1. (a) Describe a linear program whose solution (a, b) describes the line with minimum L_1 error.

Solution:

$$\boxed{\begin{array}{lll} \text{minimize} & \displaystyle\sum_{i=1}^{n} R_i & \\ \text{subject to} & ax_i + b + R_i \le y_i & \text{for all } i \\ & ax_i + b - R_i \ge y_i & \text{for all } i \end{array}}$$

For each index i, the two constraints involving x_i and y_i are equivalent to the non-linear inequality

$$R_i \ge |y_i - ax_i - b|.$$

If we fix the variables a and b, the objective function $\sum_i R_i$ is minimized by setting $R_i = |y_i - ax_i - b|$ for every i. Thus, in any optimal solution, the objective function is the L_1 error of the line $y = ax + b$, as required. ■

(b) Describe a linear program whose solution (a, b) describes the line with minimum L_∞ error.

Solution:

$$\boxed{\begin{array}{lll} \text{minimize} & R & \\ \text{subject to} & ax_i + b + R \le y_i & \text{for all } i \\ & ax_i + b - R \ge y_i & \text{for all } i \end{array}}$$

For each index i, the two constraints involving x_i and y_i are equivalent to the non-linear inequality

$$R \ge |y_i - ax_i - b|.$$

Thus, the constraints are collectively equivalent to the inequality

$$R \ge \max_i |y_i - ax_i - b|.$$

For any fixed values of a and b, the variable R is obviously minimized when $R = \max_i |y_i - ax_i - b|$. Thus, in any optimal solution, the objective function is the L_∞ error of the line $y = ax + b$, as required. ■

2. (a) Give a linear-programming formulation of the maximum-cardinality bipartite matching problem.

Solution: The following linear program is a simple modification of the maximum-flow linear program described in the lecture notes. The variable x_{uv} is an indicator variable that equals 1 if edge uv is in the matching and 0 otherwise.

$$\boxed{\begin{array}{lrll} \text{maximize} & \displaystyle\sum_{uv\in E} x_{uv} & & \\ \text{subject to} & \displaystyle\sum_{uv\in E} x_{uv} & \le 1 & \text{for every vertex } u \in L \\ & \displaystyle\sum_{uv\in E} x_{uv} & \le 1 & \text{for every vertex } v \in R \\ & x_{uv} & \ge 0 & \text{for every edge } uv \end{array}}$$

Formally, to complete the proof that this linear-programming formulation is correct, we need to show that there is an optimal solution in which every variable x_{uv} is either 0 or 1. This property follows easily from the *maximum-flow* formulation of the matching problem! Direct the edges of G from L to R, add a source vertex s with edges to L, and add a target vertex t with edges from R. Then we can interpret x_{uv} as the flow through $u\to v$; the objective function is the value of the flow; the first constraint implies that $s\to u$ has capacity 1; the second constraint implies that $v\to t$ has capacity 1; and the final constraint implies all flow values are non-negative. Any integer flow network has an integer maximum flow. ■

(b) Now dualize the linear program from part (a). What do the dual variables represent? What does the objective function represent? What problem is this!?

Solution: The dual linear program includes a variable y_v for every vertex v and an inequality constraint for every edge.

$$\boxed{\begin{array}{lrl} \text{minimize} & \displaystyle\sum_{u\in L} y_u + \sum_{v\in R} y_v & \\ \text{subject to} & y_u + y_v \ge 1 & \text{for every edge } uv \in E \\ & y_u \ge 0 & \text{for vertex } u \in L \\ & y_v \ge 0 & \text{for vertex } v \in R \end{array}}$$

The constraints imply that each vertex variable y_u is between 0 and 1. Suppose we interpret y_v as an indicator variable; say that a vertex is *marked* if $y_v = 1$. The linear constraints imply that that every edge must have at least one marked endpoint. The objective function is the number of marked vertices. This is the ***minimum vertex cover*** problem!

