Practice Midterm 1

1. Systems A and B consist of four linear equations in four unknowns modulo 2.

$$\boxed{\begin{array}{ccccc} x_1 & + & x_2 & = & 0 \\ + & & + & & \\ y_1 & + & y_2 & = & 0 \\ = & & = & & \\ 1 & & 1 & & \end{array}} \qquad \boxed{\begin{array}{ccccc} x_1 & + & x_2 & = & 1 \\ + & & + & & \\ y_1 & + & y_2 & = & 0 \\ = & & = & & \\ 1 & & 1 & & \end{array}}$$

system A system B

   (a) Solve system A using modular Gauss-Jordan elimination. If there are multiple solutions, specify the free variables and the assignment to the remaining variables in terms of the free variables.

   Solution: Keeping the equation $x_1 + x_2 = 0$ and eliminating x_1 from the others gives the system $x_2 + y_1 = 1, x_2 + y_2 = 1, y_1 + y_2 = 0$. Now we keep $x_2 + y_1 = 1$ and eliminate x_2 to obtain $y_1 + y_2 = 0$ twice. Eliminating y_1 leaves y_2 as the free variable. Substituting back gives the solution $y_1 = y_2$, $x_2 = y_2 + 1$, $x_1 = y_2 + 1$. (As y_2 can be either 0 or 1 there are two solutions for (x_1, x_2, y_1, y_2): $(1, 1, 0, 0)$ and $(0, 0, 1, 1)$.)

   (b) In system B the value in the first equation was flipped. Produce a contradictory linear combination for system B. You may use any method you like.

   Solution: The contradictory linear combination is the sum of all the equations. As each variable appears twice the left-hand side vanishes modulo two. The right hand side equals $1 + 0 + 1 + 1 = 1$.

2. You want to apply Gradient Descent to the system

$$\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

   The eigenvalues of the matrix are $\lambda_1 = 4$ and $\lambda_2 = -2$ with corresponding eigenvectors $\mathbf{v}_1 = (1/\sqrt{2}, 1/\sqrt{2})$, $\mathbf{v}_2 = (1/\sqrt{2}, -1/\sqrt{2})$.

   (a) What is the maximum rate ρ^* below which convergence to the unique solution $(1/4, 1/4)$ is guaranteed (for any initialisation)?

   Solution: ρ^* is the inverse square of the spectral norm of A. The spectral norm of a symmetric matrix is the largest eigenvalue in absolute value, which is 16 in this case. Therefore $\rho^* = 1/16$.

   (b) You run Gradient Descent with rate $\rho = 0.01$ and initialisation $(x, y) = (1, 0)$. The distance between the state (x_t, y_t) at time t and the solution $(1/4, 1/4)$ is $\Theta(b^t)$ for some number b. Find b.

   Solution: $b = 0.92$. Let $\mathbf{x}^* = (1/4, 1/4)$ and $\mathbf{x} = (x, y)$ be the state of the algorithm at some point. The difference $\mathbf{x} - \mathbf{x}^*$ is multiplied by the matrix $B = I - 2\rho AA^T$ at every step of the algorithm. As A is symmetric, AA^T has the same eigenvectors as A^2 and therefore A, and so does $B = I - 2\rho AA^T$. The associated eigenvalues are $1 - 2\rho\lambda_1^2 = 0.68$ and $1 - 2\rho\lambda_2^2 = 0.92$.

3. The list representation of f defined over the fourth roots of unity is

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

(a) Find the polynomial representation of f. You may use any method you like. Explain your steps.

Solution: We can calculate the Fourier coefficients by averaging the product of f and the corresponding basis function:

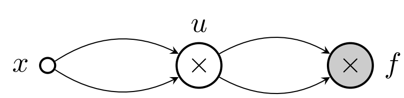
$$\begin{aligned}
\hat{f}(0) &= \text{average of } f(x) = \frac{1+1}{4} = \frac{1}{2} \\
\hat{f}(1) &= \text{average of } f(x) \cdot x = \frac{i + (-i)}{4} = 0 \\
\hat{f}(2) &= \text{average of } f(x) \cdot x^2 = \frac{i^2 + (-i)^2}{4} = -\frac{1}{2} \\
\hat{f}(3) &= \text{average of } f(x) \cdot x^3 = \frac{i^3 + (-i)^3}{4} = 0
\end{aligned}$$

so $f(x) = \frac{1}{2} - \frac{1}{2}x^2$.

