefly recall some basic facts. Consider a linear approximation problem

$$AX \approx B, \quad \text{or, equivalently,} \quad [B|A] \begin{bmatrix} -I_d \\ X \end{bmatrix} \approx 0, \quad (1.1)$$

where $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{n \times d}$, $B \in \mathbb{R}^{m \times d}$, and $A^T B \neq 0$, without any further assumption on the positive integers m , n , d . The matrices A , B , $[B|A]$, and X are called the *system matrix*, the *right-hand side (or the observation) matrix*, the *extended (or data) matrix*, and the *matrix of unknowns*, respectively. We will focus on incompatible problems, i.e., $\mathcal{R}(B) \not\subset \mathcal{R}(A)$, although, the compatible case is not strictly excluded. Consider the orthogonal transformations

$$\widehat{A}\widehat{X} \equiv (P^T A Q)(Q^T X R) \approx (P^T B R) \equiv \widehat{B}, \quad (1.2)$$

where $P^{-1} = P^T$, $Q^{-1} = Q^T$, $R^{-1} = R^T$; or, equivalently,

$$[\widehat{B}|\widehat{A}] \begin{bmatrix} -I_d \\ \widehat{X} \end{bmatrix} \equiv \left(P^T [B|A] \begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix} \right) \left(\begin{bmatrix} R^T & 0 \\ 0 & Q^T \end{bmatrix} \begin{bmatrix} -I_d \\ X \end{bmatrix} R \right) \approx 0. \quad (1.3)$$

We call problems (1.1) and (1.2)–(1.3) *orthogonally invariant* and require that X solves (1.1) if and only if $\widehat{X} = Q^T X R$ solves (1.2)–(1.3). Within this paper we investigate the most common case of the *total least squares problem* (TLS)

$$\min_{X, E, G} \| [G|E] \|_F \quad \text{subject to} \quad (A + E)X = B + G, \quad (1.4)$$

* This work has been supported by the GAČR grant No. P201/13-06684S.

[†] Faculty of Mathematics and Physics, Charles University in Prague, Prague, Czech Republic, and Institute of Computer Science, AS CR, Prague, Czech Republic (e-mail: hnetynkova@cs.cas.cz).

[‡] Department of Mathematics, Technical University of Liberec, Liberec, Czech Republic, and Institute of Computer Science, AS CR, Prague, Czech Republic (e-mail: martin.plesinger@tul.cz). The research of this author has been partly supported by the ESF grant CZ.1.07/2.3.00/30.0065.

[§] Faculty of Mathematics and Physics, Charles University in Prague, Prague, Czech Republic (e-mail: strakos@karlin.mff.cuni.cz). The research of this author has been partly supported by the ERC-CZ project LL1202.

where $\|\cdot\|_F$ denotes the Frobenius norm.

The first steps in generalizing the core problem concept for $d > 1$ were done by Björck in the series of talks [2], [3], [4], and also in the unpublished manuscript [5], by Sima in [18] and Sima and Van Huffel in [19], and by Plešinger in [17]. Following these works and the paper [10] classifying situations which can occur when $d > 1$, the paper [12] provides a rigorous extension of the core problem concept to TLS with multiple right-hand sides (1.4). The data reduction used there is based on the singular value decomposition (SVD) of the matrix A . The present paper investigates the *band generalization of the Golub–Kahan bidiagonalization* (or simply *band algorithm*), proposed by Björck and proves that the output of this algorithm also provides a core problem in the sense of [12].

The paper is organized in the following way. Section 2 recalls the background results. Section 3 describes the band generalization of the Golub–Kahan bidiagonalization. Section 4 introduces generalized Jacobi matrices and analyzes properties of the band subproblem. Section 5 concludes the paper.

Throughout the text $\mathcal{R}(M)$ and $\mathcal{N}(M)$ denote the range and null-space of a matrix M , respectively; I_ℓ (or just I) denotes an $\ell \times \ell$ identity matrix; e_k denotes the k th column of I ; $0_{\ell,\xi}$ (or just 0) denotes an $\ell \times \xi$ zero matrix; and $\|v\|$ denotes the Euclidean norm of a vector v . The following convention concerning the entries of matrices will simplify the exposition:

- club (\clubsuit) stands for a nonzero entry, $\clubsuit \neq 0$;
- heart (\heartsuit) stands for a general entry which can also be zero;
- empty spaces in matrices always represent zero entries.

Throughout the paper we assume *exact arithmetic*.

2. The core problem and other background results. In order to make the text as self-consistent as possible, we briefly recall the known results used below.

2.1. Core problem. The *core problem* within the problem (1.1) is defined as follows (see [12, Definition 5.2]):

DEFINITION 2.1 (Core problem). *The subproblem $A_{11}X_1 \approx B_1$ is a core problem within the approximation problem $AX \approx B$ if $[B_1|A_{11}]$ is minimally dimensioned and A_{22} maximally dimensioned subject to the orthogonal transformations of the form*

$$P^T[B|A] \begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix} = P^T[BR|AQ] \equiv \left[\begin{array}{c|c} B_1 & 0 \\ \hline 0 & 0 \end{array} \parallel \left[\begin{array}{c|c} A_{11} & 0 \\ \hline 0 & A_{22} \end{array} \right] \right], \quad (2.1)$$

where $P^{-1} = P^T$, $Q^{-1} = Q^T$, $R^{-1} = R^T$.

Let $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$ have k distinct singular values σ_j with multiplicities r_j and the orthonormal bases of the corresponding left singular vector subspaces $U_j \in \mathbb{R}^{\overline{m} \times r_j}$, $j = 1, \dots, k$. Let $r_{k+1} \equiv \dim(\mathcal{N}(A_{11}^T))$, with $U_{k+1} \in \mathbb{R}^{\overline{m} \times r_{k+1}}$ having the orthonormal basis vectors of $\mathcal{N}(A_{11}^T)$ as its columns. Then $U \equiv [U_1, \dots, U_k, U_{k+1}] \in \mathbb{R}^{\overline{m} \times \overline{m}}$ and $U^T = U^{-1}$. The core problem $A_{11}X_1 \approx B_1$ has the following properties (see [12, p. 925]):

(CP1) The matrix $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$ is of *full column rank* equal to $\overline{n} \leq \overline{m}$.

(CP2) The matrix $B_1 \in \mathbb{R}^{\overline{m} \times \overline{d}}$ is of *full column rank* equal to $\overline{d} \leq \overline{m}$.

(CP3) The matrices $\Phi_j \equiv U_j^T B_1 \in \mathbb{R}^{r_j \times \overline{d}}$ are of *full row rank* equal to $r_j \leq \overline{d}$, for $j = 1, \dots, k+1$.

These properties guarantee minimality of the core problem; see [12, section 4]. Dimensions of any subproblem $A_{11}X_1 \approx B_1$ having the properties (CP1)–(CP3) cannot

be reduced by any orthogonal transformation of the form (2.1). Moreover

$$U^T[B_1|A_{11}] = \left[\underbrace{\begin{array}{c} \Phi_1 \\ \vdots \\ \Phi_k \\ \Phi_{k+1} \end{array}}_{\bar{d}} \left| \underbrace{\begin{array}{c} U_1^T A_{11} \\ \vdots \\ U_k^T A_{11} \\ 0 \end{array}}_{\bar{n}} \right. \begin{array}{l} \} r_1 \\ \vdots \\ \} r_k \\ \} r_{k+1} \end{array} \right] \bar{m},$$

where $[U_1, \dots, U_k]^T A_{11}$ is a square nonsingular matrix of the size $\bar{n} \times \bar{n}$, $\bar{n} = r_1 + \dots + r_k$, and Φ_{k+1} is of full row rank $r_{k+1} = \bar{m} - \bar{n}$. Thus (CP1)–(CP3) imply that the extended matrix $[B_1|A_{11}]$ is of *full row rank* equal to \bar{m} , $\max\{\bar{n}, \bar{d}\} \leq \bar{m} \leq \bar{n} + \bar{d}$.

