**ECE 7650 (Advance Matrix Algorithm)**

**REPORT**

*On*

**“FINAL EXAMINATION FALL 2016”**

*By*

**Jamiu Babatunde Mojolagbe**

(Student ID: #7804719)

The Department of Electrical and Computer Engineering

University of Manitoba

*Submitted to*

**Ian Jeffrey, PhD**

(Course Instructor)

**PROBLEM 1**

**Question 1(i)**

This question was implemented in file named “**Q1i.m**”. Random matrices of chosen dimension were used and the following results were obtained:

**Case 1:** For randomly chosen ‘**m**’

|  |  |  |
| --- | --- | --- |
| Matrix Dimension | Size of Krylov Subspace (m) | Rank of V |
| 10x10 | 5 | 6 |
| 20x20 | 12 | 13 |
| 50x50 | 35 | 12 |
| 100x100 | 98 | 8 |
| 200x200 | 131 | 7 |
| 300x300 | 211 | 7 |
| 400x400 | 313 | 1 |

**Case 2:** For fixed ‘**m**’ for different dimensions of input matrix **A**

|  |  |  |
| --- | --- | --- |
| Matrix Dimension | Size of Krylov Subspace (m) | |
| **Rank(V) for m=5** | **Rank(V) for m=10** |
| 10x10 | 6 | 10 |
| 20x20 | 6 | 11 |
| 50x50 | 6 | 11 |
| 100x100 | 6 | 11 |
| 200x200 | 6 | 11 |
| 300x300 | 6 | 11 |
| 400x400 | 6 | 11 |

**Observation:**

As it can be observed from the tables, it is obvious that choosing **V = [r Ar A2r . . . Am-1r]** resulted in linearly dependent bases **V** such that, **V** is rank deficient. While **Case 1** presented above tends to be a generic case and not totally clear, however observing **Case 2** actually led to the conclusion that for a chosen size of Krylov subspace ‘**m**’, provided that **m < dim(A)**, **rank(V) = m+1** and for **m = dim(A)**, **rank(V)** = **m** for linearly independent bases.

**Note:** *The ranks obtained are more than ‘****m****’ since dimension of* ***V*** *is actually* ***nx(m+1)*** *but in a specific case* ***Vm*** *of dimension* ***nxm*** *can be obtained from it by just removing the last column.*

**Question 1(ii)**

Arnoldi method was implemented based on Modified Gram-Schmidt process as a function called ‘**arnoldi.m**’. The driver program for this question is named ‘**Q1ii.m**’. For the sake of clarity, ‘**Raw V**’ used to refer to the bases **V** obtained from **V = [r Ar A2r . . . Am-1r]** while the ones obtained from Arnoldi iteration is termed ‘**Arnoldi V**’.

**Case 1:** For randomly chosen ‘**m**’

|  |  |  |  |
| --- | --- | --- | --- |
| Matrix Dimension | Size of Krylov Subspace (m) | Rank of | |
| **Raw V** | **Arnoldi V** |
| 10x10 | 9 | 10 | 10 |
| 20x20 | 11 | 12 | 12 |
| 50x50 | 41 | 13 | 42 |
| 100x100 | 78 | 11 | 79 |
| 200x200 | 153 | 8 | 154 |
| 300x300 | 240 | 4 | 241 |
| 400x400 | 297 | 2 | 298 |

**Case 2:** For fixed ‘**m**’ for different dimensions of input matrix **A**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Matrix Dimension | Rank | | | |
| **V for m=5** | | **V for m=10** | |
| **Raw V** | **Arnoldi V** | **Raw V** | **Arnoldi V** |
| 10x10 | 6 | 6 | 10 | 10 |
| 20x20 | 6 | 6 | 11 | 11 |
| 50x50 | 6 | 6 | 11 | 11 |
| 100x100 | 6 | 6 | 11 | 11 |
| 200x200 | 6 | 6 | 11 | 11 |
| 300x300 | 6 | 6 | 11 | 11 |
| 400x400 | 6 | 6 | 11 | 11 |

**Observation:**

From the table shown above for **Case 1**, it can be observed that using Arnoldi method, the bases **V** obtained are linearly independent such the rank of **V** are now **m+1** (for the reasons previously presented). If **Vm** is considered, that is **V** of dimension **nxm**, then **rank(V) = m**.

