Sparse block factorization of saddle point matrices

Document status and date:
Published: 01/01/2014

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Download date: 04. Oct. 2023

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computer Science

CASA-Report 14-23
August 2014

Sparse block factorization of saddle point matrices

by

S. Lungten, W.H.A. Schilders, J.M.L. Maubach

SPARSE BLOCK FACTORIZATION OF SADDLE POINT MATRICES

S. Lungten, W.H.A. Schilders, J.M.L. Maubach

April 20, 2014

Abstract

The factorization method presented in this paper takes advantage of the special structures and properties of saddle point matrices. A variant of Gaussian elimination equivalent to the Cholesky's factorization is suggested and implemented for factorizing the saddle point matrices block-wise with small blocks of order 1 and 2. The Gaussian elimination applied to these small blocks on block level also induces a block 3×3 structured factorization of which the blocks have special properties. We compare the new block factorization with the Schilders' factorization in terms of the sparsity of their factors and computational efficiency. The factorization can be used as a direct method, and also anticipate for preconditioning techniques.

1 Introduction

Indefinite matrices with special forms which occur in many scientific and engineering problems can be exploited efficiently by taking advantage of the structures and properties of their blocks. We consider symmetric indefinite linear systems of the form (see footnote [1] for the notations)

$$\underbrace{\begin{bmatrix} \mathring{\boldsymbol{A}} & \mathring{\boldsymbol{B}}^T \\ \mathring{\boldsymbol{B}} & \boldsymbol{0} \end{bmatrix}}_{\mathring{\boldsymbol{\mathcal{A}}}} \underbrace{\begin{bmatrix} \mathring{\boldsymbol{x}} \\ \mathring{\boldsymbol{y}} \end{bmatrix}}_{\mathring{\boldsymbol{u}}} = \underbrace{\begin{bmatrix} \mathring{\boldsymbol{f}} \\ \mathring{\boldsymbol{g}} \end{bmatrix}}_{\mathring{\boldsymbol{d}}} \tag{1}$$

where $\mathring{\boldsymbol{A}} \in \mathbb{R}^{n \times n}$ is symmetric, positive definite; $\mathring{\boldsymbol{B}} \in \mathbb{R}^{m \times n}$ has full rank and $m \leq n$; $\mathring{\boldsymbol{x}}, \mathring{\boldsymbol{f}} \in \mathbb{R}^n$; and $\mathring{\boldsymbol{y}}, \mathring{\boldsymbol{g}} \in \mathbb{R}^m$. In applications, the coefficient matrix $\mathring{\boldsymbol{\mathcal{A}}}$ is usually sparse and large, which can easily turn out to be a million by million. Systems of the form (1) are known as saddle point problems, which are resulted from discretization of PDEs or coupled PDEs such as the Stokes and mixed finite element methods. Saddle point systems also arise in electronic circuit simulations [30, 35], Maxwell's equations [26],

---

[1] The original system (1) undergoes a transformation, so we use the symbol '∘' on its notations in order to represent the transformed system more conveniently without the symbol '∘'.

economic models and constrained optimization problems [1, 10, 15, 16, 18, 33]. For example consider the equality-constrained quadratic programming problem:

$$\min_{\mathring{\boldsymbol{x}}} f(\mathring{\boldsymbol{x}}) = \frac{1}{2}\mathring{\boldsymbol{x}}^T\mathring{\boldsymbol{A}}\mathring{\boldsymbol{x}} - \mathring{\boldsymbol{x}}^T\mathring{\boldsymbol{f}} \quad \text{subject to} \quad \mathring{\boldsymbol{B}}\mathring{\boldsymbol{x}} = \mathring{\boldsymbol{g}}. \tag{2}$$

The Karush-Kuhn-Tucker (KKT) conditions [15, 39] for the solution to (2) give rise to the system (1), where the components of $\mathring{\boldsymbol{y}}$ are the associated Lagrange multipliers. Thus the coefficient matrix $\mathring{\mathcal{A}}$ is also known as KKT matrix and it is nonsingular if (i) $\mathring{\boldsymbol{B}}$ has full row rank and (ii) the reduced Hessian matrix, $\mathring{\boldsymbol{Z}}^T\mathring{\boldsymbol{A}}\mathring{\boldsymbol{Z}}$ is positive definite, where $\mathring{\boldsymbol{Z}} \in \mathbb{R}^{n\times(n-m)}$ is the matrix whose columns span the $\ker(\mathring{\boldsymbol{B}})$ [29, p.443].

Numerous solution methods for the saddle point systems of the form (1) can be found in the literature and many of them have focused on preconditioning techniques for Krylov subspace iterative solvers [1, 2, 3, 7, 12, 18, 23, 24, 26, 28, 31]. As a direct method against iterative solvers, various techniques on symmetric indefinite factorization $\boldsymbol{P}^T\mathring{\mathcal{A}}\boldsymbol{P} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T$ can be found in [9, 14, 21, 34, 35, 38], where $\boldsymbol{P}$ is a permutation matrix, $\boldsymbol{L}$ is unit lower triangular matrix, $\boldsymbol{D}$ is block-diagonal matrix with blocks of order 1 or 2. The permutation matrix $\boldsymbol{P}$ is introduced for (i) *pivoting dynamically* and (ii) *reducing the fill-ins* in $\boldsymbol{L}$ if $\mathring{\mathcal{A}}$ is sparse. The block diagonal pivoting strategies are mainly due to Bunch-Kaufman [4], Bunch-Parlett [5] and Bunch-Kaufman-Parlett (BKP) [6].

In this paper, we propose a different transformation $\boldsymbol{\mathcal{T}}^T\mathring{\mathcal{A}}\boldsymbol{\mathcal{T}} = \mathcal{A}$, followed by a block Gaussian elimination factorization $\boldsymbol{P_\pi}^T\mathcal{A}\boldsymbol{P_\pi} = \boldsymbol{L_b}\boldsymbol{D_b}^{-1}\boldsymbol{L_b}^T$, where:

(i) $\boldsymbol{L_b}$ is a block lower triangular with blocks of order 1 and 2, and $\boldsymbol{D_b} = \text{diag}(\boldsymbol{L_b})$ is the block diagonal part of $\boldsymbol{L_b}$ with blocks of order 1 and 2.

(ii) $\boldsymbol{\mathcal{T}}$ is an $(n+m)\times(n+m)$ transformation (possibly a permutation) matrix, which follows from a transformation of the linear constraint matrix $\boldsymbol{B}$. Operator $\boldsymbol{\mathcal{T}}$ is chosen such that the $\boldsymbol{L_b}\boldsymbol{D_b}^{-1}\boldsymbol{L_b}^T$ factorization is *stable* and has a *sparse* factor $\boldsymbol{L_b}$.

(iii) $\boldsymbol{P_\pi}$ is a simple, predefined $(n+m)\times(n+m)$ permutation matrix for *a priori pivoting* of $\mathcal{A}$.

The proposed $\boldsymbol{L_b}\boldsymbol{D_b}^{-1}\boldsymbol{L_b}^T$ factorization method exploits the structure and properties of $\mathcal{A}$. For instance, the first m blocks of the block diagonal $\boldsymbol{D_b}$ are the 2×2 pivots inheriting the same structure and properties of $\mathcal{A}$, and the remaining $n - m$ blocks are the 1×1 pivots. At scalar level, $\boldsymbol{L_b}\boldsymbol{D_b}^{-1}\boldsymbol{L_b}^T$ factorization has the same computational efficiency as that of the Cholesky's factorization for symmetric, positive definite matrices, which is shown in the appendix of this paper. Whenever we come across the features related to the scalar level factorization, we refer them to the appendix.

There are also several other block factorization methods for $\mathring{\mathcal{A}}$ with larger blocks of order n, m or $n - m$, which are mostly based on either the Schur complement matrix

2

$\mathring{\boldsymbol{B}}\mathring{\boldsymbol{A}}^{-1}\mathring{\boldsymbol{B}}^T$ or the reduced Hessian matrix $\mathring{\boldsymbol{Z}}^T\mathring{\boldsymbol{A}}\mathring{\boldsymbol{Z}}$. For example, the Schilders' factorization [35] is a block 3×3 structured factorization with blocks of order m and $n-m$, applied to $\boldsymbol{\mathcal{T}}^T\mathring{\boldsymbol{A}}\boldsymbol{\mathcal{T}}$ for a different $\boldsymbol{\mathcal{T}}$. Later in [10, 11, 13], it has been used as a basis for implicit factorization for constructing different families of constraint preconditioners for the saddle point matrices. We also produce such a 3×3 structured block factorization from $\boldsymbol{L_b}\boldsymbol{D_b}^{-1}\boldsymbol{L_b}^T$ factorization, and it is different from the ones in [2, 10, 11, 35].

The remaining Sections of the paper are organized as follows. In Section 2, we discuss the required properties for $\boldsymbol{\mathcal{T}}$ and show that such matrices exist for symmetric saddle point problems. Section 3 is the main part of the paper, in which we present the proposed factorization $\boldsymbol{P_\pi}^T\boldsymbol{\mathcal{A}}\boldsymbol{P_\pi} = \boldsymbol{L_b}\boldsymbol{D_b}^{-1}\boldsymbol{L_b}^T$. It covers the existence, sparsity and stability of $\boldsymbol{L_b}$, and the steps for solving (1) using $\boldsymbol{L_b}$. Comparison of the new block factorization with the Schilders' block factorization is discussed in Section 4. Numerical results for this comparison are provided in Section 5.

2 Determination of transformation matrix $\boldsymbol{\mathcal{T}}$

There are different ways to choose the transformation matrix $\boldsymbol{\mathcal{T}}$ depending on the requirement of the transformed matrix $\boldsymbol{B}$. For example, in [17, 35], $\boldsymbol{\mathcal{T}}$ is chosen such that it results $\boldsymbol{B}_1$ to be an $m \times m$ upper triangular matrix [◥]. With our aim to obtain a stable and sparse block $\boldsymbol{L_b}\boldsymbol{D_b}^{-1}\boldsymbol{L_b}^T$ factorization, we choose a transformation matrix

$$\boldsymbol{\mathcal{T}} = \begin{bmatrix} \boldsymbol{P} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{bmatrix}, \tag{3}$$

where $\boldsymbol{P}$ is an $n \times n$ permutation matrix and $\boldsymbol{Q}$ is an $m \times m$ orthogonal (possibly a permutation) matrix such that

$$\boldsymbol{Q}^T\mathring{\boldsymbol{B}}\boldsymbol{P} = \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 \end{bmatrix} = \begin{bmatrix} ◣ & ▆ \end{bmatrix}, \tag{4}$$

a *lower trapezoidal* form with $\boldsymbol{B}_1$ being an $m \times m$ nonsingular lower triangular [◣] matrix and $\boldsymbol{B}_2$ the remaining $m \times (n-m)$ part [▆]. The choice of $\boldsymbol{\mathcal{T}}$ here, also ensures that $\mathring{\boldsymbol{A}}$ is just permuted due to $\boldsymbol{P}$, which is an essential property for sparsity of the transformed saddle point matrix. In the following, we give a brief overview to determine $\boldsymbol{\mathcal{T}}$ satisfying these conditions.

Typical $\mathring{\boldsymbol{B}}$. If $\mathring{\boldsymbol{B}} = [\mathring{b}_{ij}]$ is an incidence matrix with $\mathring{b}_{ij} \in \{-1, 0, 1\}$, which has maximally two non-zeros (-1 and/or 1) in each column [30, 35], then one can obtain an $m \times m$ row permutation matrix $\boldsymbol{P_r}$ and an $n \times n$ column permutation matrix $\boldsymbol{P_c}$ such that $\boldsymbol{P_r}^T\mathring{\boldsymbol{B}}\boldsymbol{P_c} = \boldsymbol{B}$ is a lower trapezoidal form. Thus $\boldsymbol{\mathcal{T}}$ in (3) is an $(n+m) \times (n+m)$ permutation matrix with $\boldsymbol{P} = \boldsymbol{P_c}$ and $\boldsymbol{Q} = \boldsymbol{P_r}$. Systems of the form (1) with incidence matrix $\mathring{\boldsymbol{B}}$ evolve in resistor network modeling [30], the Stokes equations [9], and many other applications with network topology.

General $\mathring{B}$. If $\mathring{B}$ is of more general form, then by applying a sparse $QR-$ transformation to it, one can obtain an $m \times m$ orthogonal matrix $\boldsymbol{Q}_1$ and an $n \times n$ permutation matrix $\boldsymbol{P}_1$ such that

$$\boldsymbol{Q}_1^T \mathring{\boldsymbol{B}} \boldsymbol{P}_1 = \tilde{\boldsymbol{B}} = \begin{bmatrix} \tilde{\boldsymbol{B}}_1 & \tilde{\boldsymbol{B}}_2 \end{bmatrix} = \begin{bmatrix} \diagdown & \blacksquare \end{bmatrix}.$$

Define an $n \times n$ permutation matrix

$$\boldsymbol{P}_2 = \begin{bmatrix} \boldsymbol{I}_a & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n-m} \end{bmatrix},$$

where $\boldsymbol{I}_a = [e_m, \ \ldots, \ e_1]$ with e_i being the ith unit vector in $\mathbb{R}^m$. Then

$$\boldsymbol{I}_a \tilde{\boldsymbol{B}} \boldsymbol{P}_2 = \begin{bmatrix} \boldsymbol{I}_a \tilde{\boldsymbol{B}}_1 \boldsymbol{I}_a & \boldsymbol{I}_a \tilde{\boldsymbol{B}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 \end{bmatrix} = \begin{bmatrix} \diagdown & \blacksquare \end{bmatrix}.$$

Hence, the required $\boldsymbol{\mathcal{T}}$ in (3) is determined by choosing an $m \times m$ orthogonal matrix $\boldsymbol{Q} = \boldsymbol{Q}_1 \boldsymbol{I}_a$ and an $n \times n$ permutation matrix $\boldsymbol{P} = \boldsymbol{P}_1 \boldsymbol{P}_2$. More details about sparse QR transformation can be found in [8, p. 82-95].

3   Block $\boldsymbol{L_b D_b}^{-1} \boldsymbol{L_b}^T$ factorization

Although the transformed saddle point matrix $\boldsymbol{\mathcal{A}}$ has special block structures with blocks of order m and $n-m$, we do **not** factorize it by using these blocks directly. This is because more amount of computational work has to be spent on computing the inverses, products and sums of large blocks during the factorization. In addition, this approach has to come through a certain type of conjecture and requires separate algorithms for the block matrix operations. This may also lead to a number of factorizations, which are slightly different from each other. Our aim is to first exploit the structure of $\boldsymbol{\mathcal{A}}$, partition it into small blocks of order 1 and 2, and then do a unique factorization in exact arithmetic using simple, inexpensive and robust algorithm.

