Problem Set 9 Solutions

Problem 1. Suppose we have a baseball league with 22 teams, and we are given a schedule in which each team plays 14 games.

(a) Assuming all teams are in different cities, prove that we can arrange for each team to play the same number of home and away games).

Consider the graph (N, E) whose nodes are the 22 teams, and whose edges are the games to be played. The number of edges incident each node is 14. The total number of edges in the graph is $(22 \cdot 14)/2 = 154$. Since the degree of each node is even, the graph has an euler cycle $e_1, e_2, \ldots, e_{154}$. For each edge $e_i = (x_i, y_i)$, let the corresponding game be a home game for x_i and an away game for y_i. In other words, when we enter a node, the corresponding incoming edge is an away game, and when we exit it the corresponding outgoing edge is a home game for that team. Since the euler path must enter and exit any given node the same number of times, the corresponding team will play the same number of home and away games.

(b) Prove we can split the season into two halves (one before and one after the all-star break) in which each team plays seven games.

Notice that the euler path we found in part (a) has even length. Therefore we can color the edges in the path with alternating colors, say blue and red. Since each time we enter and exit a node in the euler cycle the traversed edges have different colors, the edges incident any given node will be 7 blue and 7 red. Divide the set of all games into two halves based on the color of the corresponding edges. Each team will play 7 game in each half of the season.

Problem 2. For a simple undirected graph with edges $e_1, \ldots, e_m$, consider the following algorithm:

initially, unmark every edge
for $i = 1$ to m
  if there does not exist a path of marked edges between the endpoints of e_i
    mark edge e_i

Prove that when the algorithm finishes, the set of marked edges forms a spanning forest of the input graph. (Hint: it may help to consider the graphs consisting of the edges $e_1, \ldots, e_k$ for various values of k.)

Spanning Forest, is a forest that contains every vertex of G and connects every pair of vertices that were connected in G.

Loop invariant: At any point in the loop there is at most one marked path between any two vertices. (No cycle.)

1. *Proof:* by induction on the number of for-loop iterations

   Base case: No iteration. Initially all edges are unmarked, so there is no path between any two vertices, thus there is is at most 1 marked path between any two vertices.

   Inductive Hypothesis: Assume that on the i^{th} iteration there is at most one marked path between any two vertices.

   Inductive Step: Prove true on the $(i+1)^{th}$ iteration.

   On the $(i+1)^{th}$ iteration, the algorithm will mark the edge if there is no marked path between its endpoints otherwise it will leave the edge unmarked.

   If the edge is left unmarked then the invariant holds true by the inductive hypothesis.

   If the edge is marked then there is 1 marked path between the endpoints of the edge due to the edge marking pre-condition. Therefore, the invariant holds true.

2. Since the for-loop loops over all the edges then all the vertices of G (that have edges) will be contained in the forest.

Problem 3. Consider the following state machine on a graph G. A state is specified by labeling each vertex of G as *uncolored*, *red*, or *blue*. Initially, all vertices are uncolored except for one vertex v which is red.

make-blue(w)
Precondition:
a neighbor of w is red
Effect:
w becomes blue

make-red(w)
Precondition:
a neighbor of w is blue
Effect:
w becomes red

(a) Prove that if G is bipartite and the above code reaches a state where all vertices are colored, then the coloring is a legal coloring (that is, all neighbors of a red vertex are blue, and vice versa). You'll want to use some facts from the last homework to do this.

It can be easily proved by induction on the number of transitions that if a node w ever gets colored with red then there is an even length path from v to w, and if it get colored with blue there is an odd length path from v to w. Now assume that the state machine reaches a state in which all nodes are colored and assume for contradiction that two neighboring nodes (a, b) have the same color. If they are red, there is an even length path from v to a, a length 1 path from a to b and an even length path from b to v. The concatenation of these three paths gives an odd length cycle.

Analogously, if a and b are blue, there are odd length paths from v to a and b respectively and they close with the edge (a, b) an odd length cycle. In both case we have a contradiction because in the last homework we proved that bipartite graphs do not have odd length cycle.

(b) For what kinds of bipartite graphs is it possible to reach a state in which all vertices are colored? A brief explanation without proof is sufficient.