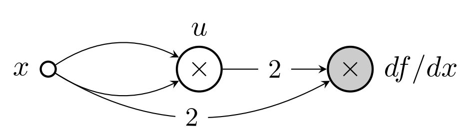
(b) What is the linear approximation of f that minimizes the average square error?

Solution: It is $\hat{f}(0) + \hat{f}(1)x$, which is the constant $1/2$. (The average square error of this approximation is $1/2$.)

4. Show the result of applying backpropagation to the circuit below. Your circuit may use plus and times gates. If you applied any simplifications (e.g., $1 \times x$ was replaced by x) explain them.



Solution: The circuit has gates $df/df = 1$, df/du, and df/dx with connections $df/du = u \times df/df + df/df \times u$ which we simplify to $df/df \times (2u)$ and further to $2u$, and $df/dx = x \times df/du + df/du \times x$, which we simplify to $df/du \times (2x)$. The f-gate can be removed as it it does not affect the output df/dx. The resulting simplified circuit is



Practice Midterm 2

1. Apply Gauss-Jordan Elimination to find a contradictory linear combination for the system of equations below. Explain all the steps.

$$\begin{aligned}
x - y &= 1 \\
y - z &= 1 \\
z - x &= 1
\end{aligned}$$

Solution: We are looking for a linear combination of the equations for which the left-hand side is identically zero and the right-hand side is 1. Let a, b, c be the coefficients. Then $(a-c)x + (c-b)z + (b-a)y = a+b+c$, so we need $a - c = 0, c - b = 0, b - a = 0, a + b + c = 1$. Eliminating a from the last three equations we obtain $b - c = 0$ (twice) and $b + 2c = 1$. Eliminating b we get $3c = 1$ and substituting back we find $a = b = c = 1/3$. Indeed, scaling each equation by $1/3$ and adding everything up gives the contradiction $0 = 1$.

2. You apply Power Iteration on symmetric matrix A with initialization $\mathbf{x}$.

(a) Let $\mathbf{x}$ be the state at the end of step t and $\mathbf{y}$ be the state in step $t+1$ before normalization. Prove that the spectral norm of A is at least $\|\mathbf{y}\|/\|\mathbf{x}\|$, regardless of initialization.

Solution: The spectral norm is the largest possible value of $\|A\mathbf{z}\|/\|\mathbf{z}\|$ for all possible $\mathbf{z}$. If we instantiate $\mathbf{z}$ with $\mathbf{x}$, we get that $\|\mathbf{y}\|/\|\mathbf{x}\| = \|A\mathbf{x}\|/\|\mathbf{x}\|$ cannot be larger than this maximum.

(b) Use part (a) to argue that the following matrix has spectral norm greater than 3.

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

Solution: Suppose we instantiate Power Iteration with $\mathbf{x} = (1,1,1)$. Then $\mathbf{y} = A\mathbf{x} = (3,4,3)$ and $\|A\| \geq \|\mathbf{y}\|/\|\mathbf{x}\| = \sqrt{3^2+4^2+3^2}/\sqrt{3} \geq 3$.

3. The function $f(x,y,z)$ evaluates to 1 if all its inputs are $+1$, and to -1 if at least one of them is a -1. Apply the Fast Fourier-Walsh algorithm to calculate the Fourier coefficients of f. Show all the steps. You may shortcut the execution if an intermediate function simplifies to a constant.

Solution: The function $f(x,y,-1)$ is always -1. The function $f(x,y,+1)$ is 1 if $x=y=+1$ and -1 otherwise. Applying the next round of recursion we get $f(x,+1,+1) = x$ and $f(x,-1,+1) = -1$, so

$$f(x,y,+1) = f(x,+1,+1)\cdot\frac{1+y}{2} + f(x,-1,+1)\cdot\frac{1-y}{2} = x\frac{1+y}{2} + (-1)\frac{1-y}{2} = -\frac{1}{2} + \frac{1}{2}x + \frac{1}{2}y + \frac{1}{2}xy$$