3.1   Block partitioning

We consider the transformation matrix $\boldsymbol{\mathcal{T}}$ of size $(n + m) \times (n + m)$ defined in (3), which is applied to the saddle point matrix $\mathring{\boldsymbol{\mathcal{A}}}$ as follows:

$$\underbrace{\begin{bmatrix} \boldsymbol{P}^T & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}^T \end{bmatrix}}_{\boldsymbol{\mathcal{T}}^T} \underbrace{\begin{bmatrix} \mathring{\boldsymbol{A}} & \mathring{\boldsymbol{B}}^T \\ \mathring{\boldsymbol{B}} & \boldsymbol{0} \end{bmatrix}}_{\mathring{\boldsymbol{\mathcal{A}}}} \underbrace{\begin{bmatrix} \boldsymbol{P} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{bmatrix}}_{\boldsymbol{\mathcal{T}}} = \underbrace{\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{\mathcal{A}}}, \tag{5}$$

where $\boldsymbol{B}$ is lower trapezoidal form in (4) and $\boldsymbol{A}$ is permuted form of $\mathring{\boldsymbol{A}}$ due to $\boldsymbol{P}$. Because of the lower trapezoidal form of $\boldsymbol{B}$, the transformed matrix $\boldsymbol{\mathcal{A}}$ can be partitioned

4

into a block 3×3 structure as follows:

$$
\boldsymbol{\mathcal{A}} = \begin{array}{c} m \\ n-m \\ m \end{array} \overset{\begin{array}{ccc} m & n-m & m \end{array}}{\left[\begin{array}{ccc} \boldsymbol{A_{11}} & \boldsymbol{A_{12}} & \boldsymbol{B_1}^T \\ \boldsymbol{A_{21}} & \boldsymbol{A_{22}} & \boldsymbol{B_2}^T \\ \boldsymbol{B_1} & \boldsymbol{B_2} & \boldsymbol{0} \end{array}\right]} = \left[\begin{array}{c|c|c} \blacksquare & \| & \blacksquare \\ \hline \blacksquare & \blacksquare & \blacksquare \\ \hline \blacksquare & \| & \end{array}\right].
$$

Beside $\boldsymbol{B_1}$ being a lower triangular, the blocks $\boldsymbol{A_{11}}$ and $\boldsymbol{A_{22}}$ already have nice properties – symmetric, positive definite and sparse. By taking advantage of the structures and properties of these blocks, we reorder $\boldsymbol{\mathcal{A}}$ by applying a simple permutation matrix $\boldsymbol{P_\pi}$ of size $(n+m) \times (n+m)$ such that the permuted $\boldsymbol{\mathcal{A}}$ can be partitioned into a block $n \times n$ structure with blocks of order 1 and 2. The partitioning gives four types of blocks, namely 2×2, 1×2 and its transpose 2×1, and 1×1 (scalar) blocks, residing in their respective domains such that

$$
n^2 = \underset{(2\times2\ \text{blocks})}{m^2} + \underset{(1\times2\ \&\ 2\times1\ \text{blocks})}{2m(n-m)} + \underset{(1\times1\ \text{blocks})}{(n-m)^2}.
$$

The most significant feature of this partitioned form is that its block-diagonal part is given by the direct sum

$$
\bigoplus_{k=1}^{m} \begin{bmatrix} a_{kk} & b_{kk} \\ b_{kk} & 0 \end{bmatrix} \oplus \bigoplus_{k=2m+1}^{n+m} a_{kk} \ ,
$$

which form a priori pivots for the block $\boldsymbol{L_b D_b}^{-1} \boldsymbol{L_b}^T$ factorization. Furthermore, all the 2×2 blocks of its block lower triangular part are retained in its factor $\boldsymbol{L_b}$, thereby ensuring the sparsity of first $2m$ rows of $\boldsymbol{L_b}$. To be clearer with partitioning and factorization, one needs to be cautious with the indices of the block elements, which are defined according the elements of the transformed matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, i.e., $\boldsymbol{A} = [a_{ij}]$, $1 \le i, j \le n$ and $\boldsymbol{B} = [b_{ij}]$, $1 \le i \le m$, $1 \le j \le n$. The permutation matrix $\boldsymbol{P_\pi}$ that we use here is defined as in [19, 35] as follows:

Definition 1. *Consider the positive integers n and m in* (1). *Let $\mathcal{N}_{n+m} = \{1, 2, \ldots, n+m\}$. Define a permutation $\pi : \mathcal{N}_{n+m} \to \mathcal{N}_{n+m}$ by*

$$
\pi = \begin{pmatrix} 1, & 2, & 3, & 4, & \cdots, & 2m-1, & 2m, & 2m+1, & \cdots, & n+m \\ 1, & n+1, & 2, & n+2, & \cdots, & m, & n+m, & m+1, & \cdots, & n \end{pmatrix}.
$$

An $(n+m) \times (n+m)$ permutation matrix $\boldsymbol{P_\pi}$ related to π is given by

$$
\boldsymbol{P_\pi} = \begin{bmatrix} \boldsymbol{e}_1, & \boldsymbol{e}_{n+1}, & \boldsymbol{e}_2, & \boldsymbol{e}_{n+2}, & \cdots, & \boldsymbol{e}_m, & \boldsymbol{e}_{n+m}, & \boldsymbol{e}_{m+1}, & \cdots, & \boldsymbol{e}_n \end{bmatrix}, \tag{6}
$$

where $\boldsymbol{e}_i$ is the ith unit vector of length $m+n$.

To gain an insight on how the application of $\boldsymbol{P_\pi}$ to $\mathcal{A}$ can partition it into a block $n \times n$ structure, we consider $\mathcal{A}$ with $n = 5$ and $m = 3$:

$$
\mathcal{A} = \begin{array}{c} \\ \\ 1 \\ 2 \\ m \to 3 \\ 4 \\ n \to 5 \\ n+1 \to 6 \\ 7 \\ n+m \to 8 \end{array}
\begin{bmatrix}
\begin{array}{ccccc|ccc}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & b_{11} & b_{21} & b_{31} \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & 0 & b_{22} & b_{32} \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & 0 & 0 & b_{33} \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & b_{14} & b_{24} & b_{34} \\
a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & b_{15} & b_{25} & b_{35} \\
\hline
b_{11} & 0 & 0 & b_{14} & b_{15} & 0 & 0 & 0 \\
b_{21} & b_{22} & 0 & b_{24} & b_{25} & 0 & 0 & 0 \\
b_{31} & b_{32} & b_{33} & b_{34} & b_{35} & 0 & 0 & 0
\end{array}
\end{bmatrix}. \tag{7}
$$

The numbers along the border of $\mathcal{A}$ form the domain of π that gives $\boldsymbol{P_\pi} = [\boldsymbol{e}_1,\ \boldsymbol{e}_6,\ \boldsymbol{e}_2,\ \boldsymbol{e}_7,\ \boldsymbol{e}_3,\ \boldsymbol{e}_8,\ \boldsymbol{e}_4,\ \boldsymbol{e}_5]$. Applying $\boldsymbol{P_\pi}$ symmetrically to $\mathcal{A}$, we obtain a block 5×5 structure:

$$
\boldsymbol{P_\pi}^T \mathcal{A} \boldsymbol{P_\pi} =
\begin{array}{c} \\ \\ 1 \\ n+1 \to 6 \\ 2 \\ 7 \\ m \to 3 \\ n+m \to 8 \\ 4 \\ n \to 5 \end{array}
\begin{bmatrix}
\begin{array}{cc|cc|cc|c|c}
a_{11} & b_{11} & a_{12} & b_{21} & a_{13} & b_{31} & a_{14} & a_{15} \\
b_{11} & 0 & 0 & 0 & 0 & 0 & b_{14} & b_{15} \\
\hline
a_{21} & 0 & a_{22} & b_{22} & a_{23} & b_{32} & a_{24} & a_{25} \\
b_{21} & 0 & b_{22} & 0 & 0 & 0 & b_{24} & b_{25} \\
\hline
a_{31} & 0 & a_{32} & 0 & a_{33} & b_{33} & a_{34} & a_{35} \\
b_{31} & 0 & b_{32} & 0 & b_{33} & 0 & b_{34} & b_{35} \\
\hline
a_{41} & b_{14} & a_{42} & b_{24} & a_{43} & b_{34} & a_{44} & a_{45} \\
a_{51} & b_{15} & a_{52} & b_{25} & a_{53} & b_{35} & a_{54} & a_{55}
\end{array}
\end{bmatrix}, \tag{8}
$$

where the numbers along the border of $\boldsymbol{P_\pi}^T \mathcal{A} \boldsymbol{P_\pi}$ form the range of π. For general n and m, let $\boldsymbol{F} = \boldsymbol{P_\pi}^T \mathcal{A} \boldsymbol{P_\pi}$. Then the blocks F_{ij} of order 1 and 2 for $1 \le i, j \le n$ are given by:

$$
F_{ij} = \begin{cases}
\begin{bmatrix} a_{ii} & b_{ii} \\ b_{ii} & 0 \end{bmatrix}, & 1 \le i = j \le m\ ; \\[2ex]
\begin{bmatrix} a_{ij} & 0 \\ b_{ij} & 0 \end{bmatrix}, & 1 \le j < i \le m\ ; \\[2ex]
\begin{bmatrix} a_{ij} & b_{ji} \\ 0 & 0 \end{bmatrix}, & 1 \le i < j \le m\ ;
\end{cases}
\quad \text{and} \quad
F_{ij} = \begin{cases}
\begin{bmatrix} a_{ij} \\ b_{ij} \end{bmatrix}, & 1 \le i \le m < j \le n\ ; \\[2ex]
\begin{bmatrix} a_{ij} & b_{ji} \end{bmatrix}, & 1 \le j \le m < i \le n\ ; \\[2ex]
\begin{bmatrix} a_{ij} \end{bmatrix}, & m < i, j \le n.
\end{cases}
$$

3.2 Existence, uniqueness and sparsity the block $L_b D_b^{-1} L_b^T$ factorization

Now, we have a nicely partitioned $n \times n$ block structured matrix $P_{\boldsymbol{\pi}}{}^T \mathcal{A} P_{\boldsymbol{\pi}}$ with blocks of order 1 and 2, in which the block diagonal forms a priori pivots. Based on the structure and properties of these blocks and partly due to π, we show that the block $L_b D_b^{-1} L_b^T$ factorization of $P_{\boldsymbol{\pi}}{}^T \mathcal{A} P_{\boldsymbol{\pi}}$ exists uniquely.

Theorem 1. *Suppose an $(n+m) \times (n+m)$ permutation matrix $P_{\boldsymbol{\pi}}$ is defined as in (6) and $F = P_{\boldsymbol{\pi}}^T \mathcal{A} P_{\boldsymbol{\pi}}$, where $\mathcal{A}$ is $(n+m) \times (n+m)$ transformed symmetric indefinite matrix in (5). Then there exists $(n+m) \times (n+m)$ nonsingular block lower triangular matrix L_b with blocks of order 1 and 2 such that*

$$F = L_b D_b^{-1} L_b^T \tag{9}$$

and the 2×2 blocks $L_{ij} = F_{ij}$, $1 \leq j \leq i \leq m$, where $D_b = \mathrm{diag}(L_b)$ is the block diagonal part of L_b. It is unique with respect to $P_{\boldsymbol{\pi}}$.

Proof. Similar to (39) in the appendix, one can deduce iteratively

$$L_{ij} = F_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{kk}^{-1} L_{jk}^T \tag{10}$$

where

$$L_{ij} = \begin{cases} 2 \times 2 \text{ blocks for } 1 \leq j \leq i \leq m; \\ 1 \times 2 \text{ blocks for } 1 \leq j \leq m < i \leq n; \\ 1 \times 1 \text{ blocks for } m < j \leq i \leq n. \end{cases} \tag{11}$$

We show the existence of (10) by induction on $j \leq i \leq n$ as follows:

(i) 2×2 blocks.

Note that for each $k = 1, \ldots, m$,

$$F_{kk}^{-1} = \begin{bmatrix} a_{kk} & b_{kk} \\ b_{kk} & 0 \end{bmatrix}^{-1} = \frac{1}{b_{kk}^2} \begin{bmatrix} 0 & b_{kk} \\ b_{kk} & -a_{kk} \end{bmatrix}$$

which exists since $b_{kk} \neq 0$. All the 2×2 blocks $L_{ij} = F_{ij}$, $1 \leq j \leq i \leq m$, which

is trivially shown in the following.

For $j = 1, \ldots, m$

For $i = j, \ldots, m$

$$L_{ij} = F_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{kk}^{-1} L_{jk}^T$$

$$= F_{ij} - \sum_{k=1}^{j-1} \begin{bmatrix} a_{ik} & 0 \\ b_{ik} & 0 \end{bmatrix} \frac{1}{b_{kk}^2} \begin{bmatrix} 0 & b_{kk} \\ b_{kk} & -a_{kk} \end{bmatrix} \begin{bmatrix} a_{jk} & b_{jk} \\ 0 & 0 \end{bmatrix}$$

$$= F_{ij} - \sum_{k=1}^{j-1} \begin{bmatrix} a_{ik} & 0 \\ b_{ik} & 0 \end{bmatrix} \frac{1}{b_{kk}^2} \begin{bmatrix} 0 & 0 \\ b_{kk} a_{jk} & b_{kk} b_{jk} \end{bmatrix} = F_{ij}.$$

(ii) 1×2 blocks.

Let the 1×2 blocks be $L_{ij} = \begin{bmatrix} \alpha_{ij} & \beta_{ji} \end{bmatrix}$, where $1 \leq j \leq m < i \leq n$.

From (10) and (11), we obtain

For $j = 1, \ldots, m$

For $i = m + 1, \ldots, n$

$$L_{ij} = F_{ij} - \sum_{k=1}^{j-1} \underbrace{L_{ik}}_{1 \times 2} \underbrace{L_{kk}^{-1} L_{jk}^T}_{2 \times 2}$$

$$= F_{ij} - \sum_{k=1}^{j-1} \begin{bmatrix} \alpha_{ik} & \beta_{ki} \end{bmatrix} \frac{1}{b_{kk}^2} \begin{bmatrix} 0 & b_{kk} \\ b_{kk} & -a_{kk} \end{bmatrix} \begin{bmatrix} a_{jk} & b_{jk} \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} a_{ij} & b_{ji} \end{bmatrix} - \sum_{k=1}^{j-1} \begin{bmatrix} \frac{a_{jk}}{b_{kk}} \beta_{ki} & \frac{b_{jk}}{b_{kk}} \beta_{ki} \end{bmatrix}.$$

Hence

$$\alpha_{ij} = a_{ij} - \sum_{k=1}^{j-1} \frac{a_{jk}}{b_{kk}} \beta_{ki} \text{ and } \beta_{ji} = b_{ji} - \sum_{k=1}^{j-1} \frac{b_{jk}}{b_{kk}} \beta_{ki}, \ 1 \leq j \leq m < i \leq n, \quad (12)$$

exist since $b_{kk} \neq 0$.