Hence forth, the table in **Case 2** perfectly confirm the assertion made in the “**Observation**” section of the previous question, that is **Question 1(i)** above for **Case 2**.

**Note:** *The ranks obtained are more than ‘****m****’ since dimension of* ***V*** *is actually* ***nx(m+1)*** *but in a specific case* ***Vm*** *of dimension* ***nxm*** *can be obtained from it by just removing the last column.*

**PROBLEM 2**

**Question 2(i)**

While it is necessary to say that Arnoldi process in itself does not directly access the entries of a given matrix but instead makes the matrix map vectors and as such reaches its conclusions from their images. However, it is important to show rigorously or partially rigorously that for **VmTAVm = Hm**, when **m=n** that the eigenvalues of **Hm = Hn**are equal to the eigenvalues of **A**.

Recall that square matrices say, A and B, are related by:

**B = T-1AT = TTAT**

where **T-1** => Non-singular matrix (that is, invertible)

The transformation represented by **T-1AT** above is known as similarity transformation or conjugation by matrix **T**.  
Suppose:

**A** = 0 0 **B** = 0 1

1 0 0 0

These two matrices are similar under transformation or conjugation by **T**, such that **T** is given by

**T** = 0 1

1 0

Since a linear transformation is the same as a matrix after a basis, says bi, is chosen; then it can be shown that:

**L**(Ʃλibi) **=** Ʃajiλibj

Now, changing the basis, changes the co-efficient of the matrix such that

**L**(Ʃγiei) **=** Ʃajiγiej

If operator **L**(V) = AV uses standard basis (Euclidean basis), then L is the matrix **TAT-1** with basis

**bi = Tei**

**Note**: *Notice how the eigenvalues are represented to depict that they are images of each other*.

**Question 2(ii)**

To proof that for the largest eigenvalue of **A** and **Hm** are (approximately) equal, a file named “**Q2ii.m**” was used. Random matrices were used as well as random initial vector and **Hm** was calculate from Arnoldi method previously implemented. The following results were obtained:

**Case 1:** For randomly chosen ‘**m**’

|  |  |  |  |
| --- | --- | --- | --- |
| Matrix Dimension | Size of Krylov Subspace (m) | Maximum Eigenvalue of | |
| **A** | **Hm** |
| 10x10 | 8 | -3.9721 | -3.9752 |
| 20x20 | 18 | -3.735+1.5751i | -3.7347+1.5762i |
| 50x50 | 33 | 8.3546 | 8.3544 |
| 100x100 | 93 | -2.03653+9.94841i | -2.03653+9.94841i |
| 200x200 | 171 | -2.82874+15.0757i | -2.82874+15.0757i |
| 300x300 | 266 | -4.39007+16.9614i | -4.39007+16.9614i |
| 400x400 | 333 | 19.5548+6.06034i | 19.5548+6.06034i |

**Case 2:** For fixed ‘**m**’ for different dimensions of input matrix **A**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Matrix Dimension | Maximum Eigenvalue | | | |
| **m=5** | | **m=10** | |
| **A** | **Hm** | **A** | **Hm** |
| 10x10 | 2.1836+2.2106i | -0.99642+2.7697i | -2.6912 | -2.6912 |
| 20x20 | 5.1404 | 4.9816 | 4.3142+1.3456i | -4.4952 |
| 50x50 | -8.6843 | -8.8097 | -2.5343+7.2262i | -5.2299+5.1851i |
| 100x100 | -8.62056+5.87658i | -6.8939 | 9.28855+4.45615i | 8.4777+4.5603i |
| 200x200 | 15.1315+2.55873i | -7.7372 | -8.25263+12.1035i | 8.94859+10.3865i |
| 300x300 | -9.26965+15.9366i | -10.9218+7.77496i | 18.424 | -8.13645+11.1553i |
| 400x400 | -7.43289+18.9942i | -9.83433+6.59414i | 20.7206 | -10.4521+12.1804i |

**Observation:**

Though **Case 1** uses randomly chosen ‘**m**’ but it really does so such that it is very useful for arriving at a very good conclusion here. Why? Because the randomly chosen ‘**m**’s in **Case 1** above tend to be less that the dimension of a but **are not much less**, which is good for our purpose here. Therefore, from table for **Case 1**, it can be concluded that when **m < n** but not much less, that is when the value of **m** approaches the dimension **n**, then the eigenvalues of **Hm** tend to mimic eigenvalues of **A** - and in fact some cases in **Case 1** above the eigenvalue of **A** is equal to eigenvalue of **Hm** for that given precision.