If there is no path from the original red vertex, v, to some vertex, w, then w won't be colored. Therefore, only for connected graphs the state machine can reach a state in which all vertices are colored.

Problem 4. Which of the following assertions is true? Prove your answer.

1. If f is $O(g)$ then $\ln f$ is $O(\ln g)$.

   False, even if we assume $f(x), g(x) > 1$ for all x. Consider for example the functions $f(x) = 2 \cdot e^{1/x}$ and $g(x) = e^{1/x}$. Clearly $f(x)$ is $O(g(x))$, but $lnf(x) = \ln 2 + 1/x$ is not $O(g(x)) = O(1/x)$ because for every constant C and sufficiently large x, $\ln 2 + 1/x > C/x$.

   However if we assume that $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = \infty$ the statement becomes true: there exists a constant C and x_o such that $f(x) \leq Cg(x)$ for all $x \geq x_o$.
   This implies that $\ln f(x) \leq \ln C + \ln g(x) < 2 \ln g(x)$ for sufficiently large x. Hence $\ln f(x)$ is $O(g(x))$.

2. If $\ln f$ is $O(\ln g)$ then f is $O(g)$.

   False. Let, $f = e^{2x}, g = e^x$. Note, $\ln f(x)$ is $O(\ln g(x))$ since $2x \leq Cx$ for all x.
   But, e^{2x} is not $O(e^x)$ since $lim_{x\to\infty} f(x)/g(x) = \lim_{x\to\infty} e^x = \infty$.

3. If f is $O(g)$ and g is never zero then $1/g$ is $O(1/f)$.

   True. If $f(x)$ is $O(g(x))$ then $|f(x)| \leq C|g(x)|$ for all $x \geq x_o$.
   Dividing the above inequality by $|f(x)||g(x)|$, $\frac{1}{|g(x)|} \leq C\frac{1}{|f(x)|}$ for all $x \geq x_o$.

   Therefore, $\frac{1}{g}$ is $O(\frac{1}{f})$.

Problem 5. A college loan officer offers you the following deal. He will give you \$25,000 at the start of each year for four years starting today to pay for your college tuition. In return, you must pay \$15,000 at the start of each year for fifteen years starting in four years. Suppose the annual interest rate is 5%. Is this a good deal? Explain.

Present value of what the college loan officer offers

$$\begin{aligned}
&= 25000 + 25000/1.05 + 25000/(1.05)^2 + 25000/(1.05)^3 \\
&= 25000(1 - \frac{1}{(1.05)^4})/(1 - \frac{1}{1.05}) \\
&= 25000 * 3.577 \\
&= \$93,081.20
\end{aligned}$$

Present value of what you must pay in return

$$\begin{aligned}
&= 15000/(1.05)^4 + 15000/(1.05)^5 + \ldots + 15000/(1.05)^{18} \\
&= (15000/(1.05)^4)(1 - \frac{1}{(1.05)^{12}})/(1 - \frac{1}{1.05}) \\
&= \$134,495.08
\end{aligned}$$

This is not a fair deal! You are overpaying them by $\$41,413.88$.

Problem 6. The integral method

(a) Find tight (Θ) asymptotic bounds for the summation $\sum_{i=n}^{\infty} \frac{1}{i^2}$ where $n > 0$.

$$\begin{aligned}
\int_n^{\infty} \frac{1}{x^2}dx &\leq \sum_{i=n}^{\infty} \frac{1}{i^2} \quad \int_{n-1}^{\infty} \frac{1}{x^2}dx \\
-\frac{1}{2x}\bigg|_n^{\infty} &\leq \sum_{i=n}^{\infty} \frac{1}{i^2} \leq -\frac{1}{2x}\bigg|_{n-1}^{\infty} \\
\frac{2}{n} &\leq \sum_{i=n}^{\infty} \frac{1}{i^2} \leq \frac{1}{2(n-1)} \leq \frac{1}{n}
\end{aligned}$$

for all $n \geq 2$, that is $\sum_{i=n}^{\infty} \frac{1}{i^2}$ is $\Theta(1/n)$.

