1. (15) This is a straightforward exercise on orders. An order relation $\prec$ defined on a set $P$ is a strict partial order if
   - the relation $p \prec p$ never holds (i.e., it is non-reflexive), and
   - $p \prec q \land q \prec r \Rightarrow p \prec r$ (i.e., it is transitive).

Define an order $\preceq$ on $P$ by letting $p \preceq q$ if either $p \prec q$ or $p = q$. Show that $\preceq$ is a partial order (i.e., its reflexive, antisymmetric, and transitive).

2. (35) We have seen an integer programming model for the longest common subsequence (LCS) problem in class (also see the handout (relevant pages from the book by Pevzner) for details). Implement this IP in AMPL for the instance discussed in class, where we sought the LCS between $v = ATCTGAT$ and $w = TGCTATA$.

3. (10) We have seen the tropical cubic polynomial in class:
   \[ p(x) = a \circ x^3 \oplus b \circ x^2 \oplus c \circ x \oplus d. \]

Assuming $b - a \leq c - b \leq d - c$, show that we can factorize $p(x)$ as follows:
   \[ p(x) = a \circ (x \oplus (b - a)) \circ (x \oplus (c - b)) \circ (x \oplus (d - c)). \]

Thus, the “roots” of the cubic polynomial are $b - a$, $c - b$, and $d - c$, which are exactly the points at which $p(x)$ is not linear (these are the points of the piecewise linear function at which the slopes change).

4. (15) Just as discussed in the previous problem, a tropical polynomial $p : \mathbb{R}^n \to \mathbb{R}$ in general can be expressed as the minimum of a finite set of linear functions. The roots of $p$ are then defined as the points where the (piecewise linear) function is not linear. The set of all such roots defines the hypersurface $\mathcal{H}(p)$ of the polynomial $p$. Consider the tropical line in 2D given as:
   \[ p(x, y) = a \circ x \oplus b \circ y \oplus c, \text{ where } a, b, c \in \mathbb{R}. \]

Plot the hypersurface $\mathcal{H}(p)$ (i.e., plot all the points where the given function is not linear).

5. (30) Consider the parametric sequence alignment problem between two sequences $v$ and $w$ of the same length $n$, where a match gets a score of 1, a mismatch is penalized by $\mu$, and an indel is penalized by $\sigma/2$. Since the lengths are same, we must have that # insertions = # deletions = $1/2$ # indels. Also, # matches + # mismatches + # insertions = $n$.

For an alignment with $x$ matches, $y$ mismatches, and $z$ insertions, the total score will be $x - y\mu - z\sigma = n - y(\mu + 1) - z(\sigma + 1)$, as $x + y + z = n$. The goal of this exercise to find the optimal alignments for all values of $\mu \geq 0$ and $\sigma \geq 0$.

Modify the MATLAB code you wrote for LCS (Hw4) so that it now finds the optimal global sequence alignment (i.e., just include the penalties for mismatches and indels). Consider the alignment between $v = TGTCCTTCCGGG$ and $w = ACCTCCTTCCG$ (both of length $n = 12$). Find the optimal alignment for various values of the parameter pair $(\mu, \sigma)$, and plot the regions (representing values of $(\mu, \sigma)$) for which a particular alignment is optimal.