(iii) 1×1 blocks.

Let the 1×1 blocks be $L_{ij} = \begin{bmatrix} \alpha_{ij} \end{bmatrix}$, where $m < j \leq i \leq n$.

8

From (10), (11) and (12), we obtain

For $j = m+1, \ldots, n$

  For $i = j, \ldots, n$

$$L_{ij} = F_{ij} - \sum_{k=1}^{m} \underbrace{L_{ik}}_{1\times 2} \underbrace{L_{kk}^{-1}}_{2\times 2} \underbrace{L_{jk}^{T}}_{2\times 1} - \sum_{k=m+1}^{j-1} \underbrace{L_{ik}}_{1\times 1} \underbrace{L_{kk}^{-1}}_{1\times 1} \underbrace{L_{jk}^{T}}_{1\times 1}$$

$$= a_{ij} - \sum_{k=1}^{m} \begin{bmatrix} \alpha_{ik} & \beta_{ki} \end{bmatrix} \frac{1}{b_{kk}^2} \begin{bmatrix} 0 & b_{kk} \\ b_{kk} & -a_{kk} \end{bmatrix} \begin{bmatrix} \alpha_{jk} \\ \beta_{kj} \end{bmatrix} - \sum_{k=m+1}^{j-1} \alpha_{ik}\alpha_{kk}^{-1}\alpha_{jk}$$

$$= a_{ij} - \sum_{k=1}^{m} \frac{1}{b_{kk}^2} \left(\alpha_{ik}b_{kk}\beta_{kj} + \beta_{ki}b_{kk}\alpha_{jk} - \beta_{ki}a_{kk}\beta_{kj} \right) - \sum_{k=m+1}^{j-1} \frac{\alpha_{ik}\alpha_{jk}}{\alpha_{kk}}.$$

Apparently the 1×1 blocks (scalars) are given by

$$\alpha_{ij} = a_{ij} + \sum_{k=1}^{m} \frac{\beta_{ki}\beta_{kj}}{b_{kk}^2}a_{kk} - \frac{\alpha_{ik}\beta_{kj} + \alpha_{jk}\beta_{ki}}{b_{kk}} - \sum_{k=m+1}^{j-1} \frac{\alpha_{ik}\alpha_{jk}}{\alpha_{kk}}, \quad m < j \leq i \leq n, \tag{13}$$

which exist only if $\alpha_{ii} \neq 0$ for each $i = j = m + 1, \ldots, n$. For this, let a rectangular matrix $\boldsymbol{\mathcal{H}} \in \mathbb{R}^{(n-m)\times n}$ be defined by $\boldsymbol{\mathcal{H}} = [H \quad I_{n-m}]$, where I_{n-m} is the $(n-m) \times (n-m)$ identity matrix, and $H = [h_{kj}]$ such that

$$h_{kj} = -\frac{\beta_{ji}}{b_{jj}}, \quad 1 \leq j \leq m, \quad 1 \leq k \leq n - m, \quad i = k + m.$$

Define $\boldsymbol{G} = \boldsymbol{\mathcal{H}}\boldsymbol{A}\boldsymbol{\mathcal{H}}^T$. Then the elements of the matrix $\boldsymbol{G} = [g_{rs}]$ are defined by

$$g_{rs} = a_{ij} + \sum_{k=1}^{m} \frac{\beta_{ki}\beta_{kj}}{b_{kk}^2}a_{kk} - \frac{\alpha_{ik}\beta_{kj} + \alpha_{jk}\beta_{ki}}{b_{kk}}, \quad 1 \leq r, s \leq n - m, \ i = r + m, \ j = s + m. \tag{14}$$

Computation of g_{rs} is demonstrated below for $n = 4, m = 2$ (note that $\boldsymbol{A}$ is

symmetric, positive definite):

$$
\underbrace{\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}}_{\boldsymbol{G}} = \underbrace{\begin{bmatrix} -\frac{\beta_{13}}{b_{11}} & -\frac{\beta_{23}}{b_{22}} & 1 & 0 \\ -\frac{\beta_{14}}{b_{11}} & -\frac{\beta_{24}}{b_{22}} & 0 & 1 \end{bmatrix}}_{\mathcal{H}} \underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} -\frac{\beta_{13}}{b_{11}} & -\frac{\beta_{14}}{b_{11}} \\ -\frac{\beta_{23}}{b_{22}} & -\frac{\beta_{24}}{b_{22}} \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathcal{H}^T}
$$

where

$$
\begin{aligned}
g_{11} &= a_{33} + \frac{\beta_{13}^2}{b_{11}^2} a_{11} + \frac{\beta_{23}^2}{b_{22}^2} a_{22} - 2\frac{a_{31}\beta_{13}}{b_{11}} - 2\left(a_{32} - \frac{a_{21}}{b_{11}}\beta_{13}\right)\frac{\beta_{23}}{b_{22}} \\
&= a_{33} + \frac{\beta_{13}^2}{b_{11}^2} a_{11} + \frac{\beta_{23}^2}{b_{22}^2} a_{22} - 2\frac{\alpha_{31}\beta_{13}}{b_{11}} - 2\frac{\alpha_{32}\beta_{23}}{b_{22}} \quad \text{using (12);} \\
g_{21} = g_{12} &= a_{43} + \frac{\beta_{13}\beta_{14}}{b_{11}^2} a_{11} + \frac{\beta_{23}\beta_{24}}{b_{22}^2} a_{22} - \frac{a_{31}\beta_{14} + a_{41}\beta_{13}}{b_{11}} - \left(a_{32} - \frac{a_{21}}{b_{11}}\beta_{13}\right)\frac{\beta_{24}}{b_{22}} - \left(a_{42} - \frac{a_{21}}{b_{11}}\beta_{14}\right)\frac{\beta_{23}}{b_{22}} \\
&= a_{43} + \frac{\beta_{13}\beta_{14}}{b_{11}^2} a_{11} + \frac{\beta_{23}\beta_{24}}{b_{22}^2} a_{22} - \frac{a_{31}\beta_{14} + a_{41}\beta_{13}}{b_{11}} - \frac{\alpha_{32}\beta_{24} + \alpha_{42}\beta_{23}}{b_{22}} \quad \text{again by the fact of (12);} \\
g_{22} &= a_{44} + \frac{\beta_{14}^2}{b_{11}^2} a_{11} + \frac{\beta_{24}^2}{b_{22}^2} a_{22} - 2\frac{a_{41}\beta_{24}}{b_{11}} - 2\left(a_{42} - \frac{a_{21}}{b_{11}}\beta_{14}\right)\frac{\beta_{24}}{b_{22}} \\
&= a_{44} + \frac{\beta_{14}^2}{b_{11}^2} a_{11} + \frac{\beta_{24}^2}{b_{22}^2} a_{22} - 2\frac{\alpha_{41}\beta_{24}}{b_{11}} - 2\frac{\alpha_{42}\beta_{24}}{b_{22}}
\end{aligned}
$$

Combining (13) and (14), we get

$$
\alpha_{ij} = g_{rs} - \sum_{k=m+1}^{j-1} \frac{\alpha_{ik}\alpha_{jk}}{\alpha_{kk}}, \quad 1 \le r, s \le n - m, \ i = r + m, \ j = s + m. \quad (15)
$$

From (15), it is evident that the 1×1 block elements are from a lower triangular matrix $\boldsymbol{L}$ such that $\boldsymbol{L}\boldsymbol{D}^{-1}\boldsymbol{L}^T = \boldsymbol{G}$, where $\boldsymbol{D} = \text{diag}(\boldsymbol{L})$. Since the matrix H has full rank $n - m$, the matrix $\boldsymbol{G}$ is symmetric positive definite. Hence by the theorem given in the appendix, $\boldsymbol{L}\boldsymbol{D}^{-1}\boldsymbol{L}^T = \boldsymbol{G}$ exists and $\alpha_{ii} > 0$ for each $i = j = m + 1, \ldots, n$. The uniqueness also follows from the same theorem in the appendix.

$\square$

From (i), (ii) and (iii), the blocks of $\boldsymbol{L_b}$ are given by

$$
L_{ij} = \begin{cases} \begin{bmatrix} a_{ii} & b_{ii} \\ b_{ii} & 0 \end{bmatrix}, & 1 \le i = j \le m \ ; \\[2ex] \begin{bmatrix} a_{ij} & 0 \\ b_{ij} & 0 \end{bmatrix}, & 1 \le j < i \le m \ ; \\[2ex] \begin{bmatrix} \alpha_{ij} & \beta_{ji} \end{bmatrix}, & 1 \le j \le m < i \le n \ ; \\[2ex] \begin{bmatrix} \alpha_{ij} \end{bmatrix}, & m < j \le i \le n. \end{cases} \quad (16)
$$

For example, $\boldsymbol{L_b} \boldsymbol{D_b}^{-1} \boldsymbol{L_b}^T$ factorization of $\boldsymbol{F}$ in (8) gives:

$$
\boldsymbol{L_b} = \left[\begin{array}{cc|cc|cc|cc}
a_{11} & b_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\
b_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
a_{21} & 0 & a_{22} & b_{22} & 0 & 0 & 0 & 0 \\
b_{21} & 0 & b_{22} & 0 & 0 & 0 & 0 & 0 \\
\hline
a_{31} & 0 & a_{32} & 0 & a_{33} & b_{33} & 0 & 0 \\
b_{31} & 0 & b_{32} & 0 & b_{33} & 0 & 0 & 0 \\
\hline
\alpha_{41} & \beta_{14} & \alpha_{42} & \beta_{24} & \alpha_{43} & \beta_{34} & \alpha_{44} & 0 \\
\alpha_{51} & \beta_{15} & \alpha_{52} & \beta_{25} & \alpha_{53} & \beta_{35} & \alpha_{54} & \alpha_{55}
\end{array} \right]. \tag{17}
$$

Obviously, only the 1×2 and 1×1 blocks are required to be computed, while the 2×2 blocks are straight away from the ones in $\boldsymbol{F}$. Perhaps the computational cost spent on transformation of $\mathring{\boldsymbol{B}}$ to $\boldsymbol{B}$ is compensated here. Consequently, the first $2m$ rows of $\boldsymbol{L_b}$ are sparse since their elements are directly from the sparse blocks $\boldsymbol{A_{11}}$ and $\boldsymbol{B_1}$. The pivots are given by

$$
\boldsymbol{D_b} = \mathrm{diag}(\boldsymbol{L_b}) = \bigoplus_{k=1}^{m} \begin{bmatrix} a_{kk} & b_{kk} \\ b_{kk} & 0 \end{bmatrix} \oplus \bigoplus_{k=2m+1}^{n+m} \alpha_{kk},
$$

in which only the 1×1 pivots are updated comparing to a priori pivots. Algorithm 1 gives a MATLAB version for the factorization $\boldsymbol{F} = \boldsymbol{L_b} \boldsymbol{D_b}^{-1} \boldsymbol{L_b}^T$. The function $\boldsymbol{s}$ involved in the algorithm is defined by

$$
\boldsymbol{s}(l) = \begin{cases} [l, \quad l+1] & \text{if } l \text{ is a row index} \text{ and } 1 \leq l \leq m, \\ [l, \quad l+1]^T & \text{if } l \text{ is a column index and } 1 \leq l \leq m, \\ [l] & \text{if } m+1 \leq l \leq n. \end{cases}
$$

It determines the positions of elements of the lth block. We use $\boldsymbol{V}(\boldsymbol{s}(k), :)$ to optimize the computational complexity (see appendix).

Below, in Definition 2, we introduce a permutation $\boldsymbol{\sigma}$ that swaps the two rows of every 2×2 block row of $\boldsymbol{L_b}$. This is done in order to obtain a lower triangular matrix $\boldsymbol{L}$, which can be used to solve the system (1) through backward substitution.

Definition 2. *Let a permutation $\sigma : \mathcal{N}_{n+m} \to \mathcal{N}_{n+m}$ be defined by*

$$
\sigma = \begin{pmatrix} 1, & 2, & \cdots, & 2m-1, & 2m, & 2m+1, & \cdots, & n+m \\ 2, & 1, & \cdots, & 2m, & 2m-1, & 2m+1, & \cdots, & n+m \end{pmatrix}.
$$

The related permutation matrix $\boldsymbol{P_\sigma}$ of size $(n+m) \times (n+m)$ is given by

$$
\boldsymbol{P_\sigma} = \begin{bmatrix} \boldsymbol{e}_2, & \boldsymbol{e}_1, & \cdots, & \boldsymbol{e}_{2m}, & \boldsymbol{e}_{2m-1}, & \boldsymbol{e}_{2m+1}, & \cdots, & \boldsymbol{e}_{n+m} \end{bmatrix}. \tag{18}
$$

It is easy to see that

$$
\boldsymbol{P_\sigma}^T \boldsymbol{P_\pi}^T \boldsymbol{\mathcal{A}} \boldsymbol{P_\pi} \boldsymbol{P_\sigma} = \boldsymbol{L} \boldsymbol{D_b}^{-1} \boldsymbol{L}^T, \tag{19}
$$

Algorithm 1 For a block structured symmetric indefinite matrix $\boldsymbol{F} \in \mathbb{R}^{(n+m)\times(n+m)}$ with blocks of order 1 and 2, the algorithm computes a nonsingular block lower triangular matrix $\boldsymbol{L_b} \in \mathbb{R}^{(n+m)\times(n+m)}$ with blocks of order 1 and 2 such that $\boldsymbol{F} = \boldsymbol{L_b}\boldsymbol{D_b}^{-1}\boldsymbol{L_b}^T$. For $j = 1,\ldots,n$, the jth block column $\boldsymbol{L_b}\big(\boldsymbol{s}(j) : n + m, \ j\big)$ overwrites $\boldsymbol{F}\big(\boldsymbol{s}(j) : n + m, \ \boldsymbol{s}(j)\big)$.

