Assignment 2

1. Given the numbers 10, 13, 8, 5 and 8, what is the average of these five numbers? Given the weights 0.1, 0.4, 0.2, 0.2 and 0.1, what is the weighted average of these numbers.

$\frac{10+13+8+5+8}{5} = 8.8$ and $0.1 \cdot 10 + 0.4 \cdot 13 + 0.2 \cdot 8 + 0.2 \cdot 5 + 0.1 \cdot 8 = 9.6$

2. We discussed that a convex combination (a weighted average where all weights are between zero and one and sum up to one) must have the property that the weighted average must be some value between the minimum and maximum values being averaged. Suppose you got 40 on the mid-term examination and 60 on the final examination of this course. Find weights so that, with these grades, you a) get zero in this course, and b) get 100 in this course.

Clearly, these weights cannot represent a convex combination. Instead, we are looking values of α such that $40\alpha + 60\;\;1 - \alpha\;\; = 0$ and $40\alpha + 60\;\;1 - \alpha\;\; = 100$. Solving these, we get that $\alpha = 3$ and $\alpha = -2$, so the weighted averages are 3 and –2, and –2 and 3, respectively.

3. Suppose you had a sensor and you would like to record the average of the last *n* readings. One approach would be to have *n* memory locations that store the last *n* readings, and then calculate the sum of those *n* readings and dividing the result by *n*. If *n* was large, this would require significant effort with each step, and even if *n* is reasonably small (say, $n = 10$), it still would require processing power. Devise a scheme so that the average can be calculated with only a fixed number of arithmetic operations with each new reading, instead of *n* operations.

Note that if $s_n = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$ and $s_{n+1} = \frac{x_2 + x_3 + x_4 + \cdots + x_{n+1}}{5}$, we can find s_{n+1} from s_n using the calculation $s_{n+1} \leftarrow s_n - \frac{x_1}{n} + \frac{x_{n+1}}{n}$, which is an $\Theta(1)$ calculation. All we would have to do is store the *n* most recent entries in an array.

4. We saw how we can approximate the square root of two by iterating $x/2 + 1/x$. Demonstrate that this is true by showing that the square root of two is a solution to $x/2 + 1/x = x$.

$\frac{\sqrt{2}}{2} + \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} = \frac{2\sqrt{2}}{2} = \sqrt{2}.$

5. Given the value *n*, what does iterating $x/2 + n/(2x)$ appear to converge to? Prove this by showing that the solution you propose is indeed a solution to $x/2 + n/(2x) = x$.

If we try this with $n = 5$, then starting with $x_0 = 5$, we see it converges in seven iterations:

5.0, 3.0, 2.3333333333333333333, 2.2380952380952380952, 2.2360688956433637285, 2.2360679774999781940, 2.2360679774997896964, 2.2360679774997896964

and it seems to converge to $\sqrt{5}$, so it stands to reason that using an arbitrary n would see the sequence converge to $\sqrt{n}$. This is true, as $\frac{\sqrt{n}}{2} + \frac{n}{2\sqrt{n}} = \frac{\sqrt{n}}{2} + \frac{\sqrt{n}}{2} = \frac{2\sqrt{n}}{2} = \sqrt{n}.$

6. Suppose we wanted to solve $x^3 - x - 1 = 0$. Find some rewriting of this equation so that fixed-point iteration converges to the real root of this equation.

One rewriting is $x = \sqrt{\frac{x+1}{x}}$ and starting with $x_0 = 1.0$, after approximately 30 iterations, it converges to 1.3247179572447460260.

Another rewriting by ashen is $x = \sqrt[3]{x+1}$ and starting $x_0 = 1.0$, after approximately 24 iterations, it converges to 1.3247179572447460260.

7. Suppose we want to solve $x^3 - x^2 - x - 1 = 0$. Find some rewriting of this equation so that fixed-point iteration converges to the real root of this equation.

One rewriting is $x = \frac{x^2+x+1}{x^2}$ and starting with $x_0 = 1.0$, after approximately 100 iterations, it converges to 1.8392867552141611325.

Another rewriting by John Jekel is $x = \sqrt[3]{x^2+x+1}$ and starting $x_0 = 1.0$, after approximately 50 iterations, it converges to 1.8392867552141611325.

8. Apply Gaussian elimination with partial pivoting and then apply backward substitution to find the solution to

$$\left(\begin{array}{cc|c} -2.4 & -5.6 & 4.0 \\ 3.0 & 2.0 & 5.0 \end{array}\right).$$

First, swap Row 1 and Row 2 to put the largest entry in absolute value at the pivot location.

Then add 0.8 times Row 1 onto Row 2. The matrix will then be in row echelon form.

Assuming the solution vector is u, we then solve for $u_2 = -2$, and the substitute this into the first equation to get that $u_1 = 3$.