2.2. Golub–Kahan bidiagonalization. Consider first $d = 1$, i.e., the single right-hand side problem $Ax \approx b$. Using the initial vectors $q_0 = 0$ and $p_1 = b/\gamma_1$, where $\gamma_1 = \|b\|$, the Golub–Kahan iterative bidiagonalization¹ computes for $j = 1, 2, \dots$

$$q_j \alpha_j = A^T p_j - q_{j-1} \gamma_j, \quad (2.2)$$

$$p_{j+1} \gamma_{j+1} = A q_j - p_j \alpha_j, \quad (2.3)$$

such that $\|q_j\| = \|p_{j+1}\| = 1$, and $\alpha_j > 0$, $\gamma_{j+1} > 0$. The matrices

$$P_j \equiv [p_1, \dots, p_j] \in \mathbb{R}^{m \times j}, \quad Q_j \equiv [q_1, \dots, q_j] \in \mathbb{R}^{n \times j},$$

have orthonormal columns, $P_j^T P_j = Q_j^T Q_j = I_j$; see [9]. The iterative process (2.2)–(2.3) terminates when the right-hand side of one of the equations becomes zero, i.e., either $q_j \alpha_j = 0$ (in the incompatible case) or $p_{j+1} \gamma_{j+1} = 0$ (in the compatible case) for some j . Consider that $Ax \approx b$ is incompatible, $b \notin \mathcal{R}(A)$, and let $q_{\bar{n}+1} \alpha_{\bar{n}+1} = 0$. Then, denoting $P_1^{\text{cp}} \equiv P_{\bar{n}+1}$ and $Q_1^{\text{cp}} \equiv Q_{\bar{n}}$,

$$(P_1^{\text{cp}})^T [b|AQ_1^{\text{cp}}] = \left[\begin{array}{c|ccc} \gamma_1 & \alpha_1 & & \\ & \gamma_2 & \ddots & \\ & & \ddots & \alpha_{\bar{n}} \\ & & & \gamma_{\bar{n}+1} \end{array} \right] = [b_1|A_{11}] \in \mathbb{R}^{(\bar{n}+1) \times (\bar{n}+1)} \quad (2.4)$$

represents the core problem within $[b|A]$, and

$$P^T [b|A] \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} = \left[\begin{array}{c|cc} b_1 & A_{11} & 0 \\ \hline 0 & 0 & A_{22} \end{array} \right], \quad P \equiv [P_1^{\text{cp}}, P_2^{\text{cp}}], \quad Q \equiv [Q_1^{\text{cp}}, Q_2^{\text{cp}}],$$

where $P_2^{\text{cp}}, Q_2^{\text{cp}}$ are chosen such that $P^{-1} = P^T$, $Q^{-1} = Q^T$; see [15]. A generalization of the Golub–Kahan bidiagonalization for the problems with multiple right-hand sides is given in section 3 below.

2.3. Right-hand side preprocessing. In order to get an equivalent problem with the full column rank right-hand sides matrix, we preprocess B in an analogous way as in [12, section 3.1]. Let $\bar{d} \equiv \text{rank}(B) \leq \min\{m, d\}$, $B \in \mathbb{R}^{m \times d}$. Consider any decomposition of B in the form

$$B = [C, 0]R^T, \quad C \in \mathbb{R}^{m \times \bar{d}}, \quad R \in \mathbb{R}^{d \times d}, \quad (2.5)$$

¹This algorithm is for its close connection to the Lanczos algorithm also called Golub–Kahan–Lanczos bidiagonalization; see, e.g., [1], [3], [4].

where C is of *full column rank*, and R is square orthogonal, i.e. $R^{-1} = R^T$. Multiplication of (1.1) from the right by R gives

$$A(XR) \approx BR, \quad \text{where} \quad XR \equiv [Y, Y'] \in \mathbb{R}^{n \times d}, \quad Y \in \mathbb{R}^{n \times \bar{d}} \quad (2.6)$$

(if $d = \bar{d}$, then it can be considered $R = I_d$, $B = C$, $X = Y$). The original problem (1.1) is in this way split into two subproblems,

$$AY \approx C \quad \text{and} \quad AY' \approx 0, \quad (2.7)$$

where the second problem is homogeneous. Following the arguments in [15], we consider the meaningful solution $Y' \equiv 0$. In this way, the approximation problem (1.1) reduces to $AY \approx C$ in (2.7). The full column rank matrix $C \in \mathbb{R}^{m \times \bar{d}}$ is called the *preprocessed right-hand side*.

Remark 2.2. The LQ decomposition of B gives an example of (2.5); see [12, remark 3.1]. Then the full column rank matrix C is in a lower triangular column echelon form and it is called the *LQ-preprocessed right-hand side*. Alternatively, one can use the SVD of B (see [12, section 3.1]) in the form

$$B = S[\Theta, 0]R^T, \quad S \in \mathbb{R}^{m \times \bar{d}}, \quad \Theta \in \mathbb{R}^{\bar{d} \times \bar{d}}, \quad R \in \mathbb{R}^{d \times d}, \quad (2.8)$$

where S has mutually orthonormal columns and the square nonsingular Θ contains the singular values of B on the diagonal. Then $C \equiv S\Theta$ has (nonzero) mutually orthogonal columns and it is called the *SVD-preprocessed right-hand side*.

3. Band generalization of the Golub–Kahan bidiagonalization. Consider the problem $AY \approx C$, where $C \in \mathbb{R}^{m \times \bar{d}}$ is of full column rank obtained above. As an extension of (2.4), we want to reduce $[C|A]$ to the upper triangular band matrix with (at most) $\bar{d} + 1$ nonzero diagonals (all entries above the \bar{d} th superdiagonal are zero).

3.1. Basic structure of the band algorithm. First, the right-hand side C is transformed to the upper triangular form. Consider the QR decomposition

$$C = P_{(0)}F, \quad F = \begin{bmatrix} F_1 \\ 0 \end{bmatrix}, \quad F_1 = \begin{bmatrix} \gamma_{1,1} & \beta_{1,2} & \cdots & \beta_{1,\bar{d}} \\ & \gamma_{2,2} & \ddots & \vdots \\ & & \ddots & \beta_{\bar{d}-1,\bar{d}} \\ & & & \gamma_{\bar{d},\bar{d}} \end{bmatrix}, \quad (3.1)$$

where $P_{(0)} \in \mathbb{R}^{m \times m}$, $P_{(0)}^{-1} = P_{(0)}^T$, and F_1 is the upper triangular square matrix with a positive diagonal, $\gamma_{j,j} > 0$, $j = 1, \dots, \bar{d}$. If C is the *SVD-preprocessed right-hand side*, then $F_1 = \Theta$ is diagonal containing the singular values of the original right-hand side B , and the first \bar{d} columns of $P_{(0)}$ are the columns of S ; see (2.8). It should be noted that the matrix $P_{(0)}$ as well as the matrices $P_{(k)}$ below, which are all $\in \mathbb{R}^{m \times m}$, are distinct from the matrices $P_j \in \mathbb{R}^{m \times j}$ used above in description of the Golub–Kahan iterative bidiagonalization. Denote $L_{(0)} \equiv P_{(0)}^T A$, then

$$P_{(0)}^T [C|A] = [F|L_{(0)}]. \quad (3.2)$$

It remains to transform $L_{(0)}$ to a lower triangular band matrix with (at most) $\bar{d} + 1$ nonzero diagonals (all entries below the \bar{d} th subdiagonal are zero). This can be done,

e.g., by multiplications of $L_{(0)}$ with suitable Householder matrices $H_{Q,j}$, $H_{P,j}$, $j = 1, 2, \dots$ from the right and left, respectively. Let for $k = 1, 2, \dots$

$$P_{(k)} = P_{(0)}H_{P,1}H_{P,2}\dots H_{P,k} \in \mathbb{R}^{m \times m}, \quad Q_{(k)} = H_{Q,1}H_{Q,2}\dots H_{Q,k} \in \mathbb{R}^{n \times n} \quad (3.3)$$

be orthogonal matrices yielding a transformation

$$P_{(k)}^T[C|AQ_{(k)}] = [F|P_{(k)}^T AQ_{(k)}] \quad (3.4)$$

$$= \left[\begin{array}{cccc|cccccc} \gamma_{1,1} & \beta_{1,2} & \cdots & \beta_{1,\bar{d}} & \alpha_{1,\bar{d}+1} & & & & & & \\ & \gamma_{2,2} & \ddots & \vdots & \beta_{2,\bar{d}+1} & \ddots & & & & & \\ & & \ddots & \beta_{\bar{d}-1,\bar{d}} & \vdots & \ddots & \alpha_{k,\bar{d}+k} & & & & \\ & & & \gamma_{\bar{d},\bar{d}} & \beta_{\bar{d},\bar{d}+1} & & \beta_{k+1,\bar{d}+k} & \heartsuit & \cdots & \heartsuit & \\ & & & & \gamma_{\bar{d}+1,\bar{d}+1} & \ddots & \vdots & \vdots & \vdots & \vdots & \\ & & & & & \ddots & \beta_{k+\bar{d}-1,\bar{d}+k} & \heartsuit & \cdots & \heartsuit & \\ & & & & & & \gamma_{\bar{d}+k,\bar{d}+k} & \heartsuit & \cdots & \heartsuit & \\ & & & & & & & \heartsuit & \cdots & \heartsuit & \\ & & & & & & & \vdots & \vdots & \vdots & \\ & & & & & & & \heartsuit & \cdots & \heartsuit & \end{array} \right], \quad (3.5)$$

$$\text{with } \alpha_{j,\bar{d}+j} > 0, \quad \gamma_{\bar{d}+j,\bar{d}+j} > 0, \quad \text{for } j = 1, \dots, k. \quad (3.6)$$

Denote

$$L_{(j)} \equiv P_{(j)}^T AQ_{(j)} = H_{P,j}^T L_{(j-1)} H_{Q,j}, \quad (3.7)$$

$$\text{and } L_{(j-)} \equiv P_{(j-1)}^T AQ_{(j)} = L_{(j-1)} H_{Q,j}, \quad j = 1, \dots, k. \quad (3.8)$$

