1. You apply gradient descent to the following (inconsistent) univarate linear system:

$$x = 0$$
$$x = 1$$

(a) Write the sum-of-squares loss f. For which value of $x^\star$ is $f(x^\star)$ minimized?

Solution: The sum-of-squares loss is $f(x) = x^2 + (x-1)^2$. Its gradient is $f'(x) = 2x + 2(x-1) = 2(2x-1)$. The loss is minimized when the gradient is zero, namely at $x^\star = 1/2$.

(b) Let x_t be the state after t steps of gradient descent with rate ρ. What is $(x_t - x^\star)/(x_{t-1} - x^\star)$?

Solution: It equals $1 - 4\rho$. The state evolution equation is $x_t = x_{t-1} - \rho f'(x_{t-1}) = x_{t-1} - 2\rho(2x_{t-1} - 1)$. Therefore $x_t - 1/2 = (x_{t-1} - 1/2) - 2\rho(2x_{t-1} - 1) = (1 - 4\rho)(x_{t-1} - 1/2)$.

(c) Assuming $x_0 = 0$ and $\rho = 1/8$, what is x_{20}?

Solution: By part (b),

$$x_{20} - 1/2 = (1 - 4\rho)^{20}(x_0 - 1/2) = (1/2)^{20}(-1/2) = -(1/2)^{21}.$$

2. You apply subspace iteration on the matrix

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}$$

Your initial state is the standard basis $\mathbf{x}_1 = (1,0,0)$, $\mathbf{x}_2 = (0,1,0)$, $\mathbf{x}_3 = (0,0,1)$.

(a) Ignoring $\mathbf{x}_3$ for now, show how $\mathbf{x}_1$ and $\mathbf{x}_2$ evolve in the first three rounds of subspace iteration.

Solution: Round 1: $A\mathbf{x}_1$ and $A\mathbf{x}_2$ are the first two columns of A. They are already orthogonal but not of unit norm. After orthogonalization the state is $\mathbf{x}_1 = (0, 1/\sqrt{5}, 2/\sqrt{5})$ and $\mathbf{x}_2 = (1,0,0)$.
Round 2: $A\mathbf{x}_1$ now becomes $(\sqrt{5}, 0, 0)$ and $A\mathbf{x}_2$ becomes $(0,1,2)$. After orthogonalization $\mathbf{x}_1 = (1,0,0)$ and $\mathbf{x}_2 = (0, 1/\sqrt{5}, 2/\sqrt{5})$. They switch places from round 2.
Round 3: They now switch places again, so the outcome is the same as after round 1. The state keeps iterating between $\mathbf{x}_1 = (0, 1/\sqrt{5}, 2/\sqrt{5})$, $\mathbf{x}_2 = (1,0,0)$ and $\mathbf{x}_1 = (1,0,0)$, $\mathbf{x}_2 = (0, 1/\sqrt{5}, 2/\sqrt{5})$.

(b) Use part (b) to find two of the three eigenvalues of A. Explain your reasoning.

Solution: The two eigenvalues are $\sqrt{5}$ and $-\sqrt{5}$. As $A\mathbf{x}_1 = \sqrt{5}\mathbf{x}_2$ and $A\mathbf{x}_2 = \sqrt{5}\mathbf{x}_1$, $\mathbf{x}_1 + \mathbf{x}_2$ is an eigenvector with eigenvalue $\sqrt{5}$ and $\mathbf{x}_1 - \mathbf{x}_2$ is an eigenvector with eigenvalue $-\sqrt{5}$.

(c) Prove that if, at any point during power iteration, $A\mathbf{x}_i$ and $A\mathbf{x}_j$ become linearly dependent (for some $i \neq j$), zero must be an eigenvalue of A.

