

# Computational Math Project

Illinois Institute of Technology

Miles Bakenhus

Ahmed Lodhika

Gunjan Sharma

Quinn Stratton

Jan-Eric Sulzbach

November 25, 2018

# Contents

<b>0</b>	<b>Introduction</b>	<b>2</b>
<b>1</b>	<b>Origin of the problem and its applications</b>	<b>2</b>
<b>2</b>	<b>Matrix Properties</b>	<b>5</b>
2.1	Tridiagonal Matrices . . . . .	5
2.2	General Banded Matrices . . . . .	6
2.3	Sparse Diagonal Matrices . . . . .	7
<b>3</b>	<b>Algorithms and Results</b>	<b>9</b>
3.1	General Banded Matrices . . . . .	9
<b>4</b>	<b>Further Study</b>	<b>10</b>
	<b>References</b>	<b>10</b>

## 0 Introduction

In this paper we discuss what motivates the study of banded matrices and show several results regarding their properties. We will also consider the application of these properties to find a more efficient implementation of the Modified Gram-Schmidt Algorithm. Finally, we will discuss potential future work.

## 1 Origin of the problem and its applications

In this part of the project we explain the motivation behind our ideas and examine the applications of our results. First, we consider the standard elliptic equation in two dimensions

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

Then the discretised equation on a grid is

$$\begin{aligned} -\Delta_h u_{ij} &= f_{ij} \quad \forall (x_i, y_i) \in \Omega_h, \quad f_{ij} = f(x_i, y_j) \\ u_{ij} &= 0 \quad \forall (x_i, y_i) \in \partial\Omega_h, \end{aligned}$$

where we use the second-order central differencing to represent the Laplace operator

$$-\Delta_h u_{ij} = \frac{1}{h^2} \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} u_{ij} = \frac{1}{h^2} (-u_{i,j+1} - u_{i,j-1} + 4u_{ij} - u_{i+1,j} - u_{i-1,j})$$

If we assume that the domain  $\Omega$  is a square, ordering the grid points from left to right and bottom to top yields the following matrix of the system  $A \in \mathbb{R}^{(h-1)^{-2} \times (h-1)^{-2}}$ :

$$A = h^2 \begin{pmatrix} 4 & -1 & 0 & \dots & 0 & -1 & & \\ -1 & 4 & -1 & & & & \ddots & \\ 0 & -1 & 4 & \ddots & & & & -1 \\ \vdots & & & \ddots & \ddots & & & 0 \\ 0 & & & & & & & \vdots \\ -1 & & & & & \ddots & \ddots & 0 \\ & \ddots & & & & \ddots & 4 & -1 \\ & & -1 & 0 & \dots & 0 & -1 & 4 \end{pmatrix}$$

And, the problem we need to solve is the linear system  $Au = f$ .

Therefore, it is important to understand how the structure of  $A$  affects the structure of the  $QR$  decomposition. Note that in this case the highest and lowest off-diagonal has a distance of order  $h$  from the diagonal.

Another example where these banded matrices show up is in the following: consider the parabolic equation in two dimensions

$$\frac{\partial u}{\partial t} = \sigma \Delta u, \quad 0 \leq x \leq X, \quad 0 \leq y \leq Y \quad 0 \leq t \leq T$$

with Dirichlet boundary condition and given initial data  $u(x, y, 0) = U^0(x, y)$ . For the numerical implementation we consider the implicit Crank-Nicolson scheme

$$\begin{aligned} & -\frac{\mu_x}{2}(U_{j-1,l}^{n+1} + U_{j+1,l}^{n+1}) - \frac{\mu_y}{2}(U_{j,l-1}^{n+1} + U_{j,l+1}^{n+1}) + (1 + \mu_x + \mu_y)U_{j,l}^{n+1} \\ & = \frac{\mu_x}{2}(U_{j-1,l}^n + U_{j+1,l}^n) + \frac{\mu_y}{2}(U_{j,l-1}^n + U_{j,l+1}^n) + (1 - \mu_x - \mu_y)U_{j,l}^n, \end{aligned}$$

for  $0 \leq j \leq J_x$ ,  $0 \leq l \leq J_y$  and  $n > 0$ . Again, we can rewrite this as a linear system  $AU^{n+1} = U^n$ , with

