Inspire…Educate…Transform.

Effective Decision Making: Optimization Simulation and Statistical Methods

Dr. K. V. Dakshinamurthy

President, International School of Engineering

Types of models

	Model Characteristics		
Category	**Form of $f(\cdot)$**	**Values of Independent Variables**	**Management Science Techniques**
Predictive Models	unknown, ill-defined	known or under decision maker's control	Regression Analysis, Time Series Analysis, Discriminant Analysis
Descriptive Models	known, well-defined	unknown or uncertain	Simulation, Queuing, PERT, Inventory Models
Prescriptive Models	known, well-defined	known or under decision maker's control	Linear Programming, Networks, Integer Programming, CPM, Goal Programming, EOQ, Nonlinear Programming

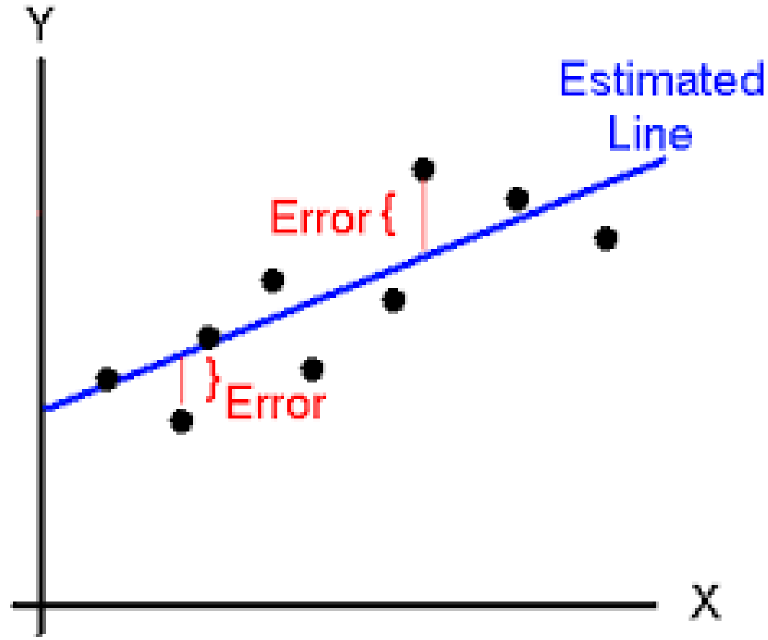
Machine Learning: Generalizing From Data via Optimization

Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

Parameters: θ_0, θ_1

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

Goal: $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$



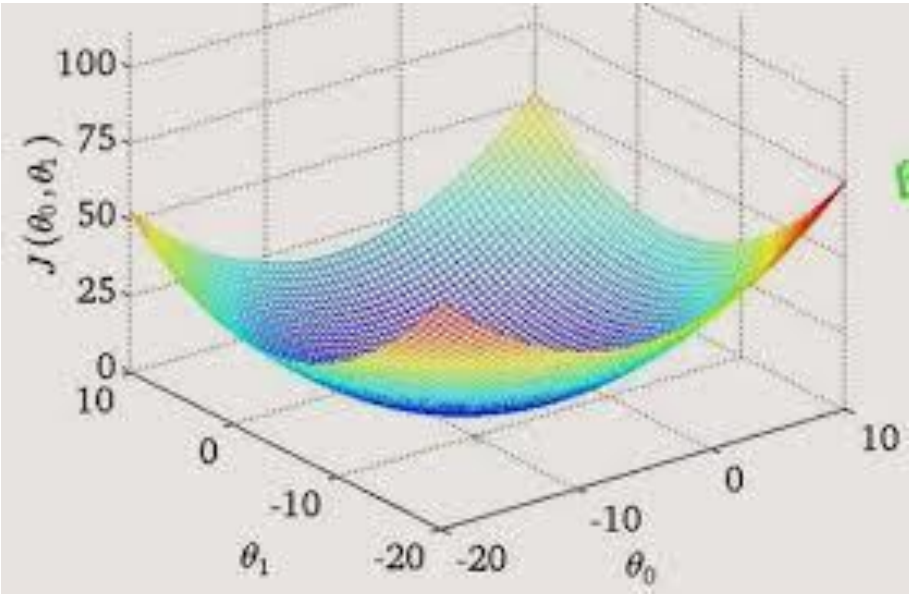
Multiple models/ Hypothesis tried. Which is correct?