This is enough for full credit. However, to prove this answer is correct, we must argue that the linear program has an *integer* optimum solution. The optimal *objective value* is equal to the optimal objective value of the primal LP, which is an integer, but that does not immediately imply that there is an optimal solution where every variable y_u is an integer.

Perhaps the easiest method[1] to show that the optimal solution to this linear program is integral uses an important property of linear programs called ***complementary slackness***. Recall that the strong duality theorem states that if x^* is the optimal solution to a linear program, there is an optimal solution y^* for the dual linear program, such that $c \cdot x^* = y^* A x^* = y^* \cdot b$. These equations have two immediate consequences:

- For any index i, if $y_i^* > 0$, then $a_i \cdot x^* = b$, where a_i denotes the ith row of A.
- For any index j, if $x_j^* > 0$, then $y^* \cdot a^j = c$, where a^j denotes the jth column of A.

More colloquially:

- If a primal variable is positive, the corresponding dual constraint is tight.
- If a dual variable is positive, the corresponding primal constraint is tight.

Equivalently, in the contrapositive:

- If a dual constraint is loose, the corresponding primal variable is zero.
- If a primal constraint is loose, the corresponding dual variable is zero.

Let x^* be the solution to the primal linear program that describes a maximum matching M, and let y^* be the optimum dual solution guaranteed by complementary slackness. Then we can immediately make two important observations:

- For each edge $uv \in M$, we have $x_{uv}^* = 1 > 0$ (the primal variable is positive) and therefore $y_u^* + y_v^* = 1$ (the corresponding dual constraint is tight).
- For each vertex v that is *not* incident to an edge in M, we have $\sum_{uv \in E} x_{uv}^* = 0 < 1$ (the primal constraint is loose) and therefore $y_v^* = 0$ (the corresponding dual variable is zero).

Now define a new solution vector $\tilde{y}$ by ***rounding*** each coordinate of y^* to the nearest integer, rounding up on the left and rounding down on the right. That is, for each vertex $u \in L$, define $\tilde{y}_u := \lceil y_u^* - 1/2 \rceil$, and for each vertex $v \in R$, define $\tilde{y}_v := \lfloor y_u^* + 1/2 \rfloor$. We make several easy observations:

- For every edge uv, we have $y_u^* + y_v^* \geq 1$, and therefore either $y_u^* \geq 1/2$ or $y_v^* > 1/2$. Thus, our rounding rules imply either $\tilde{y}_u = 1$ or $\tilde{y}_v = 1$, which implies $\tilde{y}_u + \tilde{y}_v \geq 1$.
- For every vertex v, the constraint $y_v^* \geq 0$ implies $\tilde{y} \geq 0$.
- The endpoints of each edge in the matching are rounded by exactly the same amount, but in opposite directions; in particular, we still have $\tilde{y}_u + \tilde{y}_v = 1$ for each matching edge uv.
- For each vertex v that is not incident to an edge in M, we have $\tilde{y}_v = y_v^* = 0$.

The first two observations imply that $\tilde{y}$ is a *feasible* solution to our dual LP, and the last two observations imply

$$\sum_{u \in L} \tilde{y}_u + \sum_{v \in R} \tilde{y}_v \;=\; |M| \;=\; \sum_{u \in L} y_u^* + \sum_{v \in R} y_v^*.$$

In other words, rounding y^* to $\tilde{y}$ maintains feasibility but does not change the objective value. We conclude that our dual LP has an integer optimal solution $\tilde{y}$. ■

[1]. . . to explain from scratch. The statement that the size of the largest matching and the smallest vertex cover are equal in bipartite graphs is called König's theorem; see Wikipedia for a proof in terms of the alternating-path formulation of maximum matchings. There are other relatively simple arguments using Hall's Marriage Theorem or the total unimodularity of the constraint matrix, but those would take more space to describe.