Observing **Case 2** above, however, clear things up and it can be concluded that when **m<<n** then the eigenvalues (maximums are used here) of **Hm** are not in any way approximately equals to eigenvalues of **A**.

**Question 2(iii)**

Give a matrix **Anxn**, show that when **m < n, A = VmHmV**m**T** is true or false.

Recall from Arnoldi that:

**Hm = VmTAmVm** ----------------------- (i)

where **Vm** is **nxm**

**Hm** is **mxm**

A is **nxn**

As given in the question:

**A = VmHmVmT**  ------------------------ (ii)

Put (i) into (ii)

**A =** (**VmVmT)A(VmVmT)**

When **n=m**,

then **VmVmT** = **VmTVm**= Identity matrix (**I**) => unitary matrix

therefore, **A = IAI = A**

When **n < m**

A = (**VmVmT)A(VmVmT)**

**VmV**m**T** ≠ I

thus; **A** ≠ **VmVmTAmVmVmT**

Finally, it can be concluded that for the inequality n < m, statement **A = VmHmV**m**T** is not true (that is false) for given real matrix **Anxn**.

**PROBLEM 3**

**Question 3(i)**

Deriving the directional derivative of cost functional given below in terms of the arbitrary direction vector **h** using the real scalar parameter **ɛ**:

Recall that the directional derivative is given by:

Now,

since

If **h = r**, then

**Question 3(ii)**

For step length **,** direction **d**, evaluate functional

Recall, r = b - Ax

Given that, d = r

Now, derivative of the function with respect to is equal to 0

Comparing the result obtain for above with the alpha in **Minimum Residual 1-D**, shows that both has the same formulas

**Question 3(iii)**

From 3(i) above,

Such that,

Therefore:

since **r = b – Ax**

since **d =**

The derivative of the function with respect to is equal to 0

Finally, the obtained result is identical to alpha in **Residual Norm Steepest Descent** algorithm.

**PROBLEM 4**

**Question 4(i)**

Recall from Arnoldi that,

**Hm = VmTAmVm**

where **Vm** is **nxm**

**Hm** is **mxm**

A is **nxn** and symmetric

Let’s consider the (**ij**)th entry of matrix **Hm**

Recall that, orthonormal basis for

Thus;

Hence, = 0 for **i > j + 1**

Since,

Such that,

Since matrix **A** given is symmetric, then we have symmetric **Hm**, thus as a result of the reason presented/proven above.

Therefore, it can be concluded that for a symmetric matrix **A**, **Hm** produce from Arnoldi process is symmetric tridiagonal.

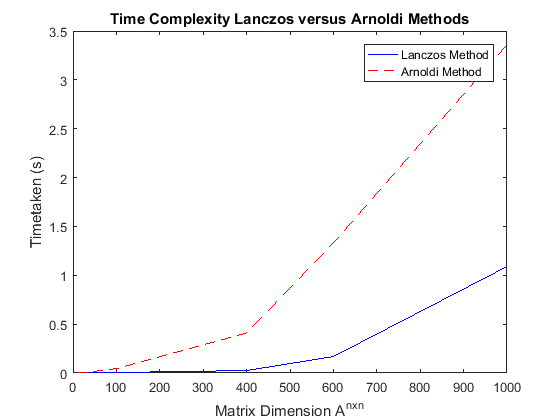
**Other Details**: For the purpose of confirmation and personal conviction, this proof was tested in file named “**Q4i.m**” for random input matrix, symmetric tridiagonal **Hm** was obtained.

**Question 4(ii)**

Lanczos method was implemented in function named “**lanczos.m**”. The driver program for this question is named “**Q4ii.m**”. The following results were obtained for different cases tested.

