# SENG 474 Assignment 2

Stephen Scinocca V00809689
Austin Beauchamp V00825419

1a)

0:
1: {3,4}, {3,4}, {2,6}, {3,4}, {3,4}
2: {1,2}, {4,6}, {4,6}, {1,2}, {4,6}
3: {1,3}, {1,3}, {1,3}
4: {3,5}, {3,5}, {1,4}, {3,5}, {1,4}, {3,5}
5: {1,5}
6: {2,3}, {2,3}, {2,3}
7: {3,6}, {3,6}
8: {2,4}, {5,6}, {2,4}, {2,4}, {5,6}, {2,4}
9: {4,5}, {4,5}, {4,5}
10: {2,5}, {2,5}

| Index | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CountMap | 0 | 5 | 5 | 3 | 6 | 1 | 3 | 2 | 6 | 3 | 2 |
| BitMap | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |

**Frequent buckets are 1,2,4,8**

1b)

| Number | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Support | 4 | 6 | 8 | 8 | 6 | 4 |
| L1 | 1 | 1 | 1 | 1 | 1 | 1 |

| Index | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CountMap | 0 | 5 | 5 | 3 | 6 | 1 | 3 | 2 | 6 | 3 | 2 |
| BitMap | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |

**Pairs counted: {3,4}, {3,5}, {2,4}**

2.

In the first pass we need to have approx S/4 buckets in the has table at max, as each bucket in the hash table stores an integer which is 4 bytes, or 4S. This is our Countmap. We can now compress this into a Bitmap, but the bitmap will still have S/4 buckets at max size, because it will have the same number of buckets as in the Countmap. This just makes it take less space for the second pass.

With S/4 buckets, we can compress that to S/32 bytes using the BitMap, and are left with S*31/32 bytes to count pairs

Assuming it takes 12 bytes to store a pair, the second pass will have at least 12 million bytes for the 1 million pairs we know about to count. Overall the memory needed for the second pass will be:

$$\frac{12{,}000{,}000\ * P}{\#\ Buckets} = \frac{12{,}000{,}000 * P}{\frac{S}{4}} = \frac{48{,}000{,}000 * P}{S}$$

But because we only have S*31/32 bytes free to count with, the formula becomes

$\frac{S*31}{32} = \frac{48{,}000{,}000 *P}{S}$

Which if we isolate for P we get:

$P < \frac{S^2}{49{,}548{,}387}$

If P is any bigger than we cannot successfully run the PCY algorithm

3a)

$$C = \begin{bmatrix} 2 & -1 \\ 2 & 1 \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad B = \begin{bmatrix} -1 & 2 \\ 3 & 1 \end{bmatrix}$$

$$CA = \begin{bmatrix} 2a_{11} - a_{21}, & 2a_{12} - a_{22} \\ 2a_{11} + a_{21}, & 2a_{12} + a_{22} \end{bmatrix}$$

$$CAB = \begin{bmatrix} -2a_{11} + a_{21} + 6a_{12} - 3a_{22}, & 4a_{11} - 2a_{21} + 2a_{12} - a_{22} \\ -2a_{11} - a_{21} + 6a_{12} + 3a_{22}, & 4a_{11} + 2a_{21} + 2a_{12} + a_{22} \end{bmatrix}$$

$$tr(CAB) = (-2a_{11} + a_{21} + 6a_{12} - 3a_{22}) + (4a_{11} + 2a_{21} + 2a_{12} + a_{22})$$

$$= 2a_{11} + 3a_{21} + 8a_{12} - 2a_{22}$$

$$f = tr(CAB)$$

3b)

$$\nabla_A f = \begin{bmatrix} \frac{df}{da_{11}} & \frac{df}{da_{12}} \\ \frac{df}{da_{21}} & \frac{df}{da_{22}} \end{bmatrix}$$

$$\frac{df}{da_{11}} = 2 \quad \frac{df}{da_{12}} = 3 \quad \frac{df}{da_{21}} = 8 \quad \frac{df}{da_{22}} = -2$$

$$\nabla_A f = \begin{bmatrix} 2 & 3 \\ 8 & -2 \end{bmatrix}$$

4a)

$$tr(AB) = \sum_{i=1}^{m} (AB)_{ii} = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} A_{ij} B_{ji} \right)$$

$$= \sum^{m} \sum^{n} A_{ij} B_{ji} = \sum^{n} \sum^{m} B_{ji} A_{ij}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ji} = \sum_{j=1}^{n} \sum_{i=1}^{m} B_{ji} A_{ij}$$

$$= \sum_{j=1}^{n} \left( \sum_{i=1}^{m} B_{ji} A_{ij} \right) = \sum_{j=1}^{n} (BA)_{jj} = tr(BA)$$

$$\therefore \quad tr(AB) = tr(BA)$$

4b)

$$tr(AB) = \sum_{i=1}^{m} a_{1i} b_{i1} + \sum_{i=1}^{m} a_{2i} b_{i2} + \cdots \sum_{i=1}^{m} a_{ni} b_{in}$$

$$\nabla_A tr(AB) = \frac{d \; tr(AB)}{d a_{ij}}$$

$$= \frac{d \sum_{i=1}^{m} a_{1i} b_{i1}}{d a_{ij}} + \frac{d \sum_{i=1}^{m} a_{2i} b_{i2}}{d a_{ij}} + \cdots \frac{d \sum_{i=1}^{m} a_{ni} b_{in}}{d a_{ij}}$$

$$= b_{ij}$$

$$\nabla_A tr(AB) = B^T$$

$$\therefore \quad \nabla_A tr(AB) = B^T$$