Solution: if $A\mathbf{x}_i = cA\mathbf{x}_j$ then $A(\mathbf{x}_i - c\mathbf{x}_j)$ must be zero. But $\mathbf{x}_i$ and $\mathbf{x}_j$ are orthogonal so $\mathbf{x}_i - c\mathbf{x}_j$ is a nonzero eigenvector with eigenvalue zero.

(d) Use part (c) to find the remaining eigenvalue of A. (**Hint:** What happens to $\mathbf{x}_3$?)

Solution: In round 1 already $A\mathbf{x}_2$ and $A\mathbf{x}_3$ are linearly dependent as they are both multiples of $(0,1,2)$. By part (c) zero is an eigenvalue of A.

3. Apply the modular Fast Fourier transform to calculate the polynomial representation of

x	1	i	-1	$-i$
$f(x)$	0	1	2	3

(a) Calculate the list representations of f^+ and f^-, where $f(x) = f^+(x^2) + xf^-(x^2)$.

Solution: $f^+(x^2) = (f(x) + f(-x))/2$ has list representation $f(1) = 1, f(-1) = 2$. $f^-(x^2) = (f(x) - f(-x))/2x$ has list representation $f(1) = -1, f(-1) = i$.

(b) Calculate the polynomial representations of f^+ and f^- over domain $\{-1, +1\}$.

Solution: g over domain $-1, 1$ has polynomial representation $g(y) = \frac{1}{2}(g(1)+g(-1)) + \frac{1}{2}(g(1)-g(-1)) \cdot y$. Applying this formula we get $f^+(y) = \frac{3}{2} - \frac{1}{2} \cdot y$ and $f^-(y) = (-\frac{1}{2} + \frac{1}{2}i) + (-\frac{1}{2} - \frac{1}{2}i) \cdot y$.

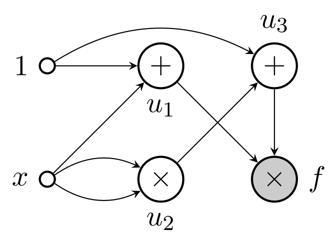
(c) Calculate the polynomial representation of f.

Solution: $f(x) = f^+(x^2) + xf^-(x^2) = \frac{3}{2} + (-\frac{1}{2} + \frac{1}{2}i) \cdot x - \frac{1}{2} \cdot x^2 + (-\frac{1}{2} - \frac{1}{2}i) \cdot x^3$.

4. Let $f(x) = 1 + x + x^2 + x^3$.

(a) Draw a circuit for f with at most two plus gates and two times gates.

Solution: If we do not restrict the fan-in, i.e. the number of incoming wires into each gate, there are many possible implementations. Here is an implementation with fan-in two that corresponds to the formula $(1 + x)(1 + x^2)$.



(b) Draw the circuit obtained by applying backpropagation to part (a). Explain any simplifications you apply.

Solution: Applying backpropagation in reverse topological order f, u_3, u_2, u_1, x we find $df/df = 1$ and

$$\frac{df}{du_3} = \frac{df}{df} \cdot \partial_{u_3}[f] = \frac{df}{df} \cdot u_1 \qquad \text{simplifies to } u_1$$
$$\frac{df}{du_2} = \frac{df}{du_3} \cdot \partial_{u_2}[u_3] = \frac{df}{du_3} \cdot 1 \qquad \text{simplifies to } u_1$$
$$\frac{df}{du_1} = \frac{df}{df} \cdot \partial_{u_1}[f] = \frac{df}{df} \cdot u_3 \qquad \text{simplifies to } u_3$$
$$\frac{df}{dx} = \frac{df}{du_1} \cdot \partial_x[u_1] + \frac{df}{du_2} \cdot \partial_x[u_2] + \frac{df}{du_2} \cdot \partial_x[u_2] = \frac{df}{du_1} \cdot 1 + \frac{df}{du_2} \cdot x + \frac{df}{du_2} \cdot x$$

which finally simplifies to $u_3 + u_1 \cdot 2x$. The (simplified) circuit is