1: **for** $j = 1 : n$
2:     **for** $k = 1 : j - 1$
3:         $\boldsymbol{V}\big(\boldsymbol{s}(k), \ :\big) = \boldsymbol{F}\big(\boldsymbol{s}(k), \ \boldsymbol{s}(k)\big)^{-1}\boldsymbol{F}\big(\boldsymbol{s}(j), \ \boldsymbol{s}(k)\big)^T$
4:     **end for**
5:     **for** $i = j : n$
6:         $\boldsymbol{F}\big(\boldsymbol{s}(i), \ \boldsymbol{s}(j)\big) = \boldsymbol{F}\big(\boldsymbol{s}(j), \ \boldsymbol{s}(j)\big) - \sum\limits_{l=1}^{j-1} \boldsymbol{F}\big(\boldsymbol{s}(i), \ \boldsymbol{s}(l)\big) \boldsymbol{V}\big(\boldsymbol{s}(l), \ :\big)$
7:     **end for**
8: **end for**

---

where $\boldsymbol{L} = \boldsymbol{P_\sigma}^T\boldsymbol{L_b}$ is a lower triangular matrix whose block elements, say $\tilde{L}_{ij}$, are given by:

$$
\tilde{L}_{ij} =
\begin{cases}
\begin{bmatrix} b_{ii} & 0 \\ a_{ii} & b_{ii} \end{bmatrix}, & 1 \leq i = j \leq m \ ; \\[12pt]
\begin{bmatrix} b_{ij} & 0 \\ a_{ij} & 0 \end{bmatrix}, & 1 \leq j < i \leq m \ ; \\[12pt]
\begin{bmatrix} \alpha_{ij} & \beta_{ji} \end{bmatrix}, & 1 \leq j \leq m < i \leq n \ ; \\[8pt]
\begin{bmatrix} \alpha_{ij} \end{bmatrix}, & m < j \leq i \leq n,
\end{cases}
\tag{20}
$$

meaning every i^{th}, $1 \leq i \leq m$, 2 by 2 block row of $\boldsymbol{L_b}$ is row-interchanged, while the other $n - m$ block rows remain unaltered. From the example of $\boldsymbol{L_b}$ in (17), we obtain

$$
\boldsymbol{P_\sigma} = \begin{bmatrix} \boldsymbol{e}_2, & \boldsymbol{e}_1, & \boldsymbol{e}_4, & \boldsymbol{e}_3, & \boldsymbol{e}_6, & \boldsymbol{e}_5, & \boldsymbol{e}_7, & \boldsymbol{e}_8 \end{bmatrix},
$$

which gives

$$
\boldsymbol{L} = \boldsymbol{P_\sigma}^T\boldsymbol{L_b} =
\left[\begin{array}{cc|cc|cc|c|c}
b_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
a_{11} & b_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
b_{21} & 0 & b_{22} & 0 & 0 & 0 & 0 & 0 \\
a_{21} & 0 & a_{22} & b_{22} & 0 & 0 & 0 & 0 \\
\hline
b_{31} & 0 & b_{32} & 0 & b_{33} & 0 & 0 & 0 \\
a_{31} & 0 & a_{32} & 0 & a_{33} & b_{33} & 0 & 0 \\
\hline
\alpha_{41} & \beta_{14} & \alpha_{42} & \beta_{24} & \alpha_{43} & \beta_{34} & \alpha_{44} & 0 \\
\hline
\alpha_{51} & \beta_{15} & \alpha_{52} & \beta_{25} & \alpha_{53} & \beta_{35} & \alpha_{54} & \alpha_{55}
\end{array}\right].
$$

Hence

$$
\boldsymbol{L}\boldsymbol{D_b}^{-1}\boldsymbol{L}^T\boldsymbol{u} = \boldsymbol{d} \quad \Rightarrow \quad \boldsymbol{L}\boldsymbol{v} = \boldsymbol{d}, \ \boldsymbol{w} = \boldsymbol{D_b}\boldsymbol{v}, \ \boldsymbol{L}^T\boldsymbol{u} = \boldsymbol{w},
$$

where $\boldsymbol{d} = \boldsymbol{P_\sigma}^T \boldsymbol{P_\pi}^T \boldsymbol{\mathcal{T}}^T \mathring{\boldsymbol{d}}$ and $\mathring{\boldsymbol{u}} = \boldsymbol{\mathcal{T}} \boldsymbol{P_\pi} \boldsymbol{P_\sigma} \boldsymbol{u}$, which solves the system (1) through backward substitution. The block diagonal $\boldsymbol{D_b}$ is not required to invert, since the vector $\boldsymbol{w}$ is directly obtained from the matrix vector product $\boldsymbol{D_b v}$. We only need to extract the block diagonal matrix $\boldsymbol{D_b}$ from $\boldsymbol{L_b}$ using the function $\boldsymbol{s}$.

3.3 Numerical stability

Pivoting strategies are required in order to address the fundamental issue on the bound of the growth factor, ρ which occurs during Gaussian elimination, defined by

$$\rho(\boldsymbol{A}) = \frac{\max\limits_{i,j,l} |a_{ij}^{(l)}|}{\max\limits_{i,j} |a_{ij}|}$$

where $a_{ij}^{(l)}$, $l = 1, \ldots, n$, are the elements of the reduced matrix, $\boldsymbol{A}^{(l)}$ at the lth step Gaussian elimination of an $n \times n$ matrix $\boldsymbol{A} = \boldsymbol{A}^{(1)}$.

In (12) and (13), the factors b_{ki}/b_{kk}, $1 \leq k \leq j - 1$, may lead to arbitrarily large updates of α_{ij} and β_{ji}, for each $1 \leq j \leq m < i \leq n$ and $m < j \leq i \leq n$. So, for stable $\boldsymbol{L_b} \boldsymbol{D_b}^{-1} \boldsymbol{L_b}^T$ factorization of $\boldsymbol{P_\pi}^T \boldsymbol{\mathcal{A}} \boldsymbol{P_\pi}$, we consider that $\mathring{\boldsymbol{B}}$ is transformed into a lower trapezoidal $\boldsymbol{B} = [\boldsymbol{B_1} \ \boldsymbol{B_2}]$ such that the diagonal elements of $\boldsymbol{B_1}$ satisfy the condition: $|b_{kk}| \geq |b_{ki}|$, $1 \leq i \leq n$ for each $k = 1, \ldots, m$. This condition is sufficient for the bound of the growth factor ρ, which is shown in Theorem 2. Since all the 2×2 blocks are directly from $\boldsymbol{\mathcal{A}}$, we have to show only the bounds for the elements of 1×2 and 1×1 blocks.

Theorem 2. *Suppose a saddle point matrix $\mathring{\boldsymbol{A}} \in \mathbb{R}^{n \times n}$ is symmetrically transformed into $\boldsymbol{\mathcal{A}}$ as in (5) such that the lower trapezoidal form $\boldsymbol{B} = [\boldsymbol{B_1} \ \boldsymbol{B_2}]$ satisfies the condition: $|b_{kk}| \geq |b_{ki}|$, $1 \leq i \leq n$ for each $k = 1, \ldots, m$. Let $\boldsymbol{P_\pi}$ be the permutation matrix in (6). If the block factorization $\boldsymbol{P_\pi}^T \boldsymbol{\mathcal{A}} \boldsymbol{P_\pi} = \boldsymbol{L_b} \boldsymbol{D_b}^{-1} \boldsymbol{L_b}^T$ runs to completion, then the growth factor ρ is bounded by*

$$\rho(\boldsymbol{\mathcal{A}}) = \frac{\max\limits_{i,j,l} \left\{ |a_{ij}^{(l)}|, |b_{ji}^{(l)}| \right\}}{\max\limits_{i,j} \left\{ |a_{ii}|, |b_{jj}| \right\}} \leq 2^{2m}, \quad 1 \leq l \leq n. \tag{21}$$

Proof. From (12), it is easy to see that

$$\text{For } j = 1, \ldots, m$$
$$\text{For } i = m+1, \ldots, n$$
$$\left[|\alpha_{ij}| \quad |\beta_{ji}| \right] \leq 2^{j-1} \left[\max\limits_{i,j} |a_{ij}| \quad |b_{jj}| \right]. \tag{22}$$

Since we do not know whether $\max\limits_{i,j} \{|a_{ij}|\}$ or $\max\limits_{j} \{|b_{jj}|\}$ is the largest, the common upper bound on the elements of 1×2 blocks, $1 \leq j \leq m < i \leq n$, is

$$\max\limits_{i,j} \left\{ |a_{ij}^{(j)}|, |b_{ji}^{(j)}| \right\} = \max\limits_{i,j} \{|\alpha_{ij}|, |\beta_{ji}|\} \leq 2^{m-1} \max\limits_{i,j} \{|a_{ij}|, |b_{jj}|\}. \tag{23}$$

The 1×1 blocks of $\boldsymbol{L_b}$ are the elements decomposed from the symmetric positive definite matrix $\boldsymbol{G} \in \mathbb{R}^{(n-m) \times (n-m)}$ defined in (14). Since, the Gaussian elimination growth factor for a symmetric positive definite matrix without pivoting is equal to 1 [37, p. 239], the bound on 1×1 blocks α_{ij}, $m < i, j < n$, in (13) is the same as that on $\boldsymbol{G}$. It suffices to show that the elements of $\boldsymbol{G}$ are bounded. From (14) and (22), for $1 \le r, s \le n - m$, $i = r + m$, $j = s + m$, we get

$$
\begin{aligned}
|g_{rs}| &\le |a_{ij}| + \sum_{k=1}^{m} \frac{|\beta_{ki}|\,|\beta_{kj}|}{|b_{kk}|^2}|a_{kk}| + \frac{|\alpha_{ik}|\,|\beta_{kj}| + |\alpha_{jk}|\,|\beta_{ki}|}{|b_{kk}|}, \\
&\le |a_{ij}| + \sum_{k=1}^{m} 2^{2(k-1)}|a_{kk}| + 2^{k-1}|\alpha_{ik}| + 2^{k-1}|\alpha_{jk}|, \\
&\le |a_{ij}| + 3 \max_{i,j}|a_{ij}| \sum_{k=1}^{m} 2^{2k-2}, \\
&\le 2^{2m} \max_{i,j}|a_{ij}|.
\end{aligned}
\tag{24}
$$

Combining (23) and (24) , we obtain

$$
\max_{i,j,l} \left\{ |a_{ij}^{(l)}|, |b_{ji}^{(l)}| \right\} \le 2^{2m} \max_{i,j} \left\{ |a_{ij}|, |b_{jj}| \right\} = 2^{2m} \max_{i,j} \left\{ |a_{ij}|, |b_{ij}| \right\}, \ 1 \le l \le n. \tag{25}
$$

$\square$

For Gaussian elimination of a matrix of size $(m + n) \times (m + n)$ with partial pivoting, Wilkinson [41] showed that $\rho \le 2^{m+n-1}$. The upper bound that we have derived in (25) is sharper for $m \le n - 1$, which complies with our assumption of $m < n$ but only with sufficient condition on $\boldsymbol{B}$ as stated in Theorem 2. We prefer to use the above mentioned transformation of $\mathring{\boldsymbol{B}}$ only if $\mathring{\boldsymbol{A}}$ is almost ill-conditioned, since it might not ensure sparse $\boldsymbol{B}$. In fact, such a transformation is equivalent to partial pivoting of $\mathring{\boldsymbol{A}}$ without having to pivot the symmetric positive definite matrix $\mathring{\boldsymbol{A}}$.

4 The new block factorization versus the Schilders' block factorization

Schilders' factors consist of a block 3×3 structure with blocks of order m and $n - m$, which are computed directly from the blocks with similar orders of the transformed saddle point matrix $\boldsymbol{\mathcal{T}}^T \mathring{\boldsymbol{A}} \boldsymbol{\mathcal{T}}$, for a different $\boldsymbol{\mathcal{T}}$. For the interest of comparison, we show that the block $\boldsymbol{L_b} \boldsymbol{D_b}^{-1} \boldsymbol{L_b}^T$ factorization can also be induced to such a block 3×3 structured factorization with blocks of order m and $n - m$. Our persuasion here is solely based on application of the inverse of the permutation π. In other words, we do not go for any further computations in order to form the blocks of order m and $n - m$ from the factors $\boldsymbol{L_b}$ and $\boldsymbol{D_b}^{-1}$. The induced block factors are different and much more sparser than the Schilders' factors for large n and m. We provide both theoretical and numerical aspects, which distinguish these two factorizations.

4.1 Induced block factorization

Let $\boldsymbol{P}_{\pi^{-1}}$ be a permutation matrix of order $(n+m) \times (n+m)$ defined by the inverse π^{-1}. Applying $\boldsymbol{P}_{\pi^{-1}}$ congruently to $\boldsymbol{F} = \boldsymbol{P_\pi}^T \boldsymbol{A} \boldsymbol{P_\pi}$ restores it to $\boldsymbol{A}$ and so the blocks $\boldsymbol{A}_{ij}$ and $\boldsymbol{B}_i$, $i = 1, 2$. However, the question here is – if $\boldsymbol{L_b}$ is reformed with a similar application of $\boldsymbol{P}_{\pi^{-1}}$ to it, will there be well-defined blocks of order m and $n - m$ in $\boldsymbol{P}_{\pi^{-1}}^T \boldsymbol{L_b} \boldsymbol{P}_{\pi^{-1}}$, which can be related to the blocks $\boldsymbol{A}_{ij}$ and $\boldsymbol{B}_i$? To answer this question, consider the example of $\boldsymbol{F}$ with $n = 5$, $m = 3$ in (7). From (17), we have:

$$
\boldsymbol{L_b} =
\begin{array}{c}
\\ \\ \\ 1 \\ n+1 \to 6 \\ 2 \\ 7 \\ m \to 3 \\ n+m \to 8 \\ 4 \\ n \to 5
\end{array}
\begin{array}{c}
\begin{array}{cccccccc}
 & \overset{n+1}{\underset{\downarrow}{}} & & & \overset{m}{\underset{\downarrow}{}} & \overset{n+m}{\underset{\downarrow}{}} & & \overset{n}{\underset{\downarrow}{}} \\
1 & 6 & 2 & 7 & 3 & 8 & 4 & 5
\end{array} \\
\left[
\begin{array}{cc|cc|cc|cc}
a_{11} & b_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\
b_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
a_{21} & 0 & a_{22} & b_{22} & 0 & 0 & 0 & 0 \\
b_{21} & 0 & b_{22} & 0 & 0 & 0 & 0 & 0 \\
a_{31} & 0 & a_{32} & 0 & a_{33} & b_{33} & 0 & 0 \\
b_{31} & 0 & b_{32} & 0 & b_{33} & 0 & 0 & 0 \\
\alpha_{41} & \beta_{14} & \alpha_{42} & \beta_{24} & \alpha_{43} & \beta_{34} & \alpha_{44} & 0 \\
\alpha_{51} & \beta_{15} & \alpha_{52} & \beta_{25} & \alpha_{53} & \beta_{35} & \alpha_{54} & \alpha_{55}
\end{array}
\right]
\end{array}
$$

and

$$
\boldsymbol{D_b}^{-1} =
\begin{array}{c}
\\ \\ \\ 1 \\ n+1 \to 6 \\ 2 \\ 7 \\ m \to 3 \\ n+m \to 8 \\ 4 \\ n \to 5
\end{array}
\begin{array}{c}
\begin{array}{cccccccc}
 & \overset{n+1}{\underset{\downarrow}{}} & & & \overset{m}{\underset{\downarrow}{}} & \overset{n+m}{\underset{\downarrow}{}} & & \overset{n}{\underset{\downarrow}{}} \\
1 & 6 & 2 & 7 & 3 & 8 & 4 & 5
\end{array} \\
\left[
\begin{array}{cc|cc|cc|cc}
0 & b_{11}^{-1} & 0 & 0 & 0 & 0 & 0 & 0 \\
b_{11}^{-1} & -a_{11}b_{11}^{-2} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & b_{22}^{-1} & 0 & 0 & 0 & 0 \\
0 & 0 & b_{22}^{-1} & -a_{22}b_{22}^{-2} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & b_{33}^{-1} & 0 & 0 \\
0 & 0 & 0 & 0 & b_{33}^{-1} & -a_{33}b_{33}^{-2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \alpha_{44}^{-1} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{55}^{-1}
\end{array}
\right]
\end{array}.
$$