9. Apply Gaussian elimination with partial pivoting and then apply backward substitution to find the solution to

$$\left(\begin{array}{ccc|c} -1.2 & 4.4 & 4.9 & -0.2 \\ 2.4 & 0.2 & 0.4 & 2.8 \\ 4.0 & 2.0 & -3.0 & -6.0 \end{array}\right)$$

First, swap Rows 1 and 3 to put the largest entry in absolute value at the pivot location.

Then add –0.6 times Row 1 onto Row 2, and add 0.3 times Row 1 onto Row 3.

Next, swap Rows 2 and 3 to put the largest entry in absolute value at the pivot location.

Then add 0.2 times Row 2 onto Row 3. The matrix will then be in row echelon form.

Performing backward substitution, we get that $u_3 = 2$, $u_2 = -2$ and $u_1 = 1$.

10. Apply Gaussian elimination with partial pivoting and then apply backward substitution to find the solution to

$$\left(\begin{array}{cccc:c} 5.0 & 2.0 & 0 & 0 & 1.0 \\ 4.5 & 7.8 & -3.0 & 0 & -17.1 \\ 0 & -4.2 & 3.3 & 5.6 & 9.4 \\ 0 & 0 & 4.0 & 2.0 & 6.0 \end{array}\right)$$

This is called a *band-diagonal* matrix, and is actually quite common in engineering.

The largest entry in Column 1 is on the diagonal, so add –0.9 times Row 1 onto Row 2.

The largest entry in Column 2 is on the diagonal, so add 0.7 times Row 2 onto Row 3.

Swap Rows 3 and 4, and then add –0.3 times Row 3 onto Row 4. The matrix is now in row echelon form.

Performing backward substitution, we get that $u_4 = -1$, $u_3 = 2$, $u_2 = -2$ and $u_1 = 1$.

11. Apply two steps of Jacobi's method to find an approximation of the solution to:

$$\begin{pmatrix} 10 & 2 \\ 2 & 10 \end{pmatrix} \mathbf{u} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

$\mathbf{u_0} = \begin{pmatrix} 0.3 \\ 0.1 \end{pmatrix}$, and thus $\mathbf{u_1} = \begin{pmatrix} 0.28 \\ 0.04 \end{pmatrix}$, and $\mathbf{u_2} = \begin{pmatrix} 0.292 \\ 0.044 \end{pmatrix}$ and you can compare this to the exact solution which is $\mathbf{u} = \begin{pmatrix} 0.291666\cdots \\ 0.041666\cdots \end{pmatrix}$.

12. Apply one step of Jacobi's method to find an approximation of the solution to:

$$\begin{pmatrix} 5 & 2 & 1 \\ 2 & 10 & -3 \\ 1 & -3 & 20 \end{pmatrix} \mathbf{u} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.$$

$\mathbf{u_0} = \begin{pmatrix} 0.4 \\ 0.1 \\ 0.05 \end{pmatrix}$, and thus $\mathbf{u_1} = \begin{pmatrix} 0.35 \\ 0.035 \\ 0.045 \end{pmatrix}$ and you can compare this to the exact solution to six digits beyond the decimal point is $\mathbf{u} = \begin{pmatrix} 0.378663 \\ 0.035170 \\ 0.036342 \end{pmatrix}$.

13. The following system is given with the solution:

$$\begin{pmatrix} 2 & 5 \\ 5 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}.$$

If you were to try to apply Jacobi's method to find the solution, does it seem to converge? Why does this happen?

It does not converge because the matrix is not diagonally dominant.

14. Given the matrix $A = \begin{pmatrix} 10 & 9 \\ 9 & 8 \end{pmatrix}$ has the solution to $A\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ as $\mathbf{u}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. What is the solution to the problem $A\mathbf{u}_2 = \begin{pmatrix} 0.9 \\ 1.1 \end{pmatrix}$? What is $\left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0.9 \\ 1.1 \end{pmatrix} \right\|_2$? What is $\|\mathbf{u}_1 - \mathbf{u}_2\|_2$. Recall that $\| \|_2$ is the 2-norm or Euclidean norm where $\|\mathbf{u}\|_2 = \sqrt{u_1^2 + u_2^2}$. Use MATLAB to find the condition number of this matrix.

```
>> A = [10 9; 9 8];          % The matrix above
>> b1 = [1.0 1.0]';          % The first right-hand target vector
>> b2 = [0.9 1.1]';          % The second right-hand target vector
>> u1 = A \ b1
     u1 =
        1
       -1
>> u2 = A \ b2
     u2 =
        2.7
       -2.9

>> norm( b1 - b2 )
     ans = 0.14142

>> norm( u1 - u2 )
     ans = 2.54951
```

Note that a relative small error between $\mathbf{b}_1$ and $\mathbf{b}_2$ results in a large error in the solutions $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. Incidentally, this is because the condition number of the matrix A is 325.99693, so a small error can be magnified by up to two-and-a-half orders of magnitude ($\log_{10}(325.99693) = 2.51321$).