and

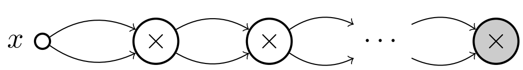
$$\begin{aligned} f(x,y,z) &= f(x,y,+1)\cdot\frac{1+z}{2} + f(x,y,-1)\cdot\frac{1-z}{2} \\ &= \left(-\frac{1}{2} + \frac{1}{2}x + \frac{1}{2}y + \frac{1}{2}xy\right)\cdot\frac{1+z}{2} + (-1)\cdot\frac{1-z}{2} \\ &= -\frac{3}{4} + \frac{1}{4}x + \frac{1}{4}y + \frac{1}{4}z + \frac{1}{4}xy + \frac{1}{4}xz + \frac{1}{4}yz + \frac{1}{4}xyz. \end{aligned}$$

The Fourier coefficients are $\hat{f}(\varnothing) = -3/4$ and $\hat{f}(S) = 1/4$ for all other S.

4. In this question all times gates take exactly two inputs. Assume n is a power of two.

(a) Draw a circuit for x^n with $\log n$ times gates.

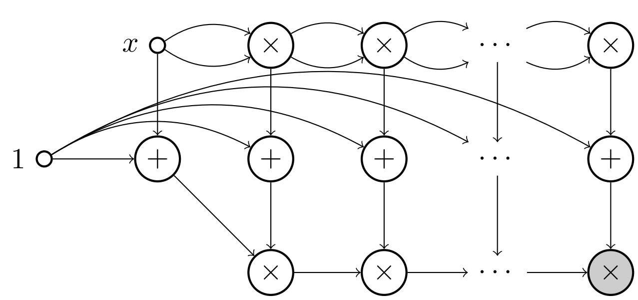
Solution:



(b) Let $f(x) = 1 + x + x^2 + \cdots + x^{n-1}$. Show that f has a circuit (with plus and times gates) of size $O((\log n)^2)$. (**Hint:** Factor f. Try $n=4$ first.)

Solution: f factors into $(1+x)(1+x^2)(1+x^4)\cdots(1+x^{n/2})$. There are $\log n$ factors. By part (a) each factor can be calculated with $O(\log n)$ gates. As f is the product of them all, it can be calculated with $O((\log n)^2)$ gates.

In fact, f can be calculated with $O(\log n)$ gates only by the following circuit:



(c) Let $g(x) = 1 + 2x + 3x^2 + \cdots + (n-1)x^{n-2}$. Show that g has a circuit of size $O((\log n)^2)$. (**Hint:** Forward/backpropagation.)

Solution: As both the times and plus gates have two incoming wires, the circuit for f has $O(\log n)$ wires (or $O((\log n)^2)$ for the less optimal construction). As g is the derivative of f, applying backpropagation to f gives a circuit for $df/dx = g$ of same asymptotic size.

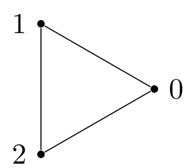
Practice Midterm 3

1. Apply Gauss-Jordan Elimination to find a *nonzero* solution to the linear system below. Explain your steps.

$$\begin{aligned} x + y + z &= 0 \\ x + 2y + 4z &= 0 \end{aligned}$$

Solution: After eliminating x we obtain $y + 3z = 0$. After eliminating y, z becomes a free variable. The solution found by Gauss-Jordan Elimination is $FREE = \{z\}$, $y = -3z$, $x = -y - z = 2z$. Setting z to any nonzero value yields a nonzero solution. For example, $z = 1$ gives $y = -3$ and $x = 2$.

2. Let A be the adjacency matrix of the 3-cycle (with zeros on the diagonal). The largest eigenvalue of A is $\lambda_1 = 2$. The corresponding eigenvector is $\mathbf{v}_1 = (1/\sqrt{3})(1,1,1)$.



(a) Let $\mathbf{x}$ be any vector that is orthogonal to $\mathbf{v}_1$. Show that the state of Power Iteration on A initialized with $\mathbf{x}$ oscillates between $\mathbf{x}$ and $-\mathbf{x}$. Assume there are no precision errors.

Solution: Let $\mathbf{x} = (x_1, x_2, x_3)$. As it is orthogonal to $\mathbf{v}_1$, $x_1 + x_2 + x_3 = 0$. Applying A to $\mathbf{x}$ gives $A\mathbf{x} = (x_2 + x_3, x_3 + x_1, x_1 + x_2) = (-x_1, -x_2, -x_3) = -\mathbf{x}$ because subtracting $x_1 + x_2 + x_3$ from every coordinate does not change $\mathbf{x}$.