The positions (indices) of the elements of $\boldsymbol{L_b}$ and $\boldsymbol{D_b}^{-1}$ are inherited from the elements of $\boldsymbol{A}$. Congruence application of π^{-1} to $\boldsymbol{L_b}$ and $\boldsymbol{D_b}^{-1}$ means taking all their elements back to the inherited positions in $\boldsymbol{A}$. For instance, with application of π^{-1} to $\boldsymbol{L_b}$ in the above example, the entries a_{32} and β_{41} are moved from their current positions $(5, 3)$ and $(7, 2)$ to their inherited positions $(3, 2)$ and $(4, 6)$, respectively. Applying π^{-1} to all other entries of $\boldsymbol{L_b}$ gives the following block 3×3 structure with blocks of order 3

and 2.

$$
\boldsymbol{P_{\pi^{-1}}}^T \boldsymbol{L_b} \boldsymbol{P_{\pi^{-1}}} =
\begin{array}{c}
\\
\\
\\
1 \\
2 \\
m \to 3 \\
4 \\
n \to 5 \\
n+1 \to 6 \\
7 \\
n+m \to 8
\end{array}
\begin{bmatrix}
\overset{1}{a_{11}} & \overset{2}{0} & \overset{\overset{m}{\downarrow}}{\underset{3}{0}} & \overset{4}{0} & \overset{\overset{n}{\downarrow}}{\underset{5}{0}} & \overset{\overset{n+1}{\downarrow}}{\underset{6}{b_{11}}} & \overset{7}{0} & \overset{\overset{n+m}{\downarrow}}{\underset{8}{0}} \\
a_{21} & a_{22} & 0 & 0 & 0 & 0 & b_{22} & 0 \\
a_{31} & a_{32} & a_{33} & 0 & 0 & 0 & 0 & b_{33} \\
\alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} & 0 & \beta_{14} & \beta_{24} & \beta_{34} \\
\alpha_{51} & \alpha_{52} & \alpha_{53} & \alpha_{54} & \alpha_{55} & \beta_{15} & \beta_{25} & \beta_{35} \\
b_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
b_{21} & b_{22} & 0 & 0 & 0 & 0 & 0 & 0 \\
b_{31} & b_{32} & b_{33} & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

and

$$
\boldsymbol{P_{\pi^{-1}}}^T \boldsymbol{D_b}^{-1} \boldsymbol{P_{\pi^{-1}}} =
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & b_{11}^{-1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & b_{22}^{-1} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & b_{33}^{-1} \\
0 & 0 & 0 & \alpha_{44}^{-1} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \alpha_{55}^{-1} & 0 & 0 & 0 \\
b_{11}^{-1} & 0 & 0 & 0 & 0 & -a_{11}b_{11}^{-2} & 0 & 0 \\
0 & b_{22}^{-1} & 0 & 0 & 0 & 0 & -a_{22}b_{22}^{-2} & 0 \\
0 & 0 & b_{33}^{-1} & 0 & 0 & 0 & 0 & -a_{33}b_{33}^{-2}
\end{bmatrix},
$$

which are indeed as expected and have blocks with special structures and properties. For general m and n, let

$$
\boldsymbol{P_{\pi^{-1}}}^T \boldsymbol{L_b} \boldsymbol{P_{\pi^{-1}}} = \mathcal{L} =
\begin{bmatrix}
\boldsymbol{L_A} & \boldsymbol{0} & \boldsymbol{D_B} \\
\boldsymbol{M} & \boldsymbol{L} & \boldsymbol{H} \\
\boldsymbol{B_1} & \boldsymbol{0} & \boldsymbol{0}
\end{bmatrix}
\text{ and } \boldsymbol{P_{\pi^{-1}}}^T \boldsymbol{D_b}^{-1} \boldsymbol{P_{\pi^{-1}}} = \mathcal{D} =
\begin{bmatrix}
\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{D_B^{-1}} \\
\boldsymbol{0} & \boldsymbol{D} & \boldsymbol{0} \\
\boldsymbol{D_B^{-1}} & \boldsymbol{0} & -\boldsymbol{D_A D_B^{-2}}
\end{bmatrix}.
$$

Then, using the above example inductively, the blocks of order m and $n - m$ can be defined as follows:

$$
\left.
\begin{aligned}
& \boldsymbol{L_A} = [a_{ij}],\ 1 \le j \le i \le m,\ m \times m \text{ lower triangular matrix;} \\
& \boldsymbol{L} = [\alpha_{ij}],\ m < j \le i \le n,\ (n-m) \times (n-m) \text{ lower triangular matrix;} \\
& \boldsymbol{M} = [\alpha_{ij}],\ 1 \le j \le m < i \le n,\ (n-m) \times m \text{ rectangular matrix;} \\
& \boldsymbol{H} = [\beta_{ji}],\ 1 \le j \le m < i \le n,\ (n-m) \times m \text{ rectangular matrix;} \\
& \boldsymbol{D_B} = \mathrm{diag}(\boldsymbol{B_1}),\ m \times m \text{ diagonal matrix;} \\
& \boldsymbol{D_A} = \mathrm{diag}(\boldsymbol{L_A}),\ m \times m \text{ diagonal matrix; and} \\
& \boldsymbol{D} = \mathrm{diag}^{-1}(\boldsymbol{L}),\ (n-m) \times (n-m) \text{ diagonal matrix.}
\end{aligned}
\right\} \quad (26)
$$

All these blocks are directly from $\boldsymbol{L_b}$, they all exist and are well-defined with respect to the permutation inverse, π^{-1}. Also $\boldsymbol{L_A}$ is the lower triangular part of the block $\boldsymbol{A}_{11}$,

and $\boldsymbol{L}$ is the lower triangular matrix decomposed from symmetric, positive definite matrix $\boldsymbol{G}$ in (14), so they are nonsingular. Using these blocks, we easily prove Lemma 1 that gives the induced block factorization of $\boldsymbol{\mathcal{A}}$ with blocks of order m and $n - m$.

Lemma 1. *Consider the transformed symmetric indefinite matrix $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{(n+m) \times (n+m)}$ in (5) and the permutation matrix $\boldsymbol{P}_\pi$ in (6). Then the factorization $\boldsymbol{P_\pi}^T \boldsymbol{\mathcal{A}} \boldsymbol{P_\pi} = \boldsymbol{L_b} \boldsymbol{D_b}^{-1} \boldsymbol{L_b}^T$ can be induced to a block factorization with blocks of order m, $n - m$ and m such that*

$$\underbrace{\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \boldsymbol{B}_1^T \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \boldsymbol{B}_2^T \\ \boldsymbol{B}_1 & \boldsymbol{B}_2 & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{\mathcal{A}}} = \underbrace{\begin{bmatrix} \boldsymbol{L}_A & \boldsymbol{0} & \boldsymbol{D}_B \\ \boldsymbol{M} & \boldsymbol{L} & \boldsymbol{H} \\ \boldsymbol{B}_1 & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{\mathcal{L}}} \underbrace{\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{D}_B^{-1} \\ \boldsymbol{0} & \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{D}_B^{-1} & \boldsymbol{0} & -\boldsymbol{D}_A \boldsymbol{D}_B^{-2} \end{bmatrix}}_{\boldsymbol{\mathcal{D}}} \underbrace{\begin{bmatrix} \boldsymbol{L}_A^T & \boldsymbol{M}^T & \boldsymbol{B}_1^T \\ \boldsymbol{0} & \boldsymbol{L}^T & \boldsymbol{0} \\ \boldsymbol{D}_B & \boldsymbol{H}^T & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{\mathcal{L}}^T},$$

(27)

where $\boldsymbol{L}_A$, $\boldsymbol{L}$, $\boldsymbol{M}$, $\boldsymbol{H}$, $\boldsymbol{D}_A$, $\boldsymbol{D}_B$ and $\boldsymbol{D}$ are as in (26).

Proof. Notice that the blocks $\boldsymbol{L}_A$, $\boldsymbol{L}$, $\boldsymbol{M}$, $\boldsymbol{H}$, $\boldsymbol{D}_A$, $\boldsymbol{D}_B$ and $\boldsymbol{D}$ defined in (26) exist due to the existence of $\boldsymbol{L_b}$, we only need to show that (27) holds. Applying $\boldsymbol{P}_{\pi^{-1}}$ congruently on $\boldsymbol{P}_\pi^T \boldsymbol{\mathcal{A}} \boldsymbol{P}_\pi$ and using Theorem 1, we obtain:

$$\begin{aligned} \boldsymbol{\mathcal{A}} &= \boldsymbol{P}_{\pi^{-1}}^T \left(\boldsymbol{P}_\pi^T \boldsymbol{\mathcal{A}} \boldsymbol{P}_\pi \right) \boldsymbol{P}_{\pi^{-1}} \\ &= \boldsymbol{P}_{\pi^{-1}}^T \left(\boldsymbol{L_b} \boldsymbol{D_b}^{-1} \boldsymbol{L_b}^T \right) \boldsymbol{P}_{\pi^{-1}} \\ &= \left(\boldsymbol{P}_{\pi^{-1}}^T \boldsymbol{L_b} \boldsymbol{P}_{\pi^{-1}} \right) \left(\boldsymbol{P}_{\pi^{-1}}^T \boldsymbol{D_b}^{-1} \boldsymbol{P}_{\pi^{-1}} \right) \left(\boldsymbol{P}_{\pi^{-1}}^T \boldsymbol{L_b}^T \boldsymbol{P}_{\pi^{-1}} \right) \\ &= \boldsymbol{\mathcal{L}} \boldsymbol{\mathcal{D}} \boldsymbol{\mathcal{L}}^T. \end{aligned}$$

$\square$

Like in the proof of the Schilders' factorization, if we compare the right- and left-hand sides of (27), the blocks can be related by the following equations. Note that the product of any two diagonal matrices is commutative.

$$\boldsymbol{A}_{11} = \boldsymbol{L}_A + \boldsymbol{L}_A^T - \boldsymbol{D}_A, \tag{28}$$

$$\boldsymbol{A}_{21} = \boldsymbol{M} + \boldsymbol{E} \boldsymbol{D}_B^{-1} \boldsymbol{L}_A^T - \boldsymbol{E} \boldsymbol{D}_A \boldsymbol{D}_B^{-1}, \tag{29}$$

$$\boldsymbol{A}_{12} = \boldsymbol{M}^T + \boldsymbol{L}_A \boldsymbol{D}_B^{-1} \boldsymbol{E}^T - \boldsymbol{D}_A \boldsymbol{D}_B^{-1} \boldsymbol{E}^T, \tag{30}$$

$$\boldsymbol{A}_{22} = \boldsymbol{M} \boldsymbol{D}_B^{-1} \boldsymbol{E}^T + \boldsymbol{L} \boldsymbol{D} \boldsymbol{L}^T + \boldsymbol{E} \boldsymbol{D}_B^{-1} \boldsymbol{M}^T - \boldsymbol{E} \boldsymbol{D}_A \boldsymbol{D}_B^{-2} \boldsymbol{E}^T, \tag{31}$$

$$\boldsymbol{B}_2 = \boldsymbol{B}_1 \boldsymbol{D}_B^{-1} \boldsymbol{E}^T \text{ or } \boldsymbol{E} = \boldsymbol{B_2}^T \boldsymbol{B_1}^{-T} \boldsymbol{D}_B. \tag{32}$$

Working out the equations (28) through (32), one can obtain the relation

$$\boldsymbol{L} \boldsymbol{D} \boldsymbol{L}^T = \begin{bmatrix} -\boldsymbol{E} & \boldsymbol{I}_{n-m} \end{bmatrix} \boldsymbol{A} \begin{bmatrix} -\boldsymbol{E} & \boldsymbol{I}_{n-m} \end{bmatrix}^T. \tag{33}$$

It is quite evident from (27), that the induced block factors $\boldsymbol{\mathcal{L}}$ and $\boldsymbol{\mathcal{D}}$ are different from the ones in [2, 10, 11, 35], since they are deduced from $\boldsymbol{\mathcal{A}}$ by using a different

transformation operator $\mathcal{T}$. Furthermore, in [2, 10, 11], suggestions for the blocks of $\mathcal{L}$ and $\mathcal{D}$ are provided such that $\mathcal{L}\mathcal{D}\mathcal{L}^T$ approximates the block $\boldsymbol{A}$ by keeping the constraint matrix $\boldsymbol{B}$ intact, which is known as implicit factorization for preconditioners. In contrary, the induced block factorization in (27) and the Schilders' factorization give the exact factorization of $\mathcal{A}$. This motivates to draw a comparison between the Schilders' factorization and the induced block factorization.