(b) Give tight (Θ) bounds for $\sum_{i=1}^{n} \frac{1}{\sqrt{i}}$

$$\begin{aligned}
\int_1^{n+1} \frac{1}{\sqrt{x}}dx &\leq \sum_{i=1}^{n+1} \frac{1}{\sqrt{i}} \leq \int_0^n \frac{1}{\sqrt{x}}dx \\
2x^{\frac{1}{2}}\bigg|_1^{n+1} &\leq \sum_{i=1}^{n} \frac{1}{\sqrt{i}} \leq 2x^{\frac{1}{2}}\bigg|_0^n \\
\sqrt{n} \leq 2\sqrt{n+1} - 2 &\leq \sum_{i=1}^{n} \frac{1}{\sqrt{i}} \leq 2\sqrt{n}
\end{aligned}$$

for all $n \geq 4$, that is $\sum_{i=1}^{n} \frac{1}{\sqrt{i}}$ is $\Theta(\sqrt{n})$.

Problem 7. Find a closed form expression $f(n)$ for $\sum_{i=1}^{n} i^2 x^i$.

We first come up with a formula for $\sum_{i=1}^{n} i^2 x^i$ and then prove it using induction. From class notes we know that

$$\sum_{i=1}^{n} i x^i = \frac{x - (n+1)x^{n+1} + n x^{n+2}}{(1-x)^2}.$$

Differentiating (with respect to x) the above expression and multiplying with x we find that

$$\sum_{i=1}^{n} i^2 x^i = \frac{x + x^2 - (n+1)^2 x^{n+1} + (2n^2 + 2n - 1)x^{n+2} - n^2 x^{n+3}}{(1-x)^3}$$

Proof by induction. Let $P(n)$ be the statement that

$$\sum_{i=1}^{n} i^2 x^i = \frac{x + x^2 - (n+1)^2 x^{n+1} + (2n^2 + 2n - 1)x^{n+2} - n^2 x^{n+3}}{(1-x)^3}.$$

Base Case. $P(1)$ is true since

$$\sum_{i=1}^{1} i^2 x^i = x = \frac{x + x^2 - 4x^2 + 3x^3 - x^4}{(1-x)^3}.$$

Inductive Step. Assume $P(n)$ is true. Then

$$\begin{aligned}
\sum_{i=1}^{n} i^2 x^i + (n+1)^2 x^{n+1} &= \\
\sum_{i=1}^{n+1} i^2 x^i &= \frac{x + x^2 - (n+1)^2 x^{n+1} + (2n^2 + 2n - 1)x^{n+2} - n^2 x^{n+3}}{(1-x)^3} + (n+1)^2 x^{n+1} \\
&= \frac{x + x^2 - (n+2)^2 x^{n+2} + (2(n+1)^2 + 2(n+1) - 1)x^{n+3} - (n+1)^2 x^{n+4}}{(1-x)^3}
\end{aligned}$$

Thus $P(n+1)$ is true. Hence, $P(n)$ is true for all n.

Problem 8. By manipulating summations, without recourse to induction, prove that

$$\sum_{k=1}^{n} H_k = (n+1)H_n - n$$

$$\begin{aligned}
\sum_{k=1}^{n} H_k &= \sum_{k=1}^{n} \sum_{j=1}^{k} \frac{1}{j} \\
&= \sum_{1 \le j \le k \le n} \frac{1}{j} \\
&= \sum_{j=1}^{n} \sum_{k=j}^{n} \frac{1}{j}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^{n} \frac{n-j+1}{j} \\
&= \sum_{k=1}^{n} \left(\frac{n+1}{j} - 1\right) \\
&= (n+1)\sum_{j=1}^{n} \frac{1}{j} - n \\
&= (n+1)H_n - n
\end{aligned}$$

Problem 9. Apparently, a standard model used by lobster fishermen is the following: the number of lobsters caught in a given year is the average of the number caught in the previous two years. Let's try out this model in the following hypothetical fishing community — in year 0, there were 100 lobsters caught, and in year 1, there were 300 lobsters caught.