“The purpose of models is not to fit the data but to sharpen the questions”

Examples of Optimization in ML

- Optimal cluster centers in k-means
- Training coefficients for regression
- Optimal smoothing parameters in forecasting
- Many data mining problems can be reformulated as optimization problems

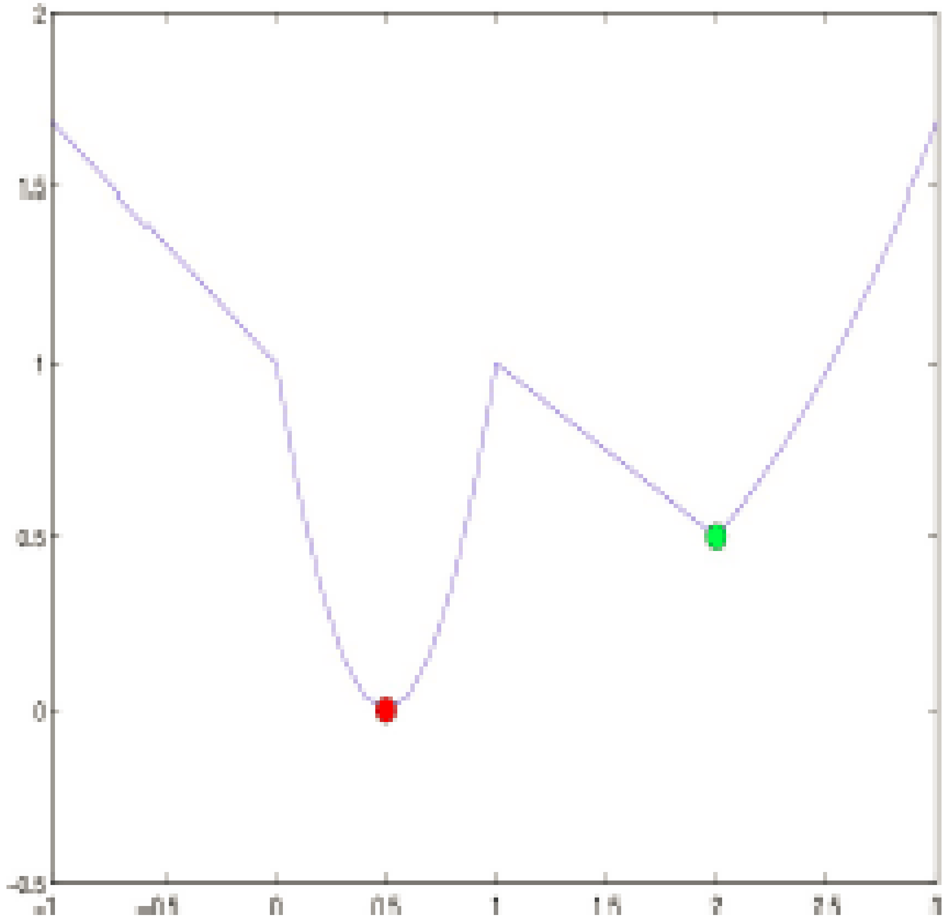
Most problems have 'Well-Behaved' Cost functions



- Closed form solution exists
- Differentiation!!
- 'Convex Optimization'

CSE 7213c

Local versus Global



Goals for today

- Recognize the optimization problem
- Setting up systematically
- Solving and analyzing using R
- Recognizing that linear optimization is relevant across several domains

Optimization

$$\text{Optimize } z = f(x_1, x_2, \dots x_n)$$

$$\text{Subject to } g_1(x_1, x_2, \dots x_n) \leq \textit{or} \geq \textit{or} = b_1$$

$$g_2(x_1, x_2, \dots x_n) \leq \textit{or} \geq \textit{or} = b_2$$

$$g_3(x_1, x_2, \dots x_n) \leq \textit{or} \geq \textit{or} = b_3$$

When Z and g are linear: Linear programming
When Z is quadratic and g is linear; Quadratic programming

LP

The problem is usually expressed in matrix form and then it becomes:

$$\begin{array}{ll} \text{Maximize} & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} & A\mathbf{x} \le \mathbf{b}, \mathbf{x} \ge 0 \end{array}$$

where A is a $m \times n$ matrix.

Process of optimizing

- Identify and name the decision variables consistently

- Mathematically define the objective/fitness function in terms of the variables

Process…

- Identify all stipulated requirements, restrictions and limitations

- Express any hidden constraints (generally non-negative or integer only like constraints)

- Identify the class of optimization it belongs to
- Pick the solution method

Case 1

Blue Ridge Hot Tubs manufactures and sells two models of hot tubs: the Aqua-Spa and the Hydro-Lux.

Howie Jones, the owner and manager of the company, needs to decide how many of each type of hot tub to produce during his next production cycle.

Howie buys prefabricated fiberglass hot tub shells from a local supplier and adds the pump and tubing to the shells to create his hot tubs. (This supplier has the capacity to deliver as many hot tub shells as Howie needs.)

Howie installs the same type of pump into both hot tubs. He will have only 200 pumps available during his next production cycle.

From a manufacturing standpoint, the main difference between the two models of hot tubs is the amount of tubing and labor required. Each Aqua-Spa requires 9 hours of labor and 12 feet of tubing. Each Hydro-Lux requires 6 hours of labor and 16 feet of tubing.

Howie expects to have 1,566 production labor hours and 2,880 feet of tubing available during the next production cycle.

Howie earns a profit of $350 on each Aqua-Spa he sells and $300 on each Hydro-Lux he sells.

He is confident that he can sell all the hot tubs he produces. The question is, how many Aqua-Spas and Hydro-Luxes should Howie produce if he wants to maximize his profits during the next production cycle

Identify the decision variables

- How many Aqua-Spas and Hydro-Luxes should be produced?
  - We will let X1 represent the number of Aqua-Spas to produce and X2 represent the number of Hydro-Luxes to produce.