4.2 Comparison with the Schilders' factorization

According to the Schilders' factorization [35, Lemma 4.1], a symmetric indefinite matrix $\mathring{\mathcal{A}} \in \mathbb{R}^{(n+m)\times(n+m)}$ is transformed into $\tilde{\mathcal{A}}$, partitioned into a block 3×3 structure and decomposed into the following form:

$$\underbrace{\begin{bmatrix} \tilde{\boldsymbol{A}}_{11} & \tilde{\boldsymbol{A}}_{12} & \tilde{\boldsymbol{B}}_1^T \\ \tilde{\boldsymbol{A}}_{21} & \tilde{\boldsymbol{A}}_{22} & \tilde{\boldsymbol{B}}_2^T \\ \tilde{\boldsymbol{B}}_1 & \tilde{\boldsymbol{B}}_2 & 0 \end{bmatrix}}_{\tilde{\mathcal{A}}} = \underbrace{\begin{bmatrix} \tilde{\boldsymbol{B}}_1^T & 0 & \tilde{\boldsymbol{L}}_1 \\ \tilde{\boldsymbol{B}}_2^T & \tilde{\boldsymbol{L}}_2 + \boldsymbol{I}_{n-m} & \tilde{\boldsymbol{M}} \\ 0 & 0 & \boldsymbol{I}_m \end{bmatrix}}_{\tilde{\mathcal{L}}} \underbrace{\begin{bmatrix} \tilde{\boldsymbol{D}}_1 & 0 & \boldsymbol{I}_m \\ 0 & \tilde{\boldsymbol{D}}_2 & 0 \\ \boldsymbol{I}_m & 0 & 0 \end{bmatrix}}_{\tilde{\mathcal{D}}} \tilde{\mathcal{L}}^T, \qquad (34)$$

where $\tilde{\boldsymbol{B}}_1$ is $m \times m$ upper triangular matrix $[\diagdown]$; $\tilde{\boldsymbol{L}}_1$ and $\tilde{\boldsymbol{L}}_2$ are respectively, $m \times m$ and $(n-m) \times (n-m)$ strictly lower triangular matrices; $\tilde{\boldsymbol{M}}$ is $(n-m) \times m$ rectangular matrix; and $\tilde{\boldsymbol{D}}_1$ and $\tilde{\boldsymbol{D}}_2$ are respectively, $m \times m$ and $(n-m) \times (n-m)$ diagonal matrices. By working out the left- and right-hand sides of (35), the blocks $\tilde{\boldsymbol{D}}_1$, $\tilde{\boldsymbol{L}}_1$ and $\tilde{\boldsymbol{M}}$ are computed from the following equations (details can be found in [35]):

$$\tilde{\boldsymbol{D}}_1 = \text{diag}\left(\tilde{\boldsymbol{B}}_1^{-T} \tilde{\boldsymbol{A}}_{11} \tilde{\boldsymbol{B}}_1^{-1}\right), \qquad (35)$$

$$\tilde{\boldsymbol{L}}_1 = \tilde{\boldsymbol{B}}_1^T \, \text{lower}\left(\tilde{\boldsymbol{B}}_1^{-T} \tilde{\boldsymbol{A}}_{11} \tilde{\boldsymbol{B}}_1^{-1}\right), \qquad (36)$$

$$\tilde{\boldsymbol{M}} = \left(\tilde{\boldsymbol{A}}_{21} - \tilde{\boldsymbol{B}}_2^T \tilde{\boldsymbol{L}}_1^T\right) \tilde{\boldsymbol{B}}_1^{-1} - \tilde{\boldsymbol{B}}_2^T \tilde{\boldsymbol{D}}_1, \qquad (37)$$

whereas the blocks $\tilde{\boldsymbol{L}}_2$ and $\tilde{\boldsymbol{D}}_2$ are to be determined from the Cholesky factorization of the reduced Hessian matrix $\boldsymbol{Z}^T \tilde{\boldsymbol{A}} \boldsymbol{Z}$, which is similar to the matrix on the right-hand side of (33), where $\boldsymbol{Z}^T = \begin{bmatrix} -\tilde{\boldsymbol{B}}_2^T \tilde{\boldsymbol{B}}_1^{-T} & \boldsymbol{I}_{n-m} \end{bmatrix}$. Although in general, the blocks $\tilde{\boldsymbol{A}}_{11}$ and $\tilde{\boldsymbol{B}}_1$ are sparse, the product $\tilde{\boldsymbol{B}}_1^{-T} \tilde{\boldsymbol{A}}_{11} \tilde{\boldsymbol{B}}_1^{-1}$ gets more fill-ins when $\tilde{\boldsymbol{B}}_1^{-1}$ is dense. Ultimately the the lower triangular block $\tilde{\boldsymbol{L}}_1$ in (36) and the rectangular block $\tilde{\boldsymbol{M}}$ in (37) turn out to be substantially full. It is shown in Table 1, that the induced blocks $\boldsymbol{D}_A$, $\boldsymbol{L}_A$ and $\boldsymbol{M}$ can be determined with a minimum involvement of the factor $\boldsymbol{B_1}^{-1}$, which leads them to be sparser than the corresponding Schilders' blocks $\tilde{\boldsymbol{D}}_1$, $\tilde{\boldsymbol{L}}_1$ and $\tilde{\boldsymbol{M}}$.

The induced block factorization in (27) has got even more advantage over the Schilders' factorization if the block $\boldsymbol{A}$ is a diagonal matrix, which occurs in applications such as resistor network modeling and some convex quadratic programming problems. It is clear from Table 1 that if $\boldsymbol{A}$ is diagonal, then the block $\boldsymbol{M} = 0$, since $\boldsymbol{L}_A = \boldsymbol{D}_A$ and

Schilders' $\tilde{\mathcal{L}}\tilde{\mathcal{D}}\tilde{\mathcal{L}}^T$	current $\mathcal{L}\mathcal{D}\mathcal{L}^T$
$\tilde{L}_1 = \tilde{B}_1^{\ T}\left(\text{strict lower}\left(\tilde{B}_1^{\ -T}\tilde{A}_{11}\tilde{B}_1^{\ -1}\right)\right)$	$L_A = \text{lower triangular}(A_{11})$
$\tilde{D}_1 = \text{diag}\left(\tilde{B}_1^{\ -T}\tilde{A}_{11}\tilde{B}_1^{\ -1}\right)$	$D_A = \text{diag}(A_{11})$
$\tilde{M} = \left(\tilde{A}_{21} - \tilde{B}_2^{\ T}\tilde{L}_1^{\ T}\right)\tilde{B}_1^{\ -1} - \tilde{B}_2^{\ T}\tilde{D}_1$	$M = A_{21} + B_2^T B_1^{-T}\left(D_A - L_A^T\right)$

Table 1: Comparison of the blocks from the Schilders' factorization and the induced block factorization in (27).

$A_{21} = 0$. Whereas that block $\tilde{M} = -\tilde{B}_2^{\ T}\left(\tilde{L}_1^{\ T}\tilde{B}_1^{\ -1} + \tilde{D}_1\right)$, which cannot be zero unless $\tilde{B}_2$ is a zero matrix.

5 Numerical experiments

We did numerical experiments on two different categories of saddle point matrices that are based on two types of constraint matrix $\mathring{B}$ as seen in Section 2. With regard to typical $\mathring{B}$, we conducted our tests on the saddle point matrices, which arise in resistor network modeling systems. With regard to more general form of $\mathring{B}$, we examined the saddle point matrices provided in the repository of the University of Florida (UF) sparse matrix collection [36], maintained by Tim Davis.

For resistor network modelings, the matrix A is a diagonal matrix with resistance values of n resistors, and $\mathring{B}$ is an incidence matrix having full row rank m. There are $m + 1$ nodes in a resistor network and one node is grounded. The row related to the ground node is deleted from the incidence matrix $\mathring{B}$, which makes the system stable [30]. As a result, $\mathring{B}$ has at most two nonzero elements in each column, which is permuted into a lower trapezoidal form. In order to understand the structures and sparsity patterns of $\mathcal{A}$, $P_\pi^{\ T}\mathcal{A}P_\pi$ and the factor L_b, we consider a small saddle point matrix $\mathring{\mathcal{A}}$ of size 113×113 from an industrial resistor network problem, consisting of 44 nodes (including the ground node) and 70 arcs. The visual representations of the resistor network and its associated saddle point matrix in this example are shown in Figure 1. The transformation operator $\mathcal{T}$ applied to $\mathring{\mathcal{A}}$ here is a permutation matrix of order 113. The transformed matrix $\mathcal{A}$ and its block partitioned form $P_\pi^{\ T}\mathcal{A}P_\pi$ with blocks of order 1 and 2 are shown in Figure 2. From Figure 3, we see that the block lower triangular factor L_b and its block diagonal part D_b contain the same structure of $P_\pi^{\ T}\mathcal{A}P_\pi$. We also obtained the factor $\mathcal{L}$ from L_b, which is compared with the Schilders' factor $\tilde{\mathcal{L}}$ as shown in Figure 4. The induced block diagonal factor $\mathcal{D}$ and the Schilders' block diagonal factor $\tilde{\mathcal{D}}$ are shown in Figure 5.

The numerical result for larger sizes of $\mathring{\mathcal{A}}$ in resistor network problems is given in Table 2. Regarding more general form of $\mathring{B}$, we experimented on 10 saddle point matrices available in the UF sparse matrix collection coming from various applications, which

is presented in Table 3. For all the matrices that we have chosen were being able to transform the $\mathring{\boldsymbol{B}}$ part into trapezoidal form such that $\mathring{\boldsymbol{A}}$ and $\boldsymbol{A}$ have the same sparsity. All the numerical tests were done in the MATLAB R2013b.

RNC20, $n = 70$, $m = 43$ $\mathring{\boldsymbol{A}}$, nonzeros $= 328$
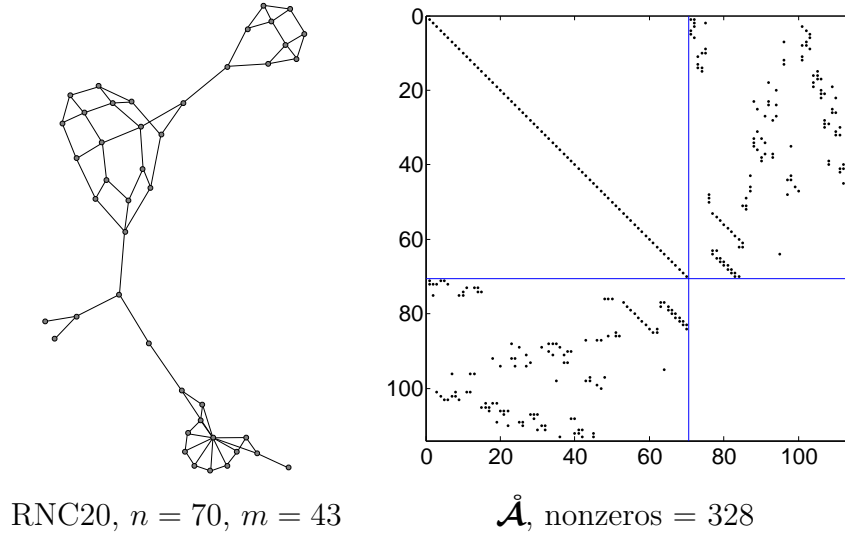
Figure 1: A realistic industrial resistor network, RNC20 and its associated saddle point matrix $\mathring{\boldsymbol{A}}$.

$\boldsymbol{A}$, nonzeros $= 328$ $\boldsymbol{P_\pi}^T \boldsymbol{A} \boldsymbol{P_\pi}$, nonzeros $= 328$
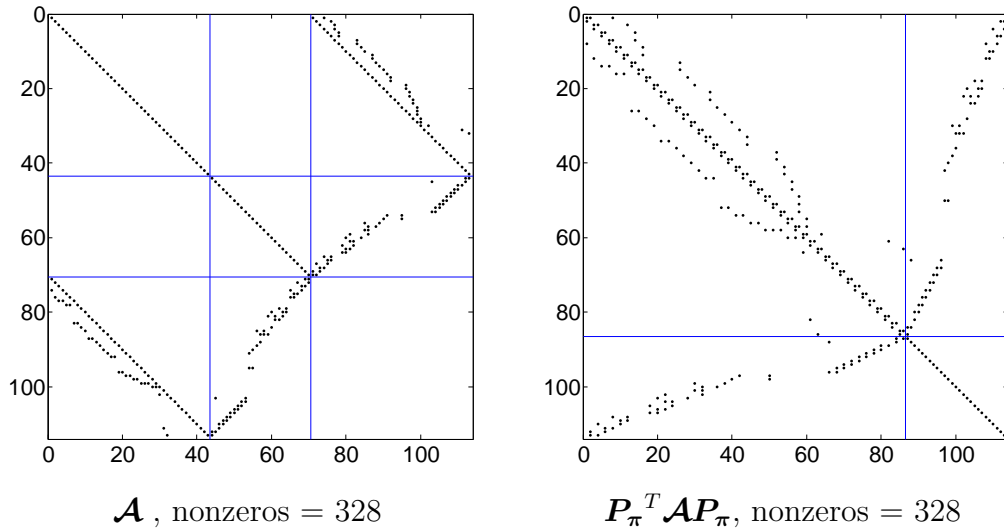
Figure 2: Transformed matrix $\boldsymbol{A}$ and the block partitioned matrix $\boldsymbol{P_\pi}^T \boldsymbol{A} \boldsymbol{P_\pi}$ with blocks of order 1 and 2.

20

$\boldsymbol{L_b}$, nonzeros = 352 $\boldsymbol{D_b} = \mathrm{diag}(\boldsymbol{L_b})$, nonzeros = 156
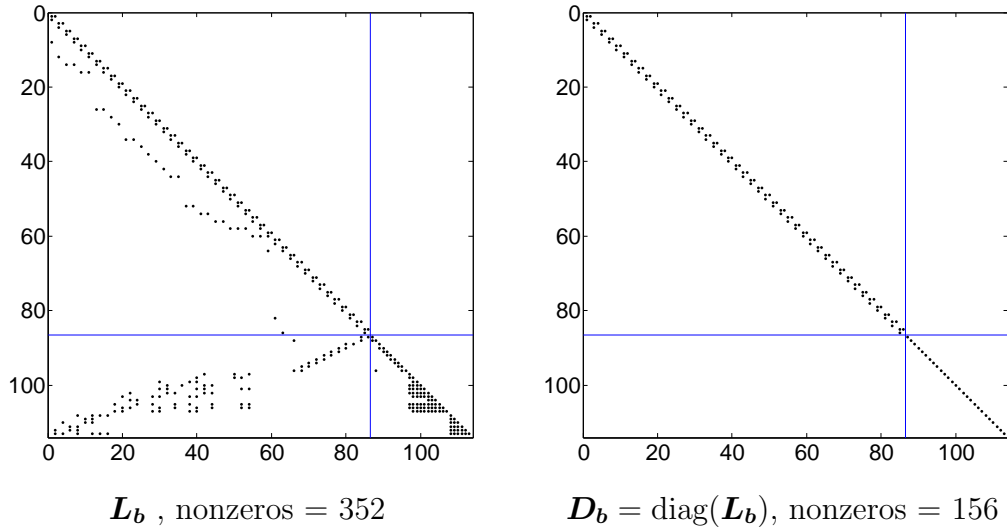
Figure 3: Factors of $\boldsymbol{P_\pi}^T \boldsymbol{\mathcal{A}} \boldsymbol{P_\pi}$—block lower triangular matrix $\boldsymbol{L_b}$ and its diagonal part with blocks of order 1 and 2.