**Case 1:** For randomly chosen ‘**m**’

|  |  |  |  |
| --- | --- | --- | --- |
| Matrix Dimension | Size of Krylov Subspace (m) | Time taken in seconds | |
| **Arnoldi Method** | **Lanczos Method** |
| 10x10 | 8 | 0.00158832017805 | 0.00113115987732 |
| 20x20 | 11 | 0.00084607676844 | 0.000893040266430 |
| 100x100 | 76 | 0.04512201520370 | 0.004989871661235 |
| 400x400 | 210 | 0.40976936152758 | 0.025197017377595 |
| 600x600 | 490 | 1.33099212957816 | 0.168348730389529 |
| 1000x1000 | 694 | 3.36032083407172 | 1.089681569138520 |



Prove that the algorithm worked was also performed and the result obtained from the various norms were as follows:

============= A[10x10] when m = 8 ====================================

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Proofs of Orthogonality of Resulting Bases \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Norm of Km from Arnoldi: 1

Norm of Km from Lanczos: 1

Error: norm(Lanczos(Km) - Arnoldi(Km)): 2.0147e-09

Inner Product (Arnoldi): <v(m), v(m)>: 1

Inner Product (Arnoldi): <v(m), v(m-1)>: -1.5404e-15

Inner Product (Arnoldi): <v(2), v(m)>: -5.3257e-12

Inner Product (Lanczos): <v(m), v(m)>: 1

Inner Product (Lanczos): <v1(m), v1(m-1)>: -1.3878e-17

Inner Product (Lanczos): <v1(2), v1(m)>: 6.4199e-14

=============================================================================

============= A[20x20] when m = 11 ===============

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Proofs of Orthogonality of Resulting Bases \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Norm of Km from Arnoldi: 1

Norm of Km from Lanczos: 1

Error: norm(Lanczos(Km) - Arnoldi(Km)): 2.8877e-06

Inner Product (Arnoldi): <v(m), v(m)>: 1

Inner Product (Arnoldi): <v(m), v(m-1)>: -7.7716e-16

Inner Product (Arnoldi): <v(2), v(m)>: 1.6214e-08

Inner Product (Lanczos): <v(m), v(m)>: 1

Inner Product (Lanczos): <v1(m), v1(m-1)>: 4.3021e-16

Inner Product (Lanczos): <v1(2), v1(m)>: -1.3128e-14

=============================================================================

============= A[100x100] when m = 76 ===============

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Proofs of Orthogonality of Resulting Bases \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Norm of Km from Arnoldi: 1.5959

Norm of Km from Lanczos: 1.4142

Error: norm(Lanczos(Km) - Arnoldi(Km)): 2.18

Inner Product (Arnoldi): <v(m), v(m)>: 1

Inner Product (Arnoldi): <v(m), v(m-1)>: 7.9353e-15

Inner Product (Arnoldi): <v(2), v(m)>: 4.1302e-06

Inner Product (Lanczos): <v(m), v(m)>: 1

Inner Product (Lanczos): <v1(m), v1(m-1)>: 8.2391e-15

Inner Product (Lanczos): <v1(2), v1(m)>: -0.0012023

=============================================================================

============= A[400x400] when m = 210 ===============

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Proofs of Orthogonality of Resulting Bases \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Norm of Km from Arnoldi: 1.8428

Norm of Km from Lanczos: 1.7321

Error: norm(Lanczos(Km) - Arnoldi(Km)): 2.5176

Inner Product (Arnoldi): <v(m), v(m)>: 1

Inner Product (Arnoldi): <v(m), v(m-1)>: 6.7521e-16

Inner Product (Arnoldi): <v(2), v(m)>: 0.046588

Inner Product (Lanczos): <v(m), v(m)>: 1

Inner Product (Lanczos): <v1(m), v1(m-1)>: -2.8027e-14

Inner Product (Lanczos): <v1(2), v1(m)>: 0.012046

=============================================================================

============= A[600x600] when m = 490 ===============

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Proofs of Orthogonality of Resulting Bases \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Norm of Km from Arnoldi: 2.0865

Norm of Km from Lanczos: 2.4495

Error: norm(Lanczos(Km) - Arnoldi(Km)): 3.0614

Inner Product (Arnoldi): <v(m), v(m)>: 1

Inner Product (Arnoldi): <v(m), v(m-1)>: 1.1645e-13

Inner Product (Arnoldi): <v(2), v(m)>: 3.639e-08

Inner Product (Lanczos): <v(m), v(m)>: 1

Inner Product (Lanczos): <v1(m), v1(m-1)>: -2.8271e-14

Inner Product (Lanczos): <v1(2), v1(m)>: -0.022383

=============================================================================

============= A[1000x1000] when m = 694 ===============

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Proofs of Orthogonality of Resulting Bases \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Norm of Km from Arnoldi: 1.9411