(b) Use part (a) to deduce the other two eigenvalues λ_2 and λ_3 of A. What are they?

Solution: $\lambda_2 = \lambda_3 = -1$. As eigenvectors are mutually orthogonal, applying A to an eigenvector $\mathbf{v}$ other than $\mathbf{v}_1$ outputs $-\mathbf{v}$, so the corresponding eigenvalue must be -1.

3. Calculate the Fourier transforms of the following functions. You may use any method you like. Write your answer in the tables provided. $+$ and $-$ are shorthand for the numbers $+1$ and -1.

(a)

x_1x_2	$++$	$-+$	$+-$	$--$
$f(x_1x_2)$	1	1	-1	-1

S	$\varnothing$	1	2	12
$\hat{f}(S)$	0	0	1	0

This is the function x_2. Its only nonzero Fourier coefficient is $\hat{f}(\{2\}) = 1$.

(b)

x_1x_2	$++$	$-+$	$+-$	$--$
$g(x_1x_2)$	0	1	1	1

S	$\varnothing$	1	2	12
$\hat{g}(S)$	$3/4$	$-1/4$	$-1/4$	$-1/4$

(**Hint:** What is $1 - g$?)
$1 - g$ is the point function point_{++}. Its polynomial representation is $(1 + x_1)/2 \cdot (1 + x_2/2) = \frac{1}{4} + \frac{1}{4}x_1 + \frac{1}{4}x_2 + \frac{1}{4}x_1x_2$. Therefore $g = \frac{3}{4} - \frac{1}{4}x_1 - \frac{1}{4}x_2 - \frac{1}{4}x_1x_2$.

(c) $h(x_1x_2x_3) = \begin{cases} f(x_1x_2), & \text{if } x_3 = -1 \\ g(x_1x_2), & \text{if } x_3 = +1. \end{cases}$

S	$\varnothing$	1	2	3	12	13	23	123
$\hat{g}(S)$	$\frac{3}{8}$	$-\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$-\frac{1}{8}$	$-\frac{1}{8}$	$-\frac{5}{8}$	$-\frac{1}{8}$

(**Hint:** Apply one of the steps of the Fast Fourier-Walsh algorithm.)

$$\begin{aligned}
h(x_1x_2x_3) &= \frac{1-x_3}{2}\cdot f + \frac{1+x_3}{2}\cdot g \\
&= \frac{1-x_3}{2}\cdot x_2 + \frac{1+x_3}{2}\cdot\left(\frac{3}{4}-\frac{1}{4}x_1-\frac{1}{4}x_2-\frac{1}{4}x_1x_2\right) \\
&= \left(\frac{1}{2}x_2-\frac{1}{2}x_2x_3\right)+\left(\frac{3}{8}-\frac{1}{8}x_1-\frac{1}{8}x_2-\frac{1}{8}x_1x_2+\frac{3}{8}x_3-\frac{1}{8}x_1x_3-\frac{1}{8}x_2x_3-\frac{1}{8}x_1x_2x_3\right) \\
&= \frac{3}{8}-\frac{1}{8}x_1+\frac{3}{8}x_2-\frac{1}{8}x_1x_2+\frac{3}{8}x_3-\frac{1}{8}x_1x_3-\frac{5}{8}x_2x_3-\frac{1}{8}x_1x_2x_3.
\end{aligned}$$

4. What is the linear approximation of f that minimizes the average square error?

   Solution: The constant $3/4$. (The approximation error is $3\cdot(1/4)^2 = 3/16$.)

5. Show the result of forward propagation on the circuit $f(x) = x(1+x(1+x))$. Your circuit may use plus and times gates. If you applied any simplifications explain them.

   Solution: I will apply the forward propagation algorithm for polynomials. Every multiplication is of the form $1+h$. Its derivative is $d1/dx + dh/dx = 0 + dh/dx$. I omit the addition of zero. Every multiplication is of the form $x \times h$ its derivative is $1\times h + x\times dh/dx$. I omit the multiplication by one and simplify this to $h + x \times dh/dx$. (The f-labeled gate is not used and could have been omitted.)