$$A = \begin{pmatrix} 1 + \mu_x + \mu_y & -\frac{\mu_x}{2} & 0 & \dots & 0 & -\frac{\mu_y}{2} & & \\ -\frac{\mu_x}{2} & 1 + \mu_x + \mu_y & -\frac{\mu_x}{2} & & & & \ddots & \\ 0 & -\frac{\mu_x}{2} & 1 + \mu_x + \mu_y & \ddots & & & & -\frac{\mu_y}{2} \\ \vdots & & & \ddots & \ddots & & & 0 \\ 0 & & & & & & & \vdots \\ -\frac{\mu_y}{2} & & & & & \ddots & \ddots & 0 \\ & & \ddots & & & \ddots & 1 + \mu_x + \mu_y & -\frac{\mu_x}{2} \\ & & & -\frac{\mu_y}{2} & 0 & \dots & 0 & -\frac{\mu_x}{2} & 1 + \mu_x + \mu_y \end{pmatrix}$$

and  $A \in \mathbb{R}^{(J_x-1)(J_y-1) \times (J_x-1)(J_y-1)}$  where the highest and lowest off-diagonal band have the distance  $J_x - 1$  from the diagonal.

*Remark 1.* In the case of three dimension, we would obtain one more non-zero sub/super-diagonal, now with a distance of  $\mathcal{O}(J_x * J_y)$  from the diagonal.

## 2 Matrix Properties

### 2.1 Tridiagonal Matrices

**Theorem 2.1.** *If  $A$  is a tridiagonal matrix, then  $R$  in the product  $A = QR$  is an upper triangular matrix with non-zero entries only in the diagonal and first two superdiagonals.*

$$A = \begin{pmatrix} a_{11} & a_{12} & & \\ a_{21} & a_{22} & a_{23} & \\ & \ddots & \ddots & \ddots \\ & & a_{mm-1} & a_{mm} \end{pmatrix}, \quad R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \\ & \ddots & \ddots & \ddots \\ & & \ddots & \ddots \\ & & & r_{mm} \end{pmatrix}$$

*Pf.* To prove the statement we will use the classical Gram-Schmidt (CGS) method for the QR decomposition.

Step 1: show that  $q_j$  has the form  $q_j = \begin{pmatrix} * \\ \vdots \\ * \\ 0 \\ \vdots \end{pmatrix} \leftarrow j + 1\text{-th entry}.$

We prove this by induction:

**Base step:** For  $j = 1$ , if we assume that  $\|a_1\| = 1$ , then  $q_1 = a_1$ . Thus

$$q_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ 0 \\ \vdots \end{pmatrix}$$

**Induction step:** Assume that the statement holds for  $j - 1$ . Then

$$v_j = a_j - \sum_{k=1}^{j-1} (q_k^* a_j) q_k \quad \text{and} \quad q_j = v_j / \|v_j\|$$

and by using the form of  $q_{j-1}$  we obtain

$$q_j = \begin{pmatrix} 0 \\ \vdots \\ a_{j-1,j} \\ a_{jj} \\ a_{j+1,j} \\ 0 \\ \vdots \end{pmatrix} - \sum_{k=1}^{j-1} \begin{pmatrix} * \\ \vdots \\ \vdots \\ * \\ 0 \\ \vdots \\ \vdots \end{pmatrix} \leftarrow k+1\text{-th entry} = \begin{pmatrix} * \\ \vdots \\ \vdots \\ \vdots \\ * \\ 0 \\ \vdots \\ \vdots \end{pmatrix} \leftarrow j+1\text{-th entry}$$

Step 2: compute  $r_{ij}$  in the CGS method.

For  $j$  from 1 to  $n$  and for  $i$  from 1 to  $j-1$ , we have  $r_{ij} = q_i^* a_j$ . Then by step 1, if  $i \leq j-3$ , we obtain  $r_{ij} = 0$  from the form of the vectors  $q_{j-3}$  and  $a_j$ :

$$0 = \begin{pmatrix} * \\ \vdots \\ * \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^* \begin{pmatrix} 0 \\ \vdots \\ 0 \\ * \\ * \\ * \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow j\text{-th entry}$$

Since, the above argument holds for all  $i \leq j-3$ ,  $R$  is non-zero only for the main diagonal and first two superdiagonals. ■