Norm of Km from Lanczos: 2.6451

Error: norm(Lanczos(Km) - Arnoldi(Km)): 3.1447

Inner Product (Arnoldi): <v(m), v(m)>: 1

Inner Product (Arnoldi): <v(m), v(m-1)>: -1.2157e-13

Inner Product (Arnoldi): <v(2), v(m)>: -2.7093e-07

Inner Product (Lanczos): <v(m), v(m)>: 1

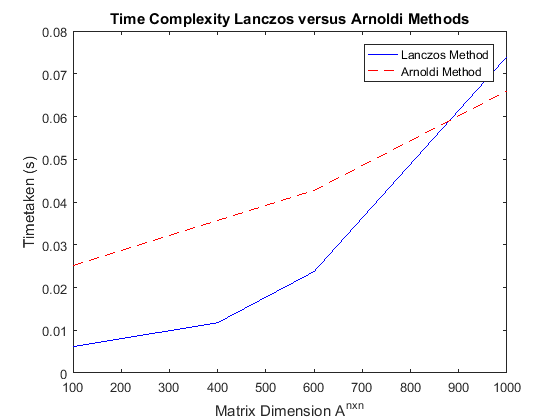
Inner Product (Lanczos): <v1(m), v1(m-1)>: -1.7757e-14

Inner Product (Lanczos): <v1(2), v1(m)>: -0.034224

=============================================================================

**Case 2:** For fixed ‘**m**’ and chosen to be 50, the following time complexity results were obtained:

|  |  |  |  |
| --- | --- | --- | --- |
| Matrix Dimension | Size of Krylov Subspace (m) | Time taken in seconds | |
| **Arnoldi Method** | **Lanczos Method** |
| 100x100 | 50 | 0.0251170326700840 | 0.00617166405982563 |
| 400x400 | 50 | 0.0356834528150765 | 0.0117577520041122 |
| 600x600 | 50 | 0.0427276106109621 | 0.0237334439910755 |
| 1000x1000 | 50 | 0.0659936210361248 | 0.0740232784851043 |



**Observation:**

From the foregoing, it can be observed from the results and plots that the implementation worked and the error tends to be very small. However, a close look at the results showed also that the bases produced are truly orthogonal but there seems to be loss of orthogonality when comparing says 1st and 500th vector as the error obtained, for **Case 1** for matrix dimensions [**600x600** **1000x1000]** for **m = [690 694],** clearly shown.

Moreover, comparing the complexity, it can be seen from the graphs and tables presented that in overall Lanczos method is fantastically faster than Arnoldi. However, when the **m** is fixed at **50** Arnoldi tends to be faster. Why? This misery was uncovered when a new test was performed and the size of Krylov subspace **m = 500** and matrix dimension is **1000x1000**. The time obtained was **2.6839** for Arnoldi and **0.79408** for Lanczos; this thereby led to the conclusion that, when **m** is chosen to **m << n** there is possibility that Arnoldi would be faster than Lanczos.

**Question 4(iii)**

Recall,

**Hm = VmTAmVm**

For symmetric trdiagonal let Hm = Tm

Such that,

---------------- (iii)

Using LU decomposition,

Where Lm = Unit lower bi-diagonal

Um = Upper bi-diagonal matrix

equation (iii) becomes:

let

Therefore,

The last column of matrix Pm is given by:

where

Recall that,

Finally, update equation is given by:

**Implementation Note**

The proved algorithm above was implemented in the file named “**lanczosfom.m**” as a function. The driver program for this function is named “**Q4iii.m**”. As contained in the driver program, random matrices can be used or an already loaded matrix and right hand side obtained from Finite Element Method solution to a particular problem. Again, for clear comparison, Restarted FOM was implemented to serve as basis for our comparison here; meanwhile the two implementation are compared to the solution obtained from MATLAB’s backslash operator which now serves as analytic solution for validation and comparison purposes. While the input matrix dimension is fixed, 157x157, random matrix generation provided a good alternative, though the former is a real time data.