The entry $\alpha_{j,\bar{d}+j}$ represents the norm of the trailing subrow (of length $n - j + 1$) of the j th row of $L_{(j-1)}$, and analogously the entry $\gamma_{\bar{d}+j,\bar{d}+j}$ represents the norm of the trailing subcolumn (of length $m - j - \bar{d} + 1$) of the j th column of $L_{(j-)}$. If the first row of $L_{(0)}$ is zero (i.e., $\alpha_{1,\bar{d}+1} = 0$), or if the trailing subcolumn of $L_{(1-)}$ of length $m - \bar{d}$ is zero (i.e., $\gamma_{\bar{d}+1,\bar{d}+1} = 0$), or if the problem does not have enough rows (i.e., $\gamma_{\bar{d}+1,\bar{d}+1}$ does not exist), then the transformation (3.4) to the form (3.5) with the condition (3.6) does not exist. In such case we formally put $k = 0$ and $Q_{(0)} \equiv I_n$. This particular case is discussed later in section 3.2. In the rest of this paragraph we for simplicity consider $\alpha_{j,\bar{d}+j} > 0$, $\gamma_{\bar{d}+j,\bar{d}+j} > 0$, $j = 1, \dots, k$. Denote $p_1, \dots, p_{\bar{d}+k}$ the first $\bar{d} + k$ columns of $P_{(k)}$ and q_1, \dots, q_k the first k columns of $Q_{(k)}$. Using (3.7) rewritten as

$$AQ_{(k)} = P_{(k)}L_{(k)} \quad \text{and} \quad A^T P_{(k)} = Q_{(k)}L_{(k)}^T, \quad (3.9)$$

we can write for Aq_j and $A^T p_j$

$$Aq_j = [p_j, p_{j+1}, \dots, p_{j+\bar{d}-1}, p_{j+\bar{d}}][\alpha_{j,\bar{d}+j}, \beta_{j+1,\bar{d}+j}, \dots, \beta_{j+\bar{d}-1,\bar{d}+j}, \gamma_{\bar{d}+j,\bar{d}+j}]^T, \quad (3.10)$$

$$A^T p_j = [q_{j-\bar{d}}, q_{j-\bar{d}+1}, \dots, q_{j-1}, q_j][\gamma_{j,j}, \beta_{j,j+1}, \dots, \beta_{j,j+\bar{d}-1}, \alpha_{j,\bar{d}+j}]^T. \quad (3.11)$$

Using the initial vectors $p_1, \dots, p_{\bar{d}}$ given by (3.1) and $q_{1-\bar{d}} = \dots = q_0 \equiv 0$, the columns of the $(\bar{d} + k) \times k$ leading principal block of $L_{(k)}$, and the columns q_1, \dots, q_k ,

and $p_{\bar{d}+1}, \dots, p_{\bar{d}+k}$ are iteratively generated by

$$q_j \alpha_{j, \bar{d}+j} \equiv A^T p_j - q_{j-\bar{d}} \gamma_{j, j} - \left(\sum_{i=1}^{\bar{d}-1} q_{j-\bar{d}+i} \beta_{j, j+i} \right), \quad (3.12)$$

$$\beta_{j+i, \bar{d}+j} \equiv p_{j+i}^T A q_j, \quad \text{for } i = 1, \dots, \bar{d} - 1, \quad (3.13)$$

$$p_{\bar{d}+j} \gamma_{\bar{d}+j, \bar{d}+j} \equiv A q_j - p_j \alpha_{j, \bar{d}+j} - \left(\sum_{i=1}^{\bar{d}-1} p_{j+i} \beta_{j+i, \bar{d}+j} \right), \quad (3.14)$$

where $\|q_j\| = \|p_{j+\bar{d}}\| = 1$, $\alpha_{j, \bar{d}+j} > 0$, $\gamma_{\bar{d}+j, \bar{d}+j} > 0$, for $j = 1, 2, \dots, k$. The β entries represent orthogonalization coefficients.

3.2. Deflation in the band algorithm. Now we focus on the case when the right-hand side of (3.12) or (3.14) becomes zero (including the case $k = 0$). Let ℓ be the *first* index for which *either* $q_\ell \alpha_{\ell, \bar{d}+\ell} = 0$ (yielding formally $\alpha_{\ell, \bar{d}+\ell} = 0$), *or* $p_{\bar{d}+\ell} \gamma_{\bar{d}+\ell, \bar{d}+\ell} = 0$ (yielding formally $\gamma_{\bar{d}+\ell, \bar{d}+\ell} = 0$), $1 \leq \ell \leq \min\{n+1, m - \bar{d} + 1\}$. The cases $\ell = n+1$ and $\ell = m - \bar{d} + 1$ represent reaching the number of columns and rows of the system matrix, respectively.

3.2.1. Upper deflation. Let for $\ell < n+1$ we get $q_\ell \alpha_{\ell, \bar{d}+\ell} = 0$. Recall that $\alpha_{\ell, \bar{d}+\ell}$ is the norm of a trailing subrow of the ℓ th row of $L_{(\ell-1)} = P_{(\ell-1)}^T A Q_{(\ell-1)}$ to the right of $\beta_{\ell, \bar{d}+\ell-1}$, therefore

$$P_{(\ell-1)}^T [C|A Q_{(\ell-1)}] = [F|L_{(\ell-1)}] = \begin{bmatrix} \ddots & & & & & & \\ \ddots & & & & & & \\ \ddots & \alpha_{\ell-1, \bar{d}+\ell-1} & & & & & \\ & \beta_{\ell, \bar{d}+\ell-1} & 0 & \cdots & 0 & & \\ & \beta_{\ell+1, \bar{d}+\ell-1} & \heartsuit & \cdots & \heartsuit & & \\ & \vdots & \vdots & & \vdots & & \end{bmatrix}. \quad (3.15)$$

In this case the Householder matrix $H_{Q, \ell}$ is constructed to transform the first row below the ℓ th row having a nonzero trailing subrow (say, the ξ th row) while producing $\alpha_{\xi, \bar{d}+\ell} > 0$.

This *upper deflation* can be easily described using (3.12)–(3.14). Consider that the $(\ell+1)$ th row of $[F|L_{(\ell-1)}]$ has the nonzero trailing subrow. The formula for computing $\alpha_{\ell+1, \bar{d}+\ell} > 0$ and q_ℓ is then given by equating the $(\ell+1)$ th (instead of the ℓ th) columns of $A^T P_{(\ell)} = Q_{(\ell)} L_{(\ell)}^T$; see also (3.9) and (3.11). Formulas (3.12)–(3.14) are then for $j = \ell, \ell+1, \dots$, modified to

$$q_j \alpha_{j+1, \bar{d}+j} \equiv A^T p_{j+1} - q_{j-\bar{d}+1} \gamma_{j, j} - \left(\sum_{i=2}^{\bar{d}-1} q_{j-\bar{d}+i} \beta_{j+1, j+i} \right), \quad (3.16)$$

$$\beta_{j+i, \bar{d}+j} \equiv p_{j+i}^T A q_j, \quad \text{for } i = 2, \dots, \bar{d} - 1, \quad (3.17)$$

$$p_{\bar{d}+j} \gamma_{\bar{d}+j, \bar{d}+j} \equiv A q_j - p_{j+1} \alpha_{j+1, \bar{d}+j} - \left(\sum_{i=2}^{\bar{d}-1} p_{j+i} \beta_{j+i, \bar{d}+j} \right). \quad (3.18)$$

The *number of summands* and *computed coefficients* β is *reduced by one* for all $j \geq \ell$. Each upper deflation changes the pattern of nonzero entries in the band matrix by reducing the *effective bandwidth* from the top by one.