## 2.2 General Banded Matrices

**Theorem 2.2.** *If  $A$  is a banded matrix with bandwidth  $2p+1$ , then  $R$  in the orthogonalization  $A = QR$  is an upper triangular matrix with non-zero entries only in the main diagonal and first  $2p$  superdiagonals.*

*Pf.* If  $A$  has bandwidth  $2p + 1$  then for  $i - j > p$ ,

$$0 = a_{ij} = \sum_{k=1}^m q_{ik} r_{kj}$$

For  $k > j$ , since  $R$  is upper triangular,  $r_{kj} = 0$ . Then, when  $i > j + p$ ,

$$0 = a_{ij} = \sum_{k=1}^j q_{ik} r_{kj}$$

This gives  $q_{ij} = 0$ . Hence, for each  $j$

$$q_j = \begin{pmatrix} q_{1,j} \\ q_{2,j} \\ \vdots \\ q_{j+p,j} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (1)$$

From (1), when  $i + p < j - p$  (i.e.,  $j - i > 2p$ ):

$$r_{ij} = q_i^* a_j = \begin{pmatrix} q_{1,i} & q_{2,i} & \dots & q_{i+p,i} & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a_{j-p,j} \\ \vdots \\ a_{j+p,j} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0$$

Therefore,  $R$  is upper triangular with its only non-zero entries in the main diagonal and  $2p$  super-diagonals. ■

### 2.3 Sparse Diagonal Matrices

Now we want to generalize the ideas from Theorems 2.1 and 2.2 to the case where  $A$  still has only three non-zero bands, but the lower band is distance  $k - 1$  from the diagonal and the upper

band is distance  $l - 1$  from the diagonal. Consider the following example for  $A$ :

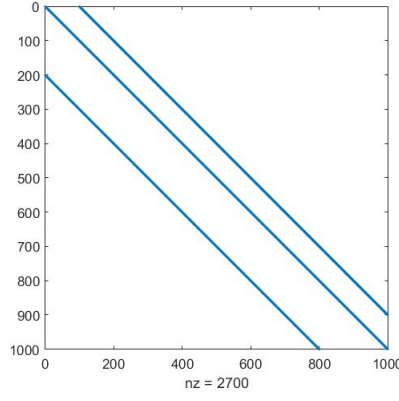
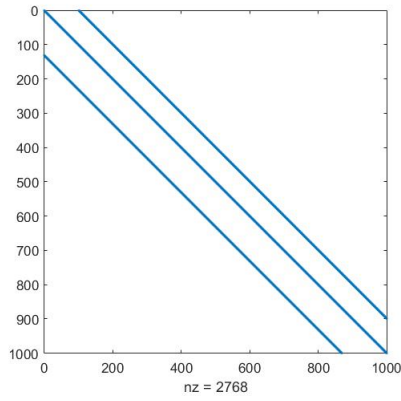


Figure 1:  $k = 200$  and  $l = 100$

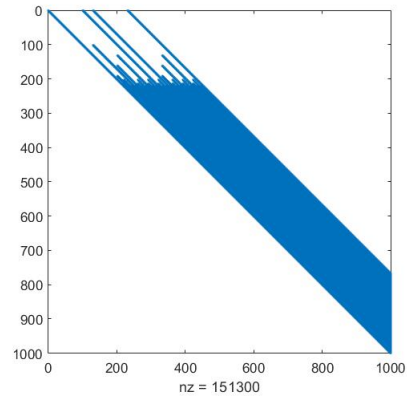
**Theorem 2.3** (General case). *The upper triangular matrix  $R$  in the  $QR$  decomposition of  $A$  has a  $k + l$ -band structure.*

*Pf.* From the CGS method we immediately see, in the worst case, the first  $j + k$  entries are non zero. Therefore, the inner product in the computation of the entries  $r_{ij}$  is only zero if  $i < j - l - k + 2$ . ■

To illustrate this, consider the example of a matrix close to the worst case, where the number of non-zero entries (nz) increases by order 50:



(a) A



(b) R

Figure 2:  $k = 131$  and  $l = 101$



A special case occurs when  $k = l$ . Again,  $R$  has only three non-zero bands: the main diagonal, the superdiagonal that has a distance  $k$  to the main diagonal, and the superdiagonal that has a distance  $2k$  to the main diagonal. For  $k = l = 100$ :