(a) Let $f(i)$ denote the number of lobsters caught in year i. Write a recurrence relation for $f(i)$ and obtain a closed-form solution.

$$f(i) = (1/2)(f(i-1) + f(i-2))$$

This is a homogenous linear recurrence. We solve the characteristic equation, $2x^2 - x - 1 = 0$, to get the roots 1 and $-1/2$. Hence the general solution is of the form $f(i) = A + B(-1/2)^i$. Using the boundary conditions $f(0) = 100$ and $f(1) = 300$ we find that the general solution is $f(i) = 700/3 - (400/3)(-1/2)^i$.

(b) Now let's consider the following variation on this model: the number of lobsters caught in a given year is the average of the number caught in the previous *three* years. To start things off, suppose that 100 were caught in year 0, 300 in year 1, and 200 in year 2.

Let $g(i)$ denote the number of lobsters caught in year i under this model. Again, write a recurrence relation for $g(i)$ and obtain a closed-form solution.

$$g(i) = (1/3)(g(i-1) + g(i-2) + g(i-3))$$

This is a homogenous linear recurrence. We solve the characteristic equation, $3x^3 - x^2 - x - 1 = 0$, to get the roots 1, $\frac{-1-\sqrt{-2}}{3}$ and $\frac{-1+\sqrt{-2}}{3}$. Hence the general solution is of the form $g(i) = A + B(\frac{-1-\sqrt{-2}}{3})^i + C(\frac{-1+\sqrt{-2}}{3})^i$. Using the boundary conditions $g(0) = 100$, $g(1) = 300$ and $g(2) = 200$ we find that the general solution is $g(i) = 650/3 + (-25/3)(7 - 4\sqrt{-2})(\frac{-1-\sqrt{-2}}{3})^i + (-25/3)(7 + 4\sqrt{-2})(\frac{-1+\sqrt{-2}}{3})^i$.

Problem 10. Consider the following recurrence:

$$T(n) = aT(n/2) + n^2 \log_2 n, \quad T(1) = 1.$$

(a) Using plug and chug, derive a solution to this recurrence in summation form. Assume n is of the form 2^m.

$$\begin{aligned}
T(n) &= aT(2^{m-1}) + 2^{2m}m \\
&= a^2T(2^{m-2}) + a \cdot 2^{2(m-1)}(m-1) + 2^{2m}m \\
&= \ldots \\
&= a^iT(2^{m-i}) + \sum_{k=0}^{i-1} a^k \cdot 2^{2(m-k)}(m-k)
\end{aligned}$$

$$
\begin{aligned}
&= \ldots \\
&= a^m T(1) + \sum_{k=0}^{m-1} a^k \cdot 2^{2(m-k)}(m-k) \\
&= a^m + \sum_{k=0}^{m-1} a^k \cdot 2^{2(m-k)}(m-k) \\
&= a^m + \sum_{k=1}^{m} a^{m-k} 2^{2k} k \\
&= a^m \cdot \left(1 + \sum_{k=1}^{m} (4/a)^k k\right) \\
&= a^{\log_2 n} \cdot \left(1 + \sum_{k=1}^{\log_2 n} (4/a)^k k\right)
\end{aligned}
$$

(b) Prove that the recurrence satisfies this summation form

$$
\begin{aligned}
aT(n/2) + n^2 \log_2 n &= a\left(a^{\log_2(n/2)} \cdot \left(1 + \sum_{k=1}^{\log_2(n/2)} (4/a)^k k\right)\right) + 2^{2\log_2 n} \log_2 n \\
&= a \cdot a^{\log_2 n - 1} \left(1 + \sum_{k=1}^{\log_2 n - 1} (4/a)^k k\right) + 4^{\log_2 n} \log_2 n \\
&= a^{\log_2 n} \left(1 + \sum_{k=1}^{\log_2 n - 1} (4/a)^k k + (4/a)^{\log_2 n} \log_2 n\right) \\
&= a^{\log_2 n} \left(1 + \sum_{k=1}^{\log_2 n} (4/a)^k k\right) \\
&= T(n)
\end{aligned}
$$