3.2.2. Lower deflation. Let for $\ell < m - \bar{d} + 1$ we get $p_{\bar{d}+\ell} \gamma_{\bar{d}+\ell, \bar{d}+\ell} = 0$. Recall that $\gamma_{\bar{d}+\ell, \bar{d}+\ell}$ is the norm of a trailing subcolumn of the ℓ th column of $L_{(\ell-)} = P_{(\ell-)}^T A Q_{(\ell)}$ below $\beta_{\ell+\bar{d}-1, \bar{d}+\ell}$, therefore

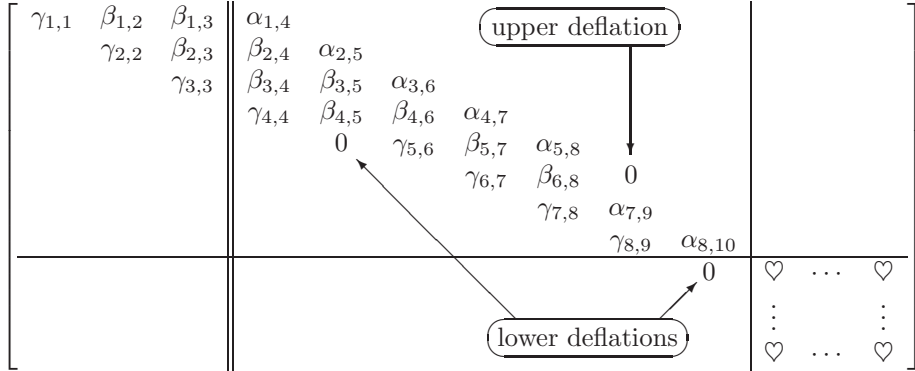
$$P_{(\ell-)}^T [C|A Q_{(\ell)}] = [F|L_{(\ell-)}] = \begin{bmatrix} \ddots & & & \vdots & \vdots \\ & \ddots & & \vdots & \vdots \\ & & \gamma_{\bar{d}+\ell-1, \bar{d}+\ell-1} & \beta_{\ell+\bar{d}-1, \bar{d}+\ell} & \heartsuit & \cdots \\ & & & 0 & \heartsuit & \cdots \\ & & & \vdots & \vdots & \\ & & & 0 & \heartsuit & \cdots \end{bmatrix}. \quad (3.19)$$

Then we take $H_{P, \ell} = I_m$. The matrix $L_{(\ell-)}$ in (3.19) is multiplied by $H_{Q, \ell+1}$ from the right, giving $\alpha_{\ell+1, \bar{d}+\ell+1}$, and the algorithm proceeds with transformation of the $(\ell + 1)$ th column (provided its trailing subcolumn is nonzero). For capturing this *lower deflation* analogously as above, it is convenient to consider a row-oriented formulation of (3.12)–(3.14). Each lower deflation modifies the pattern of nonzero entries in the band matrix by reducing the effective bandwidth from the bottom by one.

3.2.3. Band subproblem. Since the matrix $[F|L_{(k)}]$ has $(\bar{d} + 1)$ nonzero diagonals (see (3.5)), after \bar{d} deflations the effective bandwidth is reduced to one. Denote $P \in \mathbb{R}^{m \times m}$, $Q \in \mathbb{R}^{n \times n}$ the products of the resulting Householder matrices (see (3.3)) and denote $L \equiv P^T A Q$. Then

$$P^T [C|A Q] = [F|L] \equiv \left[\begin{array}{c|c|c} B_1 & A_{11} & 0 \\ \hline 0 & 0 & A_{22} \end{array} \right] \quad (3.20)$$

and the problem is decomposed in the desired subproblems, see, e.g., the following illustration



Let $A_{11} \in \mathbb{R}^{\bar{m} \times \bar{n}}$, $B_1 \in \mathbb{R}^{\bar{m} \times \bar{d}}$, and denote $P_1^{\text{cp}} \equiv [p_1, \dots, p_{\bar{m}}] \in \mathbb{R}^{m \times \bar{m}}$, $Q_1^{\text{cp}} \equiv [q_1, \dots, q_{\bar{n}}] \in \mathbb{R}^{n \times \bar{n}}$. In the rest of this paper we show that the *band subproblem*

$$(P_1^{\text{cp}})^T [C|A Q_1^{\text{cp}}] = [B_1|A_{11}] \in \mathbb{R}^{\bar{m} \times (\bar{n} + \bar{d})} \quad (3.21)$$

represents the core problem, by proving that it satisfies the properties (CP1)–(CP3); see section 2.1.

An implementation of the band algorithm with inputs A and B , and outputs A_{11} , B_1 , P_1^{cp} , Q_1^{cp} , and R (see (2.5)) can be found in Appendix A. For alternative implementations see [18, Algorithm 2.4, p. 38] or [17, Algorithm 5.1, p. 74].

4. Core problem in the band form. The equations (3.20)–(3.21) immediately give

$$A^T P_1^{\text{cp}} = Q \left[\begin{array}{c|c} A_{11}^T & \\ \hline 0 & \end{array} \right], \quad \text{and} \quad [C|AQ_1^{\text{cp}}] = P \left[\begin{array}{c|c} B_1 & A_{11} \\ \hline 0 & 0 \end{array} \right],$$

which represent QR decompositions of matrices $A^T P_1^{\text{cp}}$ and $[C|AQ_1^{\text{cp}}]$, respectively. The matrix A_{11} is in the lower triangular column echelon form with nonzero columns, thus it is of *full column rank* \bar{n} giving the property (CP1). The right-hand side B_1 is in the upper triangular form with nonzero entries on the diagonal (see (3.1)), thus it is of *full column rank* \bar{d} giving the property (CP2). Further $[B_1|A_{11}]$ is in the upper triangular row echelon form with nonzero rows, thus it is of *full row rank* \bar{m} giving the inequality

$$\max\{\bar{n}, \bar{d}\} \leq \bar{m} \leq \bar{n} + \bar{d}. \quad (4.1)$$

Note that for $\bar{d} = 1$ the matrix B_1 becomes a vector b_1 , the band algorithm becomes the standard Golub–Kahan bidiagonalization of A , the matrices $[b_1|A_{11}]$, A_{11} become bidiagonal with $[b_1|A_{11}]^T [b_1|A_{11}]$, $A_{11} A_{11}^T$, and $A_{11}^T A_{11}$, representing Jacobi matrices (symmetric tridiagonal matrices with positive subdiagonal entries). This relationship has been used in [13] and [11]. Jacobi matrices represent thoroughly studied objects with the origin in the first half of the 19th century; see the historical note 3.4.3 in [14, section 3.4, pp. 108–136]; see also [16, Chapter 7, pp. 119–150], [20, section 5, §36–§48, pp. 299–316], and [7, Chapter 1.3, pp. 10–20].

In the following we introduce generalized Jacobi matrices, discuss their spectral properties, and show their relationship to the band subproblem with $\bar{d} > 1$. In particular, we investigate bases of eigenspaces of generalized Jacobi matrices in section 4.1, and we show that $A_{11} A_{11}^T$ represents generalized Jacobi matrix in section 4.2. As a consequence, the bases of the left singular vector subspaces of A_{11} have the properties guaranteeing that the band subproblem $[B_1|A_{11}]$ satisfies also the property (CP3). Other generalizations of Jacobi matrices can be found, e.g., in [6, Chapter 3].

4.1. Generalized Jacobi matrices. Let $T \in \mathbb{R}^{n \times n}$ be a symmetric matrix with entries $t_{k,j}$. In analogy to the notation in, e.g., [8, section 4.1], we consider for $k = 1, \dots, n$

$$f(k) = \min\{j : t_{k,j} \neq 0\}, \quad \text{and} \quad h(k) = k - f(k). \quad (4.2)$$

The number $f(k)$ is the column index of the first nonzero entry in the k th row of T (provided it exists), and $h(k)$ is the distance between this and the diagonal entry. Consider the following matrices.

DEFINITION 4.1 (ρ -wedge-shaped matrix). *Let $T \in \mathbb{R}^{n \times n}$ be a symmetric matrix, and $\rho, 1 \leq \rho < n$, an integer. If $h(k)$ for $k = \rho+1, \dots, n$ is positive and nonincreasing, then we call T a ρ -wedge-shaped matrix.*

For clarity we give some examples of 3-wedge-shaped matrices:

$$\left[\begin{array}{cccccccc} \heartsuit & \heartsuit & \heartsuit & \clubsuit & & & & \\ \heartsuit & \heartsuit & \heartsuit & \heartsuit & \clubsuit & & & \\ \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \clubsuit & & \\ \clubsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \clubsuit \\ & \clubsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit \\ & & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit \\ & & & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit \\ & & & & \heartsuit & \heartsuit & \heartsuit & \heartsuit \end{array} \right], \quad \left[\begin{array}{cccccccc} \heartsuit & \heartsuit & \heartsuit & \clubsuit & & & & \\ \heartsuit & \heartsuit & \heartsuit & \heartsuit & & & & \\ \heartsuit & \heartsuit & \heartsuit & \heartsuit & \clubsuit & & & \\ \clubsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \clubsuit \\ & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit \\ & & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit \\ & & & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit \\ & & & & \heartsuit & \heartsuit & \heartsuit & \heartsuit \end{array} \right], \quad \left[\begin{array}{cccccccc} \heartsuit & \heartsuit & \heartsuit & & & & & \\ \heartsuit & \heartsuit & \heartsuit & & & & & \\ \heartsuit & \heartsuit & \heartsuit & & & & & \\ \heartsuit & \heartsuit & \heartsuit & \clubsuit & & & & \\ & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit \\ & & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit \\ & & & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit \\ & & & & \heartsuit & \heartsuit & \heartsuit & \heartsuit \\ & & & & & \heartsuit & \heartsuit & \heartsuit \end{array} \right].$$

Recall that clubs (\clubsuit) stand for nonzero entries, and hearts (\heartsuit) stand for general entries which can also be zero. Since 1-wedge-shaped matrices are symmetric tridiagonal with nonzero subdiagonal entries, the wedge-shaped matrices can be seen as a generalization of Jacobi matrices.