$\boldsymbol{\mathcal{L}}$, nonzeros = 352 $\boldsymbol{\tilde{\mathcal{L}}}$, nonzeros = 696
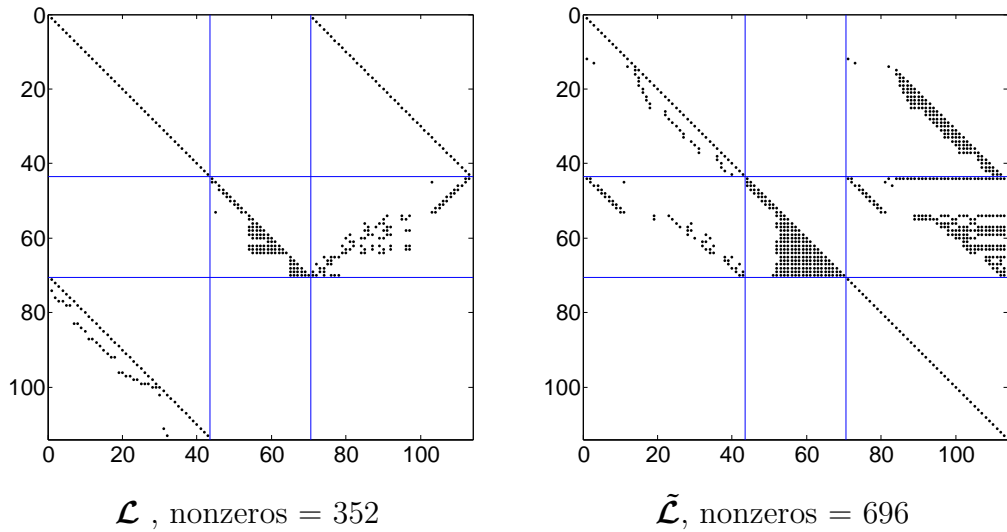
Figure 4: Induced block factor $\boldsymbol{\mathcal{L}}$ and the Schilders' factor $\boldsymbol{\tilde{\mathcal{L}}}$.

6 Conclusion

Symmetric indefinite matrices arising from saddle point problems can be congruently transformed into a special form of a block 3 by 3 structure. The transformed matrices can be exploited efficiently by taking privilege on the structures and properties of their blocks. We defined a transformation operator $\boldsymbol{\mathcal{T}}$ such that the constraint matrix $\boldsymbol{\mathring{B}}$ is transformed into a lower trapezoidal form, while the block matrix $\boldsymbol{\mathring{A}}$ is permuted only. The transformed saddle point matrix $\boldsymbol{\mathcal{A}}$ is partitioned into a block n by n structure
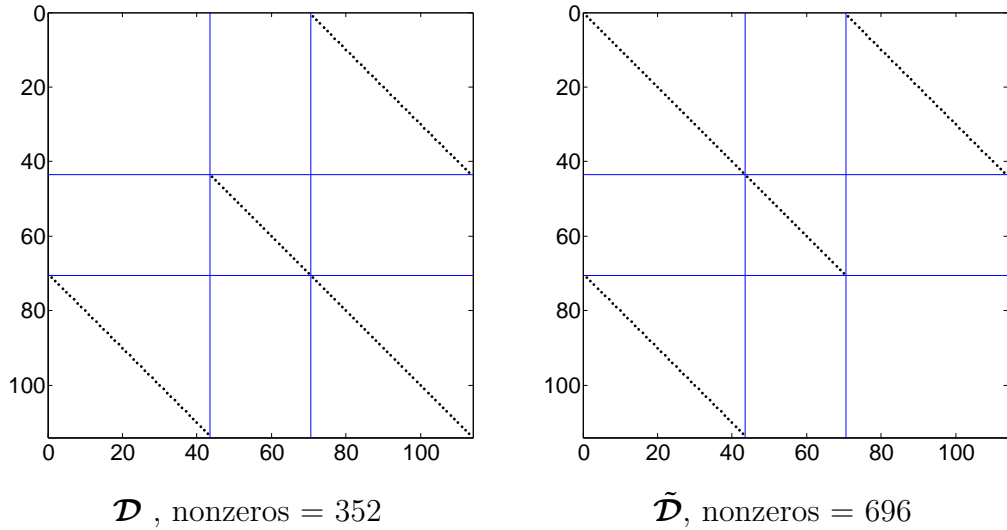
21

$\boldsymbol{\mathcal{D}}$, nonzeros = 352 $\tilde{\boldsymbol{\mathcal{D}}}$, nonzeros = 696

Figure 5: Induced block diagonal factor $\boldsymbol{\mathcal{D}}$ and the Schilders' block diagonal factor $\tilde{\boldsymbol{\mathcal{D}}}$.

network	size		nonzeros		
	n	m	$\boldsymbol{\mathcal{A}}$	$\tilde{\boldsymbol{\mathcal{L}}}$	$\boldsymbol{\mathcal{L}}$
RNB28	$1,751$	$1,058$	$8,549$	$31,137$	$13,936$
RNB27	$2,272$	$1,380$	$11,124$	$40,591$	$20,848$
RNB26	$2,430$	$1,474$	$11,934$	$50,845$	$22,240$
RNB1	$7,876$	$4,835$	$39,374$	$2,445,185$	$101,732$
RNC2	$18,762$	$11,782$	$93,804$	$12,446,959$	$192,646$
RNB5	$23,221$	$12,820$	$116,099$	$20,767,772$	$463,294$
RNB4	$25,593$	$16,052$	$127,959$	$18,212,250$	$456,245$
RNC1	$58,054$	$36,392$	$290,264$	$168,846,313$	$1,646,337$

Table 2: Nonzero counts of $\tilde{\boldsymbol{\mathcal{L}}}$ and $\boldsymbol{\mathcal{L}}$ of saddle point systems of realistic industrial resistor network problem.

with blocks of order 1 and 2 by applying a simple, predefined permutation π. Then a block $\boldsymbol{L_b}\boldsymbol{D_b}^{-1}\boldsymbol{L_b}^T$ factorization is applied to the block partitioned matrix $\boldsymbol{P_\pi}^T\boldsymbol{\mathcal{A}}\boldsymbol{P_\pi}$. The transformation operator $\boldsymbol{\mathcal{T}}$ is chosen for sparsity and stability of the factor $\boldsymbol{L_b}$, whereas the permutation π ensures a priori pivots for the factorization. We also formed a block 3 by 3 structured factorization with blocks of order m and $n-m$, which is induced from the block $\boldsymbol{L_b}\boldsymbol{D_b}^{-1}\boldsymbol{L_b}^T$ factorization. We compared the induced block factorization with the Schilders' factorization and found that it is different and has

matrix	size		nonzeros			
	n	m	$\mathring{\mathcal{A}}$	$\mathcal{A}$	$\tilde{\mathcal{L}}$	$\mathcal{L}$
ncvxqp1	$7,111$	$5,000$	$73,963$	$73,963$	$563,506$	$421,720$
tuma2	$7,515$	$5,477$	$49,365$	$49,365$	$326,722$	$245,945$
stokes64	$8,450$	$4,096$	$140,034$	$140,034$	$5,501,502$	$4,503,680$
ncvxqp9	$9,054$	$7,500$	$54,040$	$54,040$	$321,390$	$121,956$
tuma1	$13,360$	$9,607$	$87,760$	$87,760$	$713,623$	$535,932$
Qpband	$15,000$	$5,000$	$45,000$	$45,000$	$45,000$	$40,000$
mario001	$23,130$	$15,304$	$204,912$	$204,912$	$7,459,505$	$903,188$
aug3dcqp	$27,543$	$8,000$	$128,115$	$128,115$	$3,839,064$	$3,547,086$
blockqp1	$40,011$	$20,001$	$640,033$	$640,033$	$600,750,077$	$560,113$
mario002	$234,128$	$155,746$	$2,097,566$	$2,097,566$	$1,990,093,350$	$16,620,482$

Table 3: Nonzero counts of $\tilde{\mathcal{L}}$ and $\mathcal{L}$ of saddle point matrices from UF sparse matrix collections.

much more sparser block factors.

Appendix

$LD^{-1}L^T$ factorization for a symmetric, positive definite matrix

We show that $LD^{-1}L^T$ factorization for a symmetric, positive definite matrix $A \in \mathbb{R}^{n \times n}$ has the same computational efficiency with the Cholesky factorization, where L is a lower triangular matrix and $D = \mathrm{diag}(L)$ is the diagonal part of L. Since this factorization doesn't require to extract square roots like in the case of the Cholesky's, it is much more convenient for applying it to symmetric indefinite matrices. We give a rigorous proof of the existence and uniqueness of this factorization.

Theorem. *Let $A \in \mathbb{R}^{n \times n}$ be symmetric and positive definite matrix. Then there exists a unique nonsingular lower triangular matrix $L \in \mathbb{R}^{n \times n}$ such that*

$$A = LD^{-1}L^T \tag{38}$$

where $D = \mathrm{diag}(L)$ is the diagonal part of L.

Proof. We develop an algorithm for (38) and prove that the algorithm doesn't break down. Considering $A = [a_{ij}]$ of size 3×3, we have

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11}^{-1} & 0 & 0 \\ 0 & l_{22}^{-1} & 0 \\ 0 & 0 & l_{33}^{-1} \end{bmatrix} \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ 0 & l_{22} & l_{23} \\ 0 & 0 & l_{33} \end{bmatrix}.$$

Through direct computations, one can obtain

$$l_{11} = a_{11},$$

$$l_{21} = a_{21}, \quad l_{22} = a_{22} - \frac{l_{21}^2}{l_{11}},$$

$$l_{31} = a_{31}, \quad l_{32} = a_{32} - \frac{l_{31}l_{21}}{l_{11}}, \quad l_{33} = a_{33} - \frac{l_{31}^2}{l_{11}} - \frac{l_{32}^2}{l_{22}}.$$

For general n,

$$l_{ij} = a_{ij} - \sum_{k=1}^{j-1} \frac{l_{ik}l_{jk}}{l_{kk}} \ , \quad 1 \leq j \leq i \leq n \tag{39}$$

which exists if and only if the diagonal entries

$$l_{ii} = a_{ii} - \sum_{k=1}^{i-1} \frac{l_{ik}^2}{l_{kk}} \ , \quad 1 \leq i \leq n \tag{40}$$

are not equal to zero. This can be shown by induction on $i = j = 1, \ldots, n$. In the following part of the proof, let $L_{(j+1,j)} = [l_{j+1,j} \ , \ \ldots, \ l_{n,j}]^T$ denotes a jth column vector of $\boldsymbol{L}$ of length $n - j$.

For $i = j = 1$, $l_{11} = a_{11} > 0$, let $\boldsymbol{A}$ be partitioned as follows:

$$\boldsymbol{A} = \left[\begin{array}{c|c} a_{11} & \boldsymbol{A}_{21}^T \\ \hline \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{array} \right].$$

Let $\boldsymbol{0} \neq \boldsymbol{v}_1 \in \mathbb{R}^{n-1}$. Define $\boldsymbol{v} = \begin{bmatrix} v_0^T & \boldsymbol{v}_1^T \end{bmatrix}^T \in \mathbb{R}^n$ such that $v_0 = -\boldsymbol{A}_{21}^T \boldsymbol{v}_1 / a_{11}$. Then

$$0 < \boldsymbol{v}^T \boldsymbol{A} \, v = \begin{bmatrix} v_0^T & \boldsymbol{v}_1^T \end{bmatrix} \left[\begin{array}{c|c} a_{11} & \boldsymbol{A}_{21}^T \\ \hline \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{array} \right] \begin{bmatrix} v_0 \\ \boldsymbol{v}_1 \end{bmatrix}$$

$$= v_0^T a_{11} v_0 + v_0 \boldsymbol{A}_{21}^T \boldsymbol{v}_1 + \boldsymbol{v}_1^T \boldsymbol{A}_{21} v_0 + \boldsymbol{v}_1^T \boldsymbol{A}_{22} \boldsymbol{v}_1$$

$$= \frac{\boldsymbol{v}_1^T \boldsymbol{A}_{21}}{a_{11}} a_{11} \frac{\boldsymbol{A}_{21}^T \boldsymbol{v}_1}{a_{11}} - \frac{\boldsymbol{v}_1^T \boldsymbol{A}_{21}}{a_{11}} \boldsymbol{A}_{21}^T \boldsymbol{v}_1 - \frac{\boldsymbol{v}_1^T \boldsymbol{A}_{21}}{a_{11}} \boldsymbol{A}_{21}^T \boldsymbol{v}_1$$

$$= \boldsymbol{v}_1^T \boldsymbol{A}_{22} \boldsymbol{v}_1 - \boldsymbol{v}_1^T \frac{\boldsymbol{A}_{21} \boldsymbol{A}_{21}^T}{a_{11}} \boldsymbol{v}_1 = \boldsymbol{v}_1^T \underbrace{\left(\boldsymbol{A}_{22} - \frac{L_{(2,1)} L_{(2,1)}^T}{l_{11}} \right)}_{\boldsymbol{A}^{(2)}} \boldsymbol{v}_1. \tag{41}$$

(41) implies $\boldsymbol{A}^{(2)} \in \mathbb{R}^{(n-1) \times (n-1)}$ is symmetric and positive definite. The entries of $\boldsymbol{A}^{(2)}$ are:

$$a_{rs}^{(2)} = a_{ij} - \frac{l_{i1}l_{j1}}{l_{11}} \ , \quad 1 \leq r, s \leq n - 1, \quad \text{where} \ \ i = r + 1, \ j = s + 1.$$

Thus for $i = j = 2$,

$$0 < a_{11}^{(2)} = a_{22} - \frac{l_{21}^2}{l_{11}} = l_{22}.$$

24

Assume that (41) holds up to $i = j = n - 1$.

$$\text{i.e.,} \quad \boldsymbol{A}^{(n-1)} = \left(\boldsymbol{A}_{22}^{(n-2)} - \frac{L_{(n-1,n-2)} L_{(n-1,n-2)}^T}{l_{n-2,n-2}} \right) \in \mathbb{R}^{2 \times 2}$$

is symmetric and positive definite which has the entries:

$$a_{rs}^{(n-1)} = a_{ij} - \sum_{k=1}^{n-2} \frac{l_{ik} l_{jk}}{l_{kk}} , \quad 1 \leq r, s \leq 2 , \quad i = r + n - 2, \; j = s + n - 2$$

giving

$$a_{11}^{(n-1)} = a_{n-1,n-1} - \sum_{k=1}^{n-2} \frac{l_{n-1,k}^2}{l_{kk}} = l_{n-1,n-1} > 0.$$