State the objective function as a linear combination of the decision variables

- Howie earns a profit of $350 on each Aqua-Spa (X1) he sells and $300 on each Hydro-Lux (X2) he sells.

- Howie's objective of maximizing the profit he earns is stated mathematically as:

MAX: $350X_1 + 300X_2$

State the constraints as linear combinations of the decision variables.

- Only 200 pumps are available and each hot tub requires one pump

$$1X1 + 1X2 \le 200$$

- He has only 1,566 labor hours available during the next production cycle. Each Aqua-Spa he builds (each unit of X1) requires 9 labor hours and each Hydro-Lux (each unit of X2) requires 6 labor hours

$$9X1 + 6X2 \leq 1{,}566$$

- Each Aqua-Spa requires 12 feet of tubing, and each Hydro-Lux produced requires 16 feet of tubing

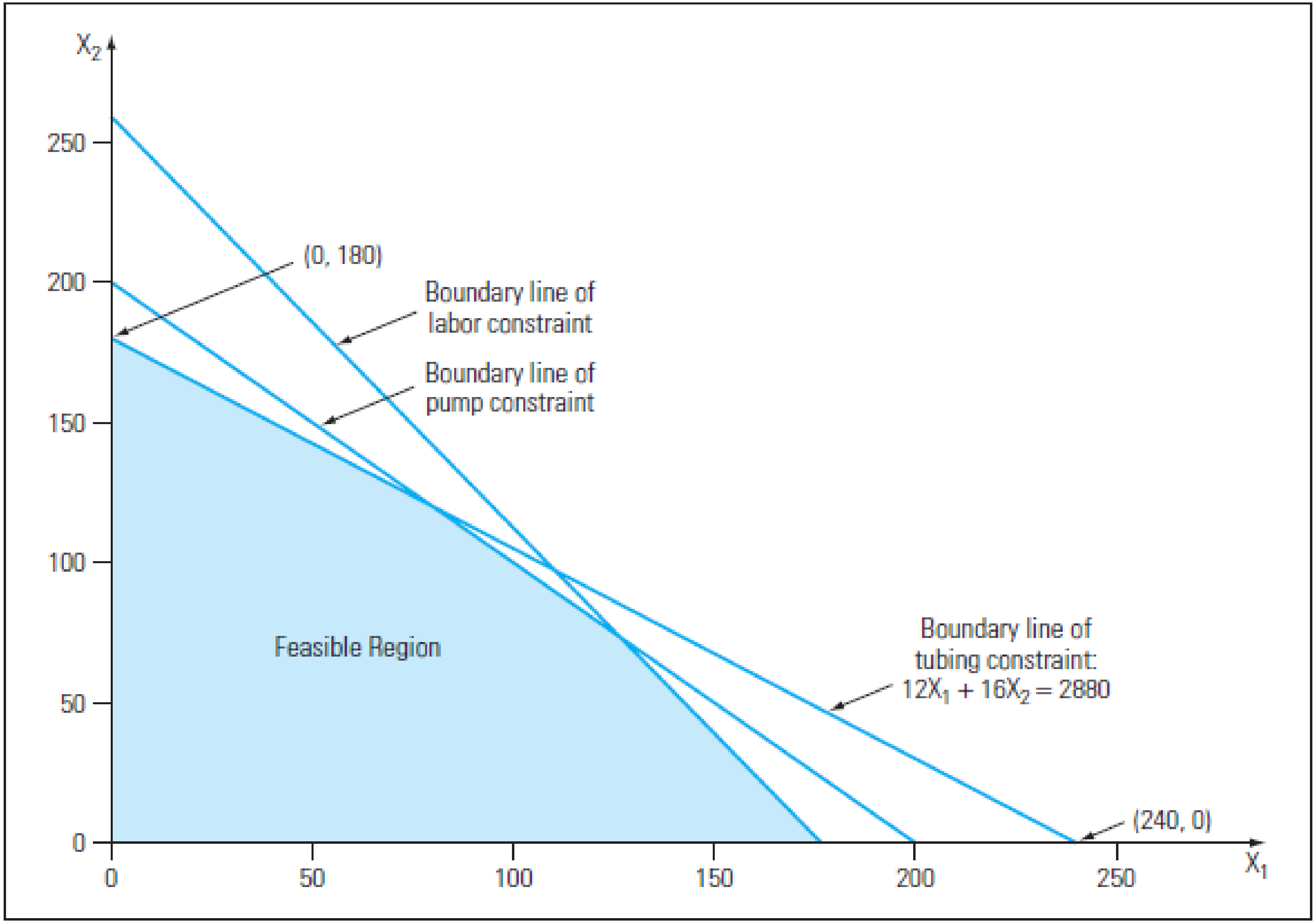
$$12X1 + 16X2 \leq 2,880$$

Hidden constraints

- There are simple lower bounds of zero on the variables X1 and X2 because it is impossible to produce a negative number of hot tubs.