Jacobi matrices have simple eigenvalues; see, e.g., [16, Lemma 7.7.1]. In the text below it is shown that multiplicities of eigenvalues of a ρ -wedge-shaped matrix are bounded by ρ . The following example of a 2-wedge-shaped matrix

$$\begin{bmatrix} 0 & 0 & 1 & & & & & \\ 0 & 0 & 0 & \ddots & & & & \\ 1 & 0 & 0 & \ddots & 1 & & & \\ & \ddots & \ddots & \ddots & 0 & & & \\ & & & & 1 & 0 & 0 & \end{bmatrix} = \begin{bmatrix} 0 & 1 & & & & & & \\ 1 & 0 & 1 & & & & & \\ & 1 & 0 & 1 & & & & \\ & & & 1 & 0 & & & \\ & & & & 1 & 0 & & \end{bmatrix} \otimes I_2 \in \mathbb{R}^{8 \times 8}$$

with eigenvalues

$$\lambda_{1,2} = -\frac{\sqrt{5}+1}{2}, \quad \lambda_{3,4} = -\frac{\sqrt{5}-1}{2}, \quad \lambda_{5,6} = \frac{\sqrt{5}-1}{2}, \quad \lambda_{7,8} = \frac{\sqrt{5}+1}{2}$$

illustrates that the bound is sharp, in the sense that the multiple eigenvalues with the multiplicity ρ can be present. This also shows that the *strict interlacing property* of eigenvalues of Jacobi matrices (see, e.g., [16, section 7.10]) does not hold for wedge-shaped matrices. Eigenvectors of Jacobi matrices have nonzero first and last entries; see, e.g., [16, Theorem 7.9.3]. The following theorem shows how to generalize the property of the nonzero first element to leading subvectors of eigenvectors of wedge-shaped matrices. This immediately gives the bound for the multiplicities of the individual eigenvalues. Subsequently we show how to generalize the property of the nonzero last element to eigenvectors of wedge-shaped matrices.

THEOREM 4.2. *Let $T \in \mathbb{R}^{n \times n}$ be a ρ -wedge-shaped matrix, $1 \leq \rho < n$. Let $\lambda \in \mathbb{R}$, $v = [\nu_1, \dots, \nu_n]^T \in \mathbb{R}^n$ be an eigenpair of T , i.e., $Tv = \lambda v$, $v \neq 0$. Then the leading subvector $[\nu_1, \dots, \nu_\rho]^T \in \mathbb{R}^\rho$ of v is nonzero.*

Proof. Because $h(k)$, $k = \rho + 1, \dots, n$ is nonincreasing, the first nonzero entry $t_{k,f(k)}$ in the k th row is also the last nonzero entry in the $(f(k))$ th column of T . Using the symmetry of T , $t_{f(k),k}$ is the last nonzero entry in the $(f(k))$ th row. Thus the $(f(k))$ th row of $Tv = \lambda v$ can for $k = \rho + 1, \dots, n$ be written as

$$\left(\sum_{\ell=1}^{k-1} t_{f(k),\ell} \nu_\ell \right) + t_{f(k),k} \nu_k = \lambda \nu_{f(k)}. \quad (4.3)$$

Let, by contradiction, $\nu_1 = \dots = \nu_\rho = 0$. Then (4.3) is for $k = \rho + 1$ reduced to

$$t_{f(\rho+1),\rho+1} \nu_{\rho+1} = \lambda \nu_{f(\rho+1)}.$$

Because $h(\rho + 1)$ is positive, $f(\rho + 1) < \rho + 1$, and $\nu_{f(\rho+1)} = 0$. Since $t_{f(\rho+1),\rho+1} \neq 0$, then $\nu_{\rho+1} = 0$. Repeating the argument gives for $k = \rho + 2, \dots, n$, $\nu_{\rho+2} = \dots = \nu_n = 0$, which contradicts $v \neq 0$. \square

This theorem has the following corollary.

COROLLARY 4.3. *Let $T \in \mathbb{R}^{n \times n}$ be a ρ -wedge-shaped matrix, $1 \leq \rho < n$. Let $\lambda \in \mathbb{R}$ be an eigenvalue of T with the multiplicity r . Let $v_\ell = [\nu_{1,\ell}, \dots, \nu_{n,\ell}]^T \in \mathbb{R}^n$,*

$\ell = 1, \dots, r$, be an arbitrary basis of the corresponding eigenspace, i.e., $TV = \lambda V$, where $V \equiv [v_1, \dots, v_r] \in \mathbb{R}^{n \times r}$. Then the leading $\rho \times r$ block of V ,

$$\Omega \equiv \begin{bmatrix} \nu_{1,1} & \cdots & \nu_{1,r} \\ \vdots & \ddots & \vdots \\ \nu_{\rho,1} & \cdots & \nu_{\rho,r} \end{bmatrix} \in \mathbb{R}^{\rho \times r}, \quad (4.4)$$

is of full column rank r .

Proof. Since $Vw = [\omega_1, \dots, \omega_n]^T$ represents for any $w \neq 0 \in \mathbb{R}^r$ an eigenvector of T , by theorem 4.2, $\Omega w = [\omega_1, \dots, \omega_\rho]^T$ is nonzero, which gives the assertion. \square

If $r > \rho$, then there exists a nontrivial linear combination of the columns of V which gives a vector with the first ρ entries zero, i.e., $\Omega \in \mathbb{R}^{\rho \times r}$ can obviously not have full column rank. This gives the bound for the multiplicities of individual eigenvalues:

COROLLARY 4.4. *An eigenvalue of a ρ -wedge-shaped matrix $T \in \mathbb{R}^{n \times n}$, $1 \leq \rho < n$, has multiplicity at most ρ .*

The following theorem generalizes the property of the last nonzero element of eigenvectors of Jacobi matrices to eigenvectors of wedge-shaped matrices. The proof is analogous to the proof of theorem 4.2.

THEOREM 4.5. *Let $T \in \mathbb{R}^{n \times n}$ be a ρ -wedge-shaped matrix, $1 \leq \rho < n$. Let $\lambda \in \mathbb{R}$, $v = [\nu_1, \dots, \nu_n]^T \in \mathbb{R}^n$ be an eigenpair of T , i.e., $Tv = \lambda v$, $v \neq 0$. Denote*

$$\begin{aligned} \{s_1, \dots, s_\rho\} &\equiv \{1, \dots, n\} \setminus \{f(k) : k = \rho + 1, \dots, n\}, \\ s_1 &> s_2 > \dots > s_\rho, \end{aligned}$$

where $f(k)$ is given by (4.2). Then the subvector $[\nu_{s_1}, \dots, \nu_{s_\rho}]^T \in \mathbb{R}^\rho$ of v is nonzero.

Proof. Since $t_{k,f(k)}$ is the first nonzero entry in the k th row of T , the k th row of $Tv = \lambda v$ can for $k = \rho + 1, \dots, n$ be written as

$$t_{k,f(k)} \nu_{f(k)} + \left(\sum_{\ell=f(k)+1}^n t_{k,\ell} \nu_\ell \right) = \lambda \nu_k. \quad (4.5)$$

Let, by contradiction, $\nu_{s_1} = \dots = \nu_{s_\rho} = 0$. Because $h(n)$ is positive, $f(n) < n$, and $\nu_\ell = 0$ for all $\ell > f(n)$, in particular, $\nu_n \equiv \nu_{s_\rho} = 0$. Thus (4.5) is for $k = n$ reduced to

$$t_{n,f(n)} \nu_{f(n)} = 0,$$

and $t_{n,f(n)} \neq 0$ gives $\nu_{f(n)} = 0$. Repeating the argument for $k = n - 1, n - 2, \dots$ up to $\rho + 1$ gives $\nu_{f(n-1)} = \nu_{f(n-2)} = \dots = \nu_{f(\rho+1)} = 0$, which contradicts $v \neq 0$. \square

Note that s_1, \dots, s_ρ represent the row (and column) indices where the effective bandwidth of T is reduced by one, and $s_\rho = n$. Both nonzero subvectors of length ρ described by theorems 4.2 and 4.5 can be observed from the pattern of a wedge-shaped matrix. As an illustration, eigenvectors of the following 3-wedge-shaped matrix of the

size 9 have nonzero subvectors $[\nu_1, \nu_2, \nu_3]^T$ and $[\nu_3, \nu_6, \nu_9]^T$:

$$\begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \\ \nu_7 \\ \nu_8 \\ \nu_9 \end{bmatrix} \begin{matrix} \longleftrightarrow \\ \longleftrightarrow \\ \longleftrightarrow \\ \\ \\ \\ \\ \\ \longleftrightarrow \end{matrix} \begin{bmatrix} \heartsuit & \heartsuit & \heartsuit & \clubsuit & & & & & \\ \heartsuit & \heartsuit & \heartsuit & \heartsuit & \clubsuit & & & & \\ \heartsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & & & & \\ \clubsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \clubsuit & & & \\ & \clubsuit & \heartsuit & \heartsuit & \heartsuit & \heartsuit & \clubsuit & & \\ & & & \clubsuit & \heartsuit & \heartsuit & \heartsuit & & \\ & & & & \clubsuit & \heartsuit & \heartsuit & \clubsuit & \\ & & & & & \heartsuit & \heartsuit & \heartsuit & \\ & & & & & & \heartsuit & \heartsuit & \heartsuit \\ & & & & & & & \heartsuit & \heartsuit \end{bmatrix} \begin{matrix} \longleftrightarrow \\ \longleftrightarrow \\ \longleftrightarrow \\ \\ \\ \\ \\ \\ \longleftrightarrow \end{matrix} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \\ \nu_7 \\ \nu_8 \\ \nu_9 \end{bmatrix}.$$