(c) By manipulating sums, derive a (simple) closed form for the summation in the case $a = 2$.
We know from class that

$$\sum_{i=1}^{t} i x^i = \frac{x - (t+1)x^{t+1} + t x^{t+2}}{(1-x)^2}$$

and in particular

$$\sum_{i=1}^{t} i 2^i = 2 - (t+1)2^{t+1} + t 2^{t+2} = 2(t-1)2^t + 2.$$

Substituting in the formula for $T(n)$ with $a = 2$ we get

$$
\begin{aligned}
T(n) &= 2^{\log_2 n} \cdot \left(1 + \sum_{k=1}^{\log_2 n} 2^k k\right) \\
&= n(3 + 2(\log_2 n - 1)2^{\log_2 n}) \\
&= 3n + 2(\log_2 n - 1)n^2
\end{aligned}
$$

(d) Derive a (messy) closed form for the summation when $a \neq 2$.
Using the formula

$$\sum_{i=1}^{t} ix^i = \frac{x - (t+1)x^{t+1} + tx^{t+2}}{(1-x)^2}$$

proved in class we get

$$\begin{aligned} T(n) &= a^{\log_2 n} \cdot \left(1 + \sum_{k=1}^{\log_2 n} (4/a)^k k\right) \\ &= a^{\log_2 n} \cdot \left(1 + \frac{(4/a) - (\log_2 n + 1)(4/a)^{\log_2 n + 1} + \log_2 n (4/a)^{\log_2 n + 2}}{(1-4/a)^2}\right) \end{aligned}$$

Optional Problem 11. "Finite Calculus" works with sums the way standard calculus works with integrals. Just like standard calculus has an inverse to the integral, namely the derivative, finite calculus has an inverse to summation, namely the *difference* operator

$$\Delta_x f(x) = f(x+1) - f(x).$$

We can modify the summation notation to be more like integration, writing

$$\Sigma_a^b g(x)\delta x = \sum_{x=a}^{b-1} g(x).$$

(a) Prove that $\Delta_x \Sigma_a^x f(y)\delta y = f(x)$ (analogue to $D \int_a^x f(y)dy = f(x)$).

$$\begin{aligned} \Delta_x \Sigma_a^x f(y)\delta y &= \Sigma_a^(x+1) f(y)\delta y - \Sigma_a^x f(y)\delta y \\ &= \sum_{y=a}^{x} g(y) - \sum_{y=a}^{x-1} g(y) \\ &= g(x) \end{aligned}$$

(b) Prove that $\Sigma_a^b(\Delta_x g(x))\delta x = f(b) - f(a)$ (analogue to $\int_a^b (Df(x))dx = f(b) - f(a)$)

$$\begin{aligned} \Sigma_a^b(\Delta_x g(x))\delta x &= \sum_{x=a}^{b-1}(g(x+1) - g(x)) \\ &= \sum_{x=a+1}^{b} g(x) - \sum_{x=a}^{b-1} g(x) \\ &= g(b) - g(a) \end{aligned}$$

(c) Define $x^{\underline{m}} = x(x-1)(x-2)\cdots(x-m+1)$. Prove that $\Delta_x x^{\underline{m}} = m x^{\underline{m-1}}$

$$\begin{aligned} \Delta_x x^{\underline{m}} &= (x+1)^{\underline{m}} - x^{\underline{m}} \\ &= (x+1)x(x-1)(x-2)\cdots(x+2-m) - x(x-1)(x-2)\cdots(x-m+1) \\ &= (x(x-1)(x-2)\cdots(x+1-m))\,((x+1) - (x-m+1)) \\ &= m x^{\underline{m-1}} \end{aligned}$$

(d) Prove that $\sum_0^n x^{\underline{m}}\,\delta x = \frac{n^{\underline{m+1}}}{m+1}$ Easily follows from part (b) and part (c):

$$\begin{aligned}
\Sigma_0^n x^{\underline{m}}\delta x &= \Sigma_0^n \Delta_x \frac{x^{\underline{m+1}}}{m+1}\delta x \\
&= \frac{n^{\underline{m+1}}}{m+1} - \frac{0^{\underline{m+1}}}{m+1} \\
&= \frac{n^{\underline{m+1}}}{m+1}
\end{aligned}$$