3. Prove that deciding whether a given integer program has a feasible solution is NP-hard.

Solution (3Color): We prove that integer programming is NP-hard by reduction from 3Color. Let $G = (V, E)$ be an arbitrary undirected graph. We define an integer program with $3V$ variables and $3E$ constraints as follows. For every vertex v, the variables r_v, g_v, and b_v indicate whether v is colored red, green, or blue.

$$
\begin{array}{lrl}
\text{maximize} & 0 & \\
\text{subject to} & r_u + r_v \le 1 & \text{for every edge } uv \\
& g_u + g_v \le 1 & \text{for every edge } uv \\
& b_u + b_v \le 1 & \text{for every edge } uv \\
& r_v + g_v + b_v = 1 & \text{for every vertex } v \\
& r_v, g_v, b_v \ge 0 & \text{for every vertex } v \\
& r_v, g_v, b_v \in \mathbb{Z} & \text{for every vertex } v
\end{array}
$$

We claim that G is 3-colorable if and only if this integer program has a feasible solution. (The objective function is irrelevant.)

- Suppose G is 3-colorable. Fix a proper 3-coloring using the colors red, green, and blue. For each vertex v, set

$$
\begin{aligned}
r_v &= [v \text{ is red}] \\
g_v &= [v \text{ is green}] \\
b_v &= [v \text{ is blue}]
\end{aligned}
$$

  Because each vertex gets exactly one of these three colors, we have $r_v + g_v + b_v = 1$ for every vertex v. Because the endpoints of each edge have different colors, the other inequalities are also satisfied. We conclude that the integer program has a feasible solution.

- Suppose the integer program has a feasible solution. The last three constraints implies that every variable is either 0 or 1. For each vertex v, color v red if $r_v = 1$, green if $g_v = 1$, and blue if $b_v = 1$; the constraint $r_v + g_v + b_v = 1$ implies the every vertex has a well-defined color. The first three constraints imply that every edge has endpoints of two different colors. We conclude that G is 3-colorable.

We can clearly construct the integer program from G in polynomial time. ∎

Solution (MinVertexCover): We prove that integer programming is NP-hard by reduction from MinVertexCover. Let $G = (V, E)$ be an arbitrary undirected graph. We define an integer program with one variable y_v for each vertex v and one inequality constraint for each edge.

$$
\begin{array}{lrl}
\text{minimize} & \displaystyle\sum_{v \in V} y_v & \\
\text{subject to} & y_u + y_v \ge 1 & \text{for every edge } uv \in E \\
& y_v \ge 0 & \text{for vertex } v \in V \\
& y_v \in \mathbb{Z} & \text{for vertex } v \in V
\end{array}
$$

The constraints imply that each variable y_u is either 0 or 1; thus, every feasible solution to the integer program specifies a subset of the vertices. The edge constraints imply that at most one endpoint of each edge belongs to this subset; thus, every feasible solution describes a vertex cover of G. Conversely, if C is any vertex cover, then setting $y_v = 1$ for all $v \in C$ and $y_v = 0$ for all $v \notin C$ gives us a feasible solution to the integer program. The objective value is the size of the vertex cover. We conclude that the optimal solution to this integer program is the minimum vertex cover.

This solution is almost identical to the solution to problem 2(b). When the graph G is bipartite, the integrality constraints $y_v \in \mathbb{Z}$ are redundant; the *linear* program has *integral* optimal solutions. Thus, when the input graph is bipartite, we can solve the minimum vertex cover problem in polynomial time! But this property is specific to bipartite graphs; if the graph G is not bipartite, the optimal LP solutions may not be integral. For example, if G is a triangle, the optimal LP solution assigns value $1/2$ to every vertex, and therefore has value $3/2$, but the smallest vertex cover has size 2. (Similarly, the optimal solution to the matching LP assigns $1/2$ to every edge, but the largest matching has size 1.) The rounding argument from our solution to 2(b) doesn't work, because we don't have a consistent method to round $1/2$! ■

Rubric: Standard NP-hardness rubric (see Homework 1). These are not the only correct solutions.