4.2. Singular values and vectors of the band subproblem. We first show the link of the band subproblem to the wedge-shaped matrices. For a positive definite matrix $M = Z^T Z$ with its (upper triangular) Cholesky factor Z it is well known that

$$\text{env}(M) = \text{env}(Z^T + Z), \quad \text{where} \quad \text{env}(T) = \{t_{k,j} : f(k) \leq j < k\} \quad (4.6)$$

denotes the so-called *envelope* of a symmetric matrix T , and $f(k)$ is given by (4.2); see, e.g., [8, section 4.2]. Analogously, using the structure of the band subproblem (3.20)–(3.21), it is reasonable to expect that the symmetric positive semidefinite matrices $A_{11}A_{11}^T$ and $[B_1|A_{11}]^T[B_1|A_{11}]$ inherit the band structure and represent wedge-shaped matrices. However, their full row rank upper triangular factors A_{11}^T and $[B_1|A_{11}]$, respectively, do not represent the Cholesky factors, in general. Thus the above mentioned result cannot be used directly; see also an example in Figure 4.1. Therefore we state and prove the following lemma.

LEMMA 4.6. *Let $A_{11}X_1 \approx B_1$, $A_{11} \in \mathbb{R}^{\overline{m} \times \overline{n}}$, $B_1 \in \mathbb{R}^{\overline{m} \times \overline{d}}$ be the band subproblem (3.20)–(3.21). If $\overline{m} > \overline{d}$, then the matrix*

$$A_{11}A_{11}^T \in \mathbb{R}^{\overline{m} \times \overline{m}},$$

is \overline{d} -wedge-shaped. Since $\overline{d} + \overline{n} > \overline{d}$, then the matrix

$$[B_1|A_{11}]^T[B_1|A_{11}] \in \mathbb{R}^{(\overline{d}+\overline{n}) \times (\overline{d}+\overline{n})},$$

is also \overline{d} -wedge-shaped.

Proof. Denote $a_k^T \equiv e_k^T A_{11}$ the k th row of A_{11} , thus $a_k^T a_j$ represents the (k, j) th entry of $A_{11}A_{11}^T$. We look for the first nonzero entry in the k th row of $A_{11}A_{11}^T$, $k = \overline{d} + 1, \dots, \overline{m}$. Denote $\varphi(j) \in \{1, \dots, \overline{m}\}$ the row index of the first nonzero entry in the j th column of A_{11} , i.e.,

$$a_{\varphi(j)}^T = [\underbrace{\heartsuit, \dots, \heartsuit}_{j-1}, \alpha_{\varphi(j), \overline{d}+j}, \underbrace{0, \dots, 0}_{\overline{n}-j}], \quad j = 1, \dots, \overline{n}. \quad (4.7)$$

Denote $\psi(k) \in \{1, \dots, \overline{n}\}$ the column index of the first nonzero entry in the k th row of A_{11} , i.e.,

$$a_k^T = [0, \dots, 0, \underbrace{\gamma_{k, \overline{d}+\psi(k)}}_{\psi(k)-1}, \underbrace{\heartsuit, \dots, \heartsuit}_{\overline{n}-\psi(k)}], \quad k = \overline{d} + 1, \dots, \overline{m}. \quad (4.8)$$

For $j = \psi(k)$ the entries $\alpha_{\varphi(j), \overline{d}+j}$ and $\gamma_{k, \overline{d}+\psi(k)}$ belong to the same column, i.e.,

$$a_k^T a_{\varphi(\psi(k))} = \gamma_{k, \overline{d}+\psi(k)} \alpha_{\varphi(\psi(k)), \overline{d}+\psi(k)} > 0. \quad (4.9)$$

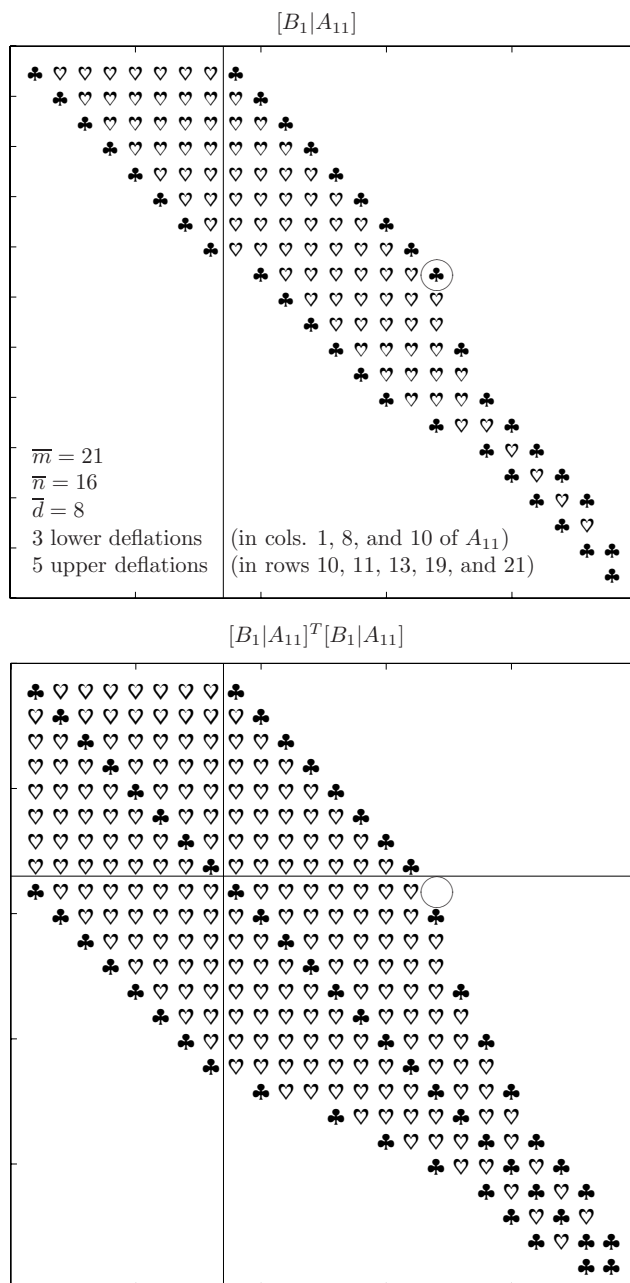


FIG. 4.1. Top: A band problem $Z \equiv [B_1|A_{11}]$ with $\overline{d} = 8$. Bottom: One can see that the positive semidefinite matrix $M \equiv [B_1|A_{11}]^T[B_1|A_{11}]$ is a 8-wedge-shaped matrix. Since the upper triangular matrix Z in the top does not represent the Cholesky factor of the matrix M in the bottom, then $\text{env}(M) \neq \text{env}(Z^T + Z)$; see, e.g., the encircled nonzero entry in the top part and the encircled zero entry in the bottom part.

Using the lower echelon form of A_{11} , all rows above $a_{\varphi(\psi(k))}^T$ ((4.7) with $j = \psi(k)$) are *structurally orthogonal* to a_k^T (4.8), thus all entries to the left of $a_k^T a_{\varphi(\psi(k))}$ (4.9) are zero. Consequently, $a_k^T a_{\varphi(\psi(k))}$ (4.9) is the *first nonzero entry* in the k th row of $A_{11}A_{11}^T$, $k = \bar{d} + 1, \dots, \bar{m}$. Thus

$$f(k) = \varphi(\psi(k)),$$

see (4.2). The matrix A_{11} has the band form with the α entries located on the top and the γ entries on the bottom of the band; see section 3.2.3 above. The row $a_{\varphi(\psi(k))}^T$ ((4.7) with $j = \psi(k)$) is always placed above the row (4.8). Thus (4.9) is on the left of the diagonal entry $a_k^T a_k$ in the k th row of $A_{11}A_{11}^T$, i.e., $\varphi(\psi(k)) < k$ and

$$h(k) = k - \varphi(\psi(k))$$

is positive for $k = \bar{d} + 1, \dots, \bar{m}$. Because both $\varphi(j)$ and $\psi(k)$ are increasing, the composed function $\varphi(\psi(k))$ is also increasing, and $h(k)$ is nonincreasing for $k = \bar{d} + 1, \dots, \bar{m}$. Consequently, $A_{11}A_{11}^T$ is a \bar{d} -wedge-shaped-matrix.