**Results and Discussions**

The following results were obtained when the implemented function was compared against Restarted Full Orthogonalization Method (FOM):

**Case 1**: For Krylov subspace, **m = 20** and error tolerance, **tol = 1e-06**

Where **x** = solution from **lanczosfom.m**, **x1** = solution from restarted FOM, **xe** = A\b

**LLU** is used here to refer to the solution we obtained based on LU

**RFOM** is used here to refer to the solution obtained from restarted FOM

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Matrix Dimension | Time taken (s) | | Proof that it works | | Number of iterations | |
| **LLU** | **RFOM** | norm(x-xe) | norm(x1-xe) | **LLU** | **RFOM** |
| 50x50 | 0.0179180 | 0.0085334 | 6.7281e-08 | 6.7281e-08 | 8 | 8 |
| 100x100 | 0.0025287 | 0.0036320 | 9.048e-08 | 9.048e-08 | 7 | 7 |
| 200x200 | 0.0026633 | 0.0036283 | 1.2478e-08 | 1.2478e-08 | 7 | 7 |
| 500x500 | 0.005898 | 0.005949 | 2.1109e-08 | 2.1109e-08 | 6 | 6 |

**Case 2**: For Krylov subspace, **m = 20** and error tolerance, **tol = 1e-10**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Matrix Dimension | Time taken (s) | | Proof that it works | | Number of iterations | |
| **LLU** | **RFOM** | norm(x-xe) | norm(x1-xe) | **LLU** | **RFOM** |
| 50x50 | 0.0015623 | 0.0032273 | 2.7607e-12 | 2.7608e-12 | 12 | 12 |
| 100x100 | 0.0031587 | 0.0032486 | 2.5796e-12 | 2.5798e-12 | 11 | 11 |
| 200x200 | 0.0030508 | 0.0037120 | 9.0409e-13 | 9.0451e-13 | 10 | 10 |
| 500x500 | 0.0033564 | 0.0027470 | 7.0452e-13 | 7.0452e-13 | 9 | 9 |

**Case 3**: For Krylov subspace, **m = 50** and error tolerance, **tol = 1e-6**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Matrix Dimension | Time taken (s) | | Proof that it works | | Number of iterations | |
| **LLU** | **RFOM** | norm(x-xe) | norm(x1-xe) | **LLU** | **RFOM** |
| 50x50 | 0.0015847 | 0.0019662 | 6.3966e-08 | 6.3966e-08 | 8 | 8 |
| 100x100 | 0.0023119 | 0.0034027 | 8.8003e-08 | 8.8003e-08 | 7 | 7 |
| 200x200 | 0.0025459 | 0.0024315 | 8.7779e-09 | 8.7779e-09 | 7 | 7 |
| 500x500 | 0.0043801 | 0.0046013 | 1.9508e-08 | 1.9508e-08 | 6 | 6 |

**Observation:**

It can be observed from the table that the implemented algorithm worked. It can be seen result wise that increasing the size of Krylov subspace **m** decreases the number of iterations and vice versa. It can also be said that, setting error tolerance very low increases the number of iterations.

Regarding comparison with restarted FOM, it can be observed that this approach tends to reduce the computational time when compared to the restarted FOM as shown in the table above. However, regard the error obtained, there was also slight but not obvious improvement in the error, this is not really obvious because of the chosen precision, though, **Case 2** above somehow show this (for matrix dimension 100x100).

**PROBLEM 5**

**Question 5(i)**

Solving Ax = b

Such that: , that is, over determined system

It is what noting that over determined systems fall under least square problem.

***Why QR for this system?***

QR provides efficient way in solving non-square system such that Q factor provides orthogonal bases for the span of the columns of given matrix A.

***How do we do this?***

Soving as given means, minimizing the functional

It should be noted that transforming that difference by an orthogonal matrix preserve the norm (Euclidean norm).

Therefore, =

Which can be represented in block form as:



Since the second term above contains no ‘x’ then it is not relevant to the functional minimization.

Therefore, minimizing the first term is equivalent to minimizing the norm of the first term.

Given that to be non – singular ( once A is not singular itself)



Such that the solution, x, can be obtained from the above expression by applying backward substitution. Conclusively, QR factorization reduces the over determined system to a square one. It should also be stated that need not to be computed as the QR can be obtained more efficiently if only is calculated rather than obtaining all of the Q. This is what “economy size” QR in MATLAB does.

**Question 5(ii)**

Given:

where, W = AV and V = Vm

Equation (i) can be written as:

Let such that,

Thus: simplify

Recall from least-square that,

Therefore,

Since W = AV as given above, then,

Recall from Arnoldi,



Where **Hm**is **mxm**, **Vm**is **nxm**, **Hm+1**is (**m+1)xm** and **Vm+1**is **nx(m+1)**

This result can be said to be approximately and theoretically equal to the result of minimizing the functional which gives

Where and letting