For $i = j = n$, apply similar partitioning to $\boldsymbol{A}^{(n-1)}$. Define $\boldsymbol{u} = \begin{bmatrix} u_0^T & \boldsymbol{u}_1^T \end{bmatrix}^T \in \mathbb{R}^2$ such that $\boldsymbol{u}_1 \neq 0$ and $u_0 = \left(-\boldsymbol{A}_{21}^{(n-1)} \right)^T \boldsymbol{u}_1 / a_{11}^{(n-1)}$. Then

$$0 < \boldsymbol{u}^T \boldsymbol{A}^{(n-1)} \boldsymbol{u} = \begin{bmatrix} u_0^T & \boldsymbol{u}_1^T \end{bmatrix} \left[\begin{array}{c|c} a_{11}^{(n-1)} & (\boldsymbol{A}_{21}^{(n-1)})^T \\ \hline \boldsymbol{A}_{21}^{(n-1)} & \boldsymbol{A}_{22}^{(n-1)} \end{array} \right] \begin{bmatrix} u_0 \\ \boldsymbol{u}_1 \end{bmatrix} = \boldsymbol{u}_1^T \underbrace{\left(\boldsymbol{A}_{22}^{(n-1)} - \frac{L_{(n,n-1)} L_{(n,n-1)}^T}{l_{n-1,n-1}} \right)}_{\boldsymbol{A}^{(n)}} \boldsymbol{u}_1.$$

So, $\boldsymbol{A}^{(n)} \in \mathbb{R}^{1 \times 1}$ is symmetric and positive definite that gives

$$a_{11}^{(n)} = a_{nn} - \sum_{k=1}^{n-2} \frac{l_{n,k}^2}{l_{kk}} - \frac{l_{n,n-1}^2}{l_{n-1,n-1}} = a_{nn} - \sum_{k=1}^{n-1} \frac{l_{n,k}^2}{l_{kk}} = l_{nn} > 0.$$

For uniqueness, suppose there exist two nonsingular lower triangular matrices $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ satisfying (39) for the same $\boldsymbol{A}$. Then

$$\boldsymbol{L}_1 \boldsymbol{D}_1^{-1} \boldsymbol{L}_1^T = \boldsymbol{L}_2 \boldsymbol{D}_2^{-1} \boldsymbol{L}_2^T$$
$$\iff \boldsymbol{L}_2^{-1} \boldsymbol{L}_1 \boldsymbol{D}_1^{-1} = \boldsymbol{D}_2^{-1} \boldsymbol{L}_2^T \boldsymbol{L}_1^{-T}. \tag{42}$$

In (42), the left-hand side is a lower triangular matrix while the right-hand side is an upper triangular. This is possible only if $\boldsymbol{L}_2^{-1} \boldsymbol{L}_1$ is a diagonal matrix. Therefore, let $\boldsymbol{L}_2^{-1} \boldsymbol{L}_1 = \boldsymbol{D}$, where $\boldsymbol{D}$ is a diagonal matrix and hence $\boldsymbol{D}^{-T} = \boldsymbol{L}_2^T \boldsymbol{L}_1^{-T}$. Since the diagonal of product of any two upper triangular matrices is equal to the product of their diagonals, we get

$$\boldsymbol{D}_2^T \boldsymbol{D}_1^{-T} = \text{diag}(\boldsymbol{L}_2^T) \text{diag}(\boldsymbol{L}_1^{-T}) = \text{diag}(\boldsymbol{L_2}^T \boldsymbol{L}_1^{-T}) = \boldsymbol{D}^{-T}.$$

$\therefore$ from (42),

$$\boldsymbol{D} \boldsymbol{D}_1^{-1} = \boldsymbol{D}_2^{-1} \boldsymbol{D}^{-T}$$
$$\iff \boldsymbol{D} = \boldsymbol{D}_2^{-1} \boldsymbol{D}^{-T} \boldsymbol{D}_1$$
$$\iff \boldsymbol{D} = \boldsymbol{D}_2^{-1} \boldsymbol{D}_2^T \boldsymbol{D}_1^{-T} \boldsymbol{D}_1 = \boldsymbol{I}$$
$$\iff \boldsymbol{L}_1 = \boldsymbol{L}_2.$$

$\square$

Since $l_{kk} > 0$, (40) implies that $l_{kk} \leq a_{kk}$ for each $k = 1, \ldots, n$. Again from (40),

$$a_{ii} = \sum_{k=1}^{i} \frac{l_{ik}^2}{l_{kk}} \geq \frac{l_{ik}^2}{l_{kk}} \geq \frac{l_{ik}^2}{a_{kk}} \ , \ \text{ for } 1 \leq k \leq i \leq n$$

$$\Rightarrow \quad |l_{ij}| \leq \max_{i} a_{ii} \ , \ \text{ for } 1 \leq i, j \leq n.$$

From (39), observe that every time the jth column of $\boldsymbol{L}$ is updated, the quotient l_{jk}/l_{kk} is computed repeatedly for $(n - j + 1)(j - 1)$ times. This recurrence causes extra computational cost. Therefore by computing the quotients l_{jk}/l_{kk} for $k = 1, \ldots, j - 1$ and storing them in a vector $\boldsymbol{\nu}$ of length $j - 1$ before every next update, reduces the number of divisions to $j - 1$ as shown in the Step 3 of Algorithm 2. Step 6 involves $(n - j + 1)(j - 1)$ multiplications and $(n - j + 1)(j - 1)$ additions. Therefore, Algorithm 2 requires total flop counts precisely equal to

$$\sum_{j=1}^{n} (j - 1) + 2(n - j + 1)(j - 1) = \frac{1}{3}n^3 + \frac{1}{2}n^2 - \frac{5}{6}n,$$

which is equal to the flop counts of the Cholesky's factorization.

Algorithm 2 For a given symmetric and positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, the algorithm computes a nonsingular lower triangular matrix $\boldsymbol{L} \in \mathbb{R}^{n \times n}$ such that $\boldsymbol{A} = \boldsymbol{L} \operatorname{diag}^{-1}(\boldsymbol{L}) \boldsymbol{L}^T$. For $j = 1, \ldots, n$, the column $\boldsymbol{L}(j : n, j)$ overwrites $\boldsymbol{A}(j : n, j)$.

1: **for** $j = 1 : n$
2:     **for** $k = 1 : j - 1$
3:         $\boldsymbol{\nu}(k) = \boldsymbol{A}(j, k)/\boldsymbol{A}(k, k)$
4:     **end for**
5:     **for** $i = j : n$
6:         $\boldsymbol{A}(i, j) = \boldsymbol{A}(i, j) - \sum_{l=1}^{j-1} \boldsymbol{\nu}(l) \boldsymbol{A}(i, l)$
7:     **end for**
8: **end for**

References

[1] G. Al-Jeiroudi, J. Gondzio, J. Hall, *Preconditioning indefinite systems in interior point methods for large scale linear optimization*, Technical Report MS-2006-003, School of Mathematics, University of Edinburgh, 2007.

[2] M. Benzi, G. H. Golub and J. Liesen, *Numerical solution of saddle point problems*, Acta Numerica, vol. 14, pp. 1-137, 2005.

[3] M. Benzi and J. Liu, *Block preconditioning for saddle point systems with indefinite (1, 1) block*, International Journal of Computer Mathematics, Vol. 00, No. 05, 2007.

[4] J. R. Bunch and L. Kaufman, *Some stable methods for calculating inertia and solving symmetric linear systems*, Mathematics of Computation, Vol. 31, No. 137, pp. 163-179, 1977.

[5] J. R. Bunch and B. N. Parlett, *Direct methods for solving symmetric indefinite systems of linear equations*, SIAM Journal on Matrix Analysis and Applications, Vol. 8, pp. 639-655, 1971.

[6] J. R. Bunch, L. Kaufman, and B. N. Parlett, *Decomposition of a symmetric matrix*, Numerical Mathematics, Vol. 31, pp. 31-48, 1976.

[7] Z. H. Cao, *A class of constraint preconditioners for nonsymmetric saddlevpoint problems*, Numer. Math., vol. 103, pp. 47-61, 2006.

[8] sc T. A. Davis, "Direct Methods for Sparse Linear Systems", SIAM, 2006.

[9] A. C. De Niet and F. W. Wubs *Numerically stable LDL^T-factorization of $\mathcal{F}$-type saddle point matrices*, IMA Journal of Numerical Analysis Vol. 29, pp. 208-234, 2009.

[10] H. S. Dollar, *Iterative linear algebra for constrained optimization*, DPhil (PhD) thesis, Oxford University Computing Laboratory, 2005.

[11] H. S. Dollar, N. I. M. Gould, W. H. A. Schilders and A. J. Wathen, *Implicitfactorization preconditioning and iterative solvers for regularized saddle-point systems*, SIAM J. Matrix Anal. Appl., Vol. 28, pp. 170-189, 2006.

[12] H. S. Dollar, N. I. M. Gould, W. H. A. Schilders, and A. J. Wathen, *Using constraint preconditioners with regularized saddle-point problems*, Computational Optimization and Applications, Vol. 36, pp. 249-270, 2007.

[13] H. S. Dollar and A. .J. Wathen, *Approximate factorization constraint preconditioners for saddle-point matrices*, SIAM J. SCI. COMPUT., Vol. 27, No. 5, pp. 1555-1572.

[14] I. S. Duff and S. Parlet, *Strategies for scaling and pivoting for sparse symmetric indefinite problems*, SIAM J. MATRIX ANAL. APPL., Vol. 27, No. 2, pp. 313340, 2005.

[15] A. Forsgren, P. E. Gill, M. H. Wright, *Interior Methods for Nonlinear Optimization*, SIAM Review, Vol. 44, No. 4, pp. 525597, 2002.

[16] M. P. Friedlander, D. Orban, *A primaldual regularized interior-point method for convex quadratic programs*, Springer and Mathematical Optimization Society, Math. Prog. Comp., Vol. 4, pp. 71107, 2012.

[17] E. GALLIGANI AND L. Z. MODENA, *Error analysis of an algorithm for equality-constrained quadratic programming problems*, Springer-Verlag Computing Vol. 58, pp. 47-67, 1997.

[18] P. E. GILLT, W. MURRAY, D. B. PONCELEN, AND M. A. SAUNDERS, *Preconditioners for indefinite systems arising in optimization*, SIAM J. MATRIX ANAL. APPL., Vol. 13, No. 1, pp. 292-311, 1992.

[19] G. H. GOLUB AND C. GRIEF, *On solving block-structured indefinite linear systems*, SIAM J. Sci. Comput., Vol. 24, No. 6, pp. 2076-2092, 2003.

[20] G. H. GOLUB AND C. F. VAN LOAN, "Matrix Computations", 4th edition, The Johns Hopkins University Press, Baltimore, 2013.

[21] N. I. M. GOULD *On modified factorizations for large-scale linearly constrained optimization*, SIAM J. OPTIM., Vol. 9, No. 4, pp. 1041-1063, 1999.

[22] N. J. HIGHAM, "Accuracy and Stability of Numerical Algorithms", second edition, SIAM, 2002.

[23] C. KELLER, N. I. M. GOULD, AND A. J. WATHEN, *Constraint preconditioning for indefinite linear systems*, SIAM Journal on Matrix Analysis and Applications, Vol. 21, No. 4, pp. 1300-1317, 2000.

[24] L. KOMZSIK, P. POSCHMANN, I. SHARAPOV *A preconditioning technique for indefinite linear systems*, Finite Elements in Analysis and Design, Vol. 26, pp. 253-258, 1997.

[25] N. LI AND Y. SAAD, *Crout versions of the ILU factorization with pivoting for sparse symmetric matrices*, Elec. Trans. Numer. Anal., Vol. 21, pp. 75-85, 2005.

[26] Q. LIU, *New preconditioning techniques for saddle point problems arising from the time-harmonic Maxwell equations*, Hindawi Publishing Corporation, ISRN Applied Mathematics, 2013.

[27] J. M. L. MAUBACH, W. H. A. SCHILDERS, *Micro- and macro-block factorizations for regularized saddle point systems*, Technical report, Center for Analysis, Scientific computing and Applications, Eindhoven University of Technology, April 2012.

[28] M. F. MURPHY, G. H. GOLUB, AND A. J. WATHENS *A note on preconditioning for indefinite linear systems*, SIAM J. SCI. COMPUT., Vol. 21, No. 6, pp. 19691972, 2000.

[29] J. NOCEDAL AND S. WRIGHT, "Numerical Optimization", Springer-Verlag, NewY ork, 1999.

[30] J. Rommes and W. H. A. Schilders, *Efficient Methods for Large Resistor Networks*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 29, pp. 28-39, 2010.

[31] T. Rusten and R. Winther, *A preconditioned iterative method for saddlepoint problems*, SIAM Journal on Matrix Analysis and Applications, Vol. 13, pp. 887-904, 1992.

[32] M. A. Saunders and J. A. Tomlin, *Stable reduction to KKT systems in barrier methods for linear and quadratic programming*, Linear Algebra and Applications, 431, pp. 381-395, 2009.

[33] A. Shahzad, E. C. Kerrigan, and G. A. Constantinides *A stable and efficient method for solving a convex quadratic program with application to optimal control*, SIAM J. OPTIM. Vol. 22, No. 4, pp. 13691393, 2012.

[34] Q. Schenk and K. Gätner, *On fast factorization pivoting methods for sparse symmetric indefinite systems*, Electronic Transactions on Numerical Analysis, Vol. 23, pp. 158-179, 2006.

[35] W. H. A. Schilders, *Solution of indefinite linear systems using an LQ decomposition for the linear constraints*, Linear Algebra and Applications, 431, pp. 381-395, 2009.

[36] UF Sparse Matrix Collection: GHS_indef group, http://www.cise.ufl.edu/research/sparse/matrices/GHS_indef/index.html.

[37] G. W. Stewart, "Matrix Algorithms", Vol. 1: Basic Decompositions, SIAM, 1998.

[38] M. Tuma *A note on the LDL^T decomposition of matrices from saddle-point problems*, SIAM J. Matrix Anal. Appl., Vol. 23, pp. 903-915, 2002.

[39] S. J. Wright, "Primal-dual interior-point methods", SIAM, Philadelphia, PA, 1997.

[40] D. S. Watkins, "Fundamentals of Matrix Computations", second edition, John Wiley & Sons, Inc., New York, 2002.

[41] J. H. Wilkinson, *Error analysis of direct methods of matrix inversion*, J. Assoc. Comput. Mach. Vol. 8, pp. 281330, 1961.

PREVIOUS PUBLICATIONS IN THIS SERIES:

Number	Author(s)	Title	Month
14-19	O. Krehel A. Muntean	Error control for the FEM approximation of an upscaled thermo-diffusion system with Smoluchowski interactions	June '14
14-20	J.H.M. Evers R.C. Fetecau L. Ryzhik	Anisotropic interactions in a first-order aggregation model: a proof of concept	June '14
14-21	S.F. Nemadjieu	A stable and convergent O-method for general moving hypersurfaces	July '14
14-22	A. Corbetta L. Bruno A. Muntean F. Toschi	High statistics measurements of predestrian dynamics	July '14
14-23	S. Lungten W.H.A. Schilders J.M.L. Maubach	Sparse block factorization of saddle point matrices	Aug. '14