The proof for $[B_1|A_{11}]^T[B_1|A_{11}]$ is analogous: Replace A_{11} by $[B_1|A_{11}]^T$ and exchange the roles of the α and γ entries. \square

Recall that $\bar{m} \geq \max\{\bar{n}, \bar{d}\}$; see (4.1). If $\bar{m} = \bar{d}$, then $A_{11}A_{11}^T$ has no particular structure. Note that $\bar{d} + \bar{n} = \bar{d}$ (i.e., A_{11} has no columns) occurs in the excluded case $A^T B = 0$ (after the QR decomposition (3.1)–(3.2), the band algorithm starts with \bar{d} upper deflations). The matrix $A_{11}^T A_{11} \in \mathbb{R}^{\bar{n} \times \bar{n}}$ is the trailing principal block of the \bar{d} -wedge-shaped matrix $[B_1|A_{11}]^T[B_1|A_{11}]$. If $\bar{n} > \bar{d}$, then $A_{11}^T A_{11}$ represents a \bar{d} -wedge-shaped matrix; see also Figure 4.1. If $\bar{n} \leq \bar{d}$, then $A_{11}^T A_{11}$ has no particular structure.

Now we are ready to formulate the properties of the band subproblem (3.20)–(3.21).

COROLLARY 4.7. *Let $A_{11}X_1 \approx B_1$, $A_{11} \in \mathbb{R}^{\bar{m} \times \bar{n}}$, $B_1 \in \mathbb{R}^{\bar{m} \times \bar{d}}$ be the band subproblem (3.20)–(3.21), i.e., $[B_1|A_{11}]$ and A_{11} are the upper and lower triangular band matrices, respectively, with (at most) $\bar{d} + 1$ nonzero diagonals. Then:*

- (a) *The singular values of $[B_1|A_{11}]$ and A_{11} have multiplicities at most \bar{d} .*
- (b) *Let v_1, \dots, v_r be an orthonormal basis of the right singular vector subspace of $[B_1|A_{11}]$ corresponding to a singular value with the multiplicity r (or of the null-space $\mathcal{N}([B_1|A_{11}])$ with the dimension r). Then the leading $\bar{d} \times r$ block of $[v_1, \dots, v_r] \in \mathbb{R}^{(\bar{m} + \bar{d}) \times r}$ is of full column rank r .*
- (c) *Let u_1, \dots, u_r be an orthonormal basis of the left singular vector subspace of A_{11} corresponding to a singular value with the multiplicity r (or of the null-space $\mathcal{N}(A_{11}^T)$ with the dimension r). Then the matrix*

$$\Phi \equiv [u_1, \dots, u_r]^T B_1 \in \mathbb{R}^{r \times \bar{d}}$$

is of full row rank r , i.e., the band subproblem $A_{11}X_1 \approx B_1$ satisfies the condition (CP3); see section 2.1.

Proof. Assertion (a) follows directly from lemma 4.6 and corollary 4.4, except for the case of A_{11} with $\bar{m} = \bar{d}$. Since $A_{11} \in \mathbb{R}^{\bar{m} \times \bar{n}}$ is of full column rank, $\bar{m} \geq \bar{n}$, the assertion becomes in this case trivial. Assertion (b) follows directly from lemma 4.6 and corollary 4.3. Assertion (c): The leading block $\Omega \in \mathbb{R}^{\bar{d} \times r}$ of $[u_1, \dots, u_r] \in \mathbb{R}^{\bar{m} \times r}$, is of full column rank r by corollary 4.3 (the case $\bar{m} = \bar{d}$ excluded in lemma 4.6 becomes again trivial) and $B_1 = [F_1^T, 0]^T$, where $F_1 \in \mathbb{R}^{\bar{d} \times \bar{d}}$ is nonsingular; see (3.1). Thus Φ is of full row rank r . \square

5. Concluding remarks. We have proved that the band algorithm computes the problem $A_{11}X_1 \approx B_1$ satisfying conditions (CP1)–(CP3) formulated in section 2.1. We state this result as the following theorem.

THEOREM 5.1. *The band subproblem $A_{11}X_1 \approx B_1$ (3.20)–(3.21) obtained as the output of the band algorithm described in section 3 applied on the problem $AX \approx B$ represents a core problem within $AX \approx B$ in the sense of definition 2.1. It can therefore be called the core problem in the band form.*

Assertions (a) and (b) of corollary 4.7 give some additional properties of core problems. For $\bar{d} = 1$, these properties reduce to the well known facts that singular values of $[b_1|A_{11}]$ are simple and the right singular vectors have nonzero first components, guaranteeing existence of the unique TLS solution of the core problem. The properties from corollary 4.7 might be helpful in analysis of solvability of core problems for $\bar{d} > 1$. This issue is, however, still under investigation.

Acknowledgments. The authors wish to thank Diana M. Sima for valuable discussions about the TLS problems and Miroslav Tůma for his help and discussions about envelopes of band matrices.

REFERENCES

- [1] J. L. BARLOW, *Reorthogonalization for the Golub–Kahan–Lanczos bidiagonal reduction*, Numer. Math., 124 (2013), pp. 237–278.
- [2] Å. BJÖRCK, *Bidiagonal Decomposition and Least Squares*, Presentation, Canberra, Australia, 2005.
- [3] Å. BJÖRCK, *A Band-Lanczos Generalization of Bidiagonal Decomposition*, Presentation, Conference in Honor of G. Dahlquist, Stockholm, Sweden, 2006.
- [4] Å. BJÖRCK, *A band-Lanczos algorithm for least squares and total least squares problems*, in Book of Abstracts of 4th Total Least Squares and Errors-in-Variables Modeling Workshop, Leuven, Katholieke Universiteit Leuven, Leuven, Belgium, 2006, pp. 22–23.
- [5] Å. BJÖRCK, *Block bidiagonal decomposition and least squares problems with multiple right-hand sides*, unpublished manuscript.
- [6] B. BOHNHORST, *Beiträge zur numerischen Behandlung des unitären Eigenwertproblems*, Ph.D. thesis, Universität Bielefeld, Bielefeld, Germany, 1993.
- [7] W. GAUTSCHI, *Orthogonal Polynomials, Computation and Approximation*, Oxford University Press, New York, 2004.
- [8] A. GEORGE AND J. W. H. LIU, *Computer Solution of Large Sparse Positive Definite Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1981.
- [9] G. H. GOLUB AND W. KAHAN, *Calculating the singular values and pseudo-inverse of a matrix*, SIAM J. Numer. Anal., Ser. B, 2 (1965), pp. 205–224.
- [10] I. HNĚTYNKOVÁ, M. PLEŠINGER, D. M. SIMA, Z. STRAKOŠ, AND S. VAN HUFFEL, *The total least squares problem in $AX \approx B$: A new classification with the relationship to the classical works*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 748–770.
- [11] I. HNĚTYNKOVÁ, M. PLEŠINGER, AND Z. STRAKOŠ, *Lanczos tridiagonalization, Golub–Kahan bidiagonalization and core problem*, Proc. Appl. Math. Mech., 6 (2006), pp. 717–718.
- [12] I. HNĚTYNKOVÁ, M. PLEŠINGER, AND Z. STRAKOŠ, *The core problem within a linear approximation problem $AX \approx B$ with multiple right-hand sides*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 917–931.
- [13] I. HNĚTYNKOVÁ AND Z. STRAKOŠ, *Lanczos tridiagonalization and core problems*, Linear Algebra Appl., 421 (2007), pp. 243–251.
- [14] J. LIESEN AND Z. STRAKOŠ, *Krylov Subspace Methods, Principles and Analysis*, Oxford University Press, New York, 2012.
- [15] C. C. PAIGE AND Z. STRAKOŠ, *Core problem in linear algebraic systems*, SIAM J. Matrix Anal. Appl., 27 (2006), pp. 861–875.
- [16] B. N. PARLETT, *The Symmetric Eigenvalue Problem*, SIAM, Philadelphia, 1998.
- [17] M. PLEŠINGER, *The total least squares problem and reduction of data in $AX \approx B$* , Ph.D. thesis, Technical University of Liberec, Liberec, Czech Republic, 2008.

- [18] D. M. SIMA, *Regularization techniques in model fitting and parameter estimation*, Ph.D. thesis, Katholieke Universiteit Leuven, Leuven, Belgium, 2006.
- [19] D. M. SIMA AND S. VAN HUFFEL, *Core problems in $AX \approx B$* , Technical Report, Dept. of Electrical Engineering, Katholieke Universiteit Leuven (2006).
- [20] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Oxford England, Clarendon Press, 1965 (reprint 2004).

Appendix A. Implementation of the band algorithm.