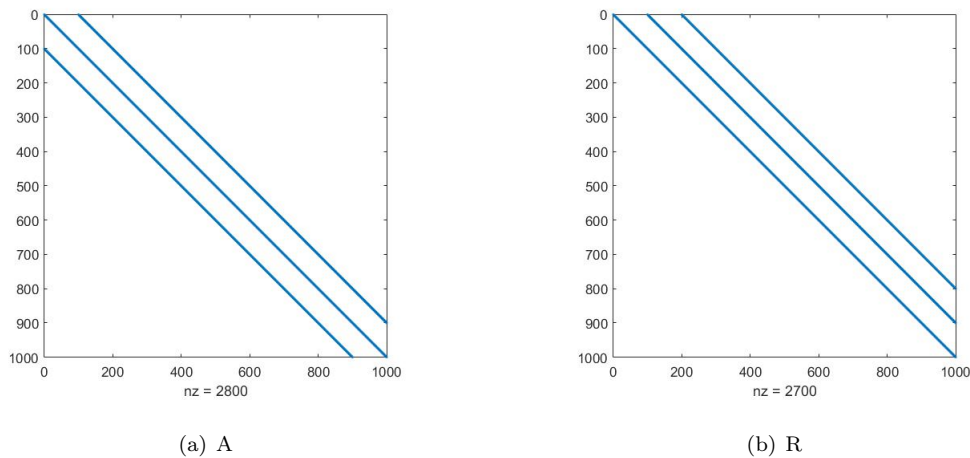


Figure 3:  $k = l = 100$

### 3 Algorithms and Results

Based on the properties proved in section 2, it is fairly straightforward to modify existing algorithms for finding the QR-factorization of a matrix, to exploit sparsity patterns.

#### 3.1 General Banded Matrices

Suppose  $A \in \mathbb{C}^{m \times n}$  with bandwidth  $2p + 1$ . Consider the QR-factorization of  $A$ ,  $A = QR$ . Then by Theorem 2.2, we know that if  $j > i + 2p$ ,  $r_{ij} = 0$ . We can alter the well-known *Modified Gram-Schmidt* (MGS) algorithm to take advantage of this fact, as seen in Algorithm 1. Below are some results for the performance of **MGS for Banded Matrices** and the performance of the standard MGS algorithm applied to the same random banded matrices.

Performance of *Modified Gram-Schmidt* **NOT COMPLETE!**

---

**Algorithm 1** MGS for Banded Matrices

---

```
1: for  $i = 1$  to  $n$  do  
2:    $r_{ii} \leftarrow \|\mathbf{a}_i\|_2$   
3:    $\mathbf{q}_i \leftarrow \mathbf{a}_i / r_{ii}$   
4:   for  $j = i + 1$  to  $\min\{i + 2p, n\}$  do  
5:      $r_{ij} \leftarrow \mathbf{q}_i^* \mathbf{a}_j$   
6:      $\mathbf{v}_j \leftarrow \mathbf{v}_j - r_{ij} \mathbf{q}_i$ 
```

[Note that this is based on the *Modified Gram-Schmidt* algorithm as described in [1]]

---

$\mathbb{C}^{10 \times 10}$	0.0024461	0.0023396	0.0023416	0.0023403	0.0023373	0.0023707
$\mathbb{C}^{500 \times 500}$	5.1399053	5.0569402	5.0526341	5.0476353	5.0430921	5.3553525
$\mathbb{C}^{750 \times 750}$	12.0286885	12.1555834	12.3704160	12.0967193	12.5868144	12.2757503
$\mathbb{C}^{1000 \times 1000}$	21.6673735	21.1107465	21.1824376	21.3300323	21.0178360	22.8876292

## 4 Further Study

There are several avenues for future work on this topic. Other algorithms, such as Householder triangularization, may be able to take advantage of the properties of banded matrices and address the poor stability of Modified Gram-Schmidt. Our report also never considers memory optimization with the use of a compressed storage format, such as Compressed Sparse Row (CSR) or diagonal form. Finally, further work may investigate the application of banded matrix properties to algorithmic parallelization of Modified Gram-Schmidt and other algorithms.

## References

- [1] Lloyd N. Trefethen and III David Bau. *Numerical Linear Algebra*. SIAM, Philadelphia, Pennsylvania, 1997.