Algorithm 1 implements the band algorithm. Assuming exact arithmetic, it returns for general input data A, B the matrices A_{11}, B_1 of the core problem in the band form, and the corresponding transformation matrices $P_1^{\text{cp}}, Q_1^{\text{cp}}$ (see (3.21)), and R (see (2.5)). For alternative implementations see [18, Algorithm 2.4, p. 38] or [17, Algorithm 5.1, p. 74]. In the algorithm p_k, q_j denote the k th and j th column of P and Q , and $P_k \equiv [p_1, \dots, p_k] \in \mathbb{R}^{m \times k}$, $Q_j \equiv [q_1, \dots, q_j] \in \mathbb{R}^{n \times j}$ (Q_0 represents a matrix with no columns); $L_{k,j} \equiv P_k^T A Q_j$ denotes the $k \times j$ leading principal block of L , in particular $L_{\bar{m}, \bar{n}} = A_{11}$ (see (3.21)), and $l_{k,j} \equiv e_k^T L e_j$ is the (k, j) th entry of L . The variables c_U and c_L are counters of the upper and lower deflations, respectively. The algorithm stops when $c_U + c_L = \bar{d} = \text{rank}(B)$; see the line 7. The indices j and k denote the number of columns and rows, respectively, of the currently computed part of the matrix L . If j or k becomes equal to n or m , respectively, then the algorithm stays in the loop of lines {7–13, 36–39, 7, etc.} or {7–26, 32–35, 38–39, 7, etc.}. The value of c_U respectively c_L increases until $c_U + c_L = \bar{d}$. The following schema illustrates (on the example given below (3.20)) how the algorithm assembles the matrix A_{11} ; the arrows represent updates in the lines 24 or 31:

$$\begin{aligned}
 L_{\bar{d},0} &= \left[\begin{array}{c} \\ \\ \\ \end{array} \right] \xrightarrow{24} \left[\begin{array}{c} \alpha_{1,4} \\ \beta_{2,4} \\ \beta_{3,4} \end{array} \right] \xrightarrow{31} \left[\begin{array}{c} \alpha_{1,4} \\ \beta_{2,4} \\ \beta_{3,4} \\ \hline \gamma_{4,4} \end{array} \right] \xrightarrow{24} \left[\begin{array}{c|c} \alpha_{1,4} & 0 \\ \beta_{2,4} & \alpha_{2,5} \\ \beta_{3,4} & \beta_{3,5} \\ \hline \gamma_{4,4} & \beta_{4,5} \end{array} \right] \xrightarrow{24} \left[\begin{array}{c|c|c} \alpha_{1,4} & 0 & 0 \\ \beta_{2,4} & \alpha_{2,5} & 0 \\ \beta_{3,4} & \beta_{3,5} & \alpha_{3,6} \\ \hline \gamma_{4,4} & \beta_{4,5} & \beta_{4,6} \\ \hline 0 & 0 & \gamma_{5,6} \end{array} \right] \\
 &\xrightarrow{31} \left[\begin{array}{c|c|c|c} \alpha_{1,4} & 0 & 0 & 0 \\ \beta_{2,4} & \alpha_{2,5} & 0 & 0 \\ \beta_{3,4} & \beta_{3,5} & \alpha_{3,6} & 0 \\ \hline \gamma_{4,4} & \beta_{4,5} & \beta_{4,6} & \alpha_{4,7} \\ \hline 0 & 0 & \gamma_{5,6} & \beta_{5,7} \end{array} \right] \xrightarrow{24} \left[\begin{array}{c|c|c|c|c} \alpha_{1,4} & 0 & 0 & 0 & 0 \\ \beta_{2,4} & \alpha_{2,5} & 0 & 0 & 0 \\ \beta_{3,4} & \beta_{3,5} & \alpha_{3,6} & 0 & 0 \\ \hline \gamma_{4,4} & \beta_{4,5} & \beta_{4,6} & \alpha_{4,7} & 0 \\ \hline 0 & 0 & \gamma_{5,6} & \beta_{5,7} & \alpha_{5,8} \\ \hline 0 & 0 & 0 & \gamma_{6,7} & \beta_{6,8} \\ \hline 0 & 0 & 0 & 0 & \gamma_{7,8} & \alpha_{7,9} \\ \hline 0 & 0 & 0 & 0 & 0 & \gamma_{8,9} & \alpha_{8,10} \end{array} \right] = L_{\bar{m}, \bar{n}} = A_{11}.
 \end{aligned}$$

The sums in (3.12), (3.14), or in (3.16), (3.18), are implemented in the lines 11 and 25, respectively. This implementation does not reflect, for simplicity, the structure of zero entries in the band matrix; i.e., as an example, the sum in the line 11 computes the full matrix-vector product of the matrix $[q_1, \dots, q_{j-1}]$ with the last column of $L_{k,j-1}^T$.

Algorithm 1 (Band generalization of the Golub–Kahan bidiagonalization).

```

1: input  $A \in \mathbb{R}^{m \times n}$ ,  $B \in \mathbb{R}^{m \times d}$  { $A^T B \neq 0$ ,  $\text{rank}(B) = \bar{d}$ }
2: compute  $B = [C, 0]R^T$  {RHS preprocessing,  $C \in \mathbb{R}^{m \times \bar{d}}$ ,  $R^T = R^{-1}$ }
3: compute  $C = P_{\bar{d}}F_1$  {QR decomposition in the economic form}
4: initialize  $Q_0 \leftarrow []$ ,  $L_{\bar{d},0} \leftarrow []$  {data arrays/matrices}
5: initialize  $c_U \leftarrow 0$ ,  $c_L \leftarrow 0$  {deflation counters}
6: initialize  $j \leftarrow 1$ ,  $k \leftarrow \bar{d}$  {control variables/indices}
7: while  $c_U + c_L < \bar{d}$  do
8:   if  $j = 1$ , then {compute an auxiliary vector}
9:      $\text{aux}_q \leftarrow A^T p_{j+c_U} = A^T p_1$ 
10:  else
11:     $\text{aux}_q \leftarrow A^T p_{j+c_U} - \sum_{i=1}^{j-1} q_i l_{j+c_U,i}$ 
12:  end
13:  if  $\text{aux}_q \neq 0$ , then
14:     $\alpha_{j+c_U, \bar{d}+j} \leftarrow \|\text{aux}_q\|$  {compute  $\alpha$  coefficient}
15:     $q_j \leftarrow \text{aux}_q / \alpha_{j+c_U, \bar{d}+j}$  {compute  $q$  vector}
16:     $Q_j \leftarrow [Q_{j-1}, q_j]$  {update of  $Q$  matrix}
17:     $\text{beta} \leftarrow []$ 
18:    if  $c_U + c_L < \bar{d} - 2$ , then
19:      for  $i = j + 1 + c_U, \dots, j + \bar{d} - 1 - c_L$  do {compute  $\beta$  coefficients}
20:         $\beta_{i, \bar{d}+j} \leftarrow p_i^T A q_j$ 
21:         $\text{beta} \leftarrow [\text{beta}, \beta_{i, \bar{d}+j}]$ 
22:      end
23:    end
24:     $L_{k,j} \leftarrow [L_{k,j-1}, [0_{1,j-1+c_U}, \alpha_{j+c_U, \bar{d}+j}, \text{beta}]^T]$  {update of  $L$  (add a col.)}
25:     $\text{aux}_p \leftarrow A q_j - \sum_{i=1}^k p_i l_{i,j}$  {compute an auxiliary vector}
26:    if  $\text{aux}_p \neq 0$ , then
27:       $k \leftarrow k + 1$ 
28:       $\gamma_{k, \bar{d}+j} \leftarrow \|\text{aux}_p\|$  {compute  $\gamma$  coefficient}
29:       $p_k \leftarrow \text{aux}_p / \gamma_{k, \bar{d}+j}$  {compute  $p$  vector}
30:       $P_k \leftarrow [P_{k-1}, p_k]$  {update of  $P$  matrix}
31:       $L_{k,j} \leftarrow [L_{k-1,j}^T, [0_{1,j-1}, \gamma_{k, \bar{d}+j}]^T]^T$  {update of  $L$  matrix (add a row)}
32:    else
33:       $c_L \leftarrow c_L + 1$  {lower deflation}
34:    end
35:     $j \leftarrow j + 1$ 
36:  else
37:     $c_U \leftarrow c_U + 1$  {upper deflation}
38:  end
39: end
40:  $\bar{m} \leftarrow k$ ,  $\bar{n} \leftarrow j - 1$ ,  $B_1 \leftarrow [F_1^T, 0_{\bar{d}, \bar{m} - \bar{d}}]^T$ ,  $A_{11} \leftarrow L_{\bar{m}, \bar{n}}$ ,  $P_1^{\text{cp}} \leftarrow P_{\bar{m}}$ ,  $Q_1^{\text{cp}} \leftarrow Q_{\bar{n}}$ 
41: output  $A_{11} \in \mathbb{R}^{\bar{m} \times \bar{n}}$ ,  $B_1 \in \mathbb{R}^{\bar{m} \times \bar{d}}$ ,  $P_1^{\text{cp}} \in \mathbb{R}^{m \times \bar{m}}$ ,  $Q_1^{\text{cp}} \in \mathbb{R}^{n \times \bar{n}}$ ,  $R \in \mathbb{R}^{d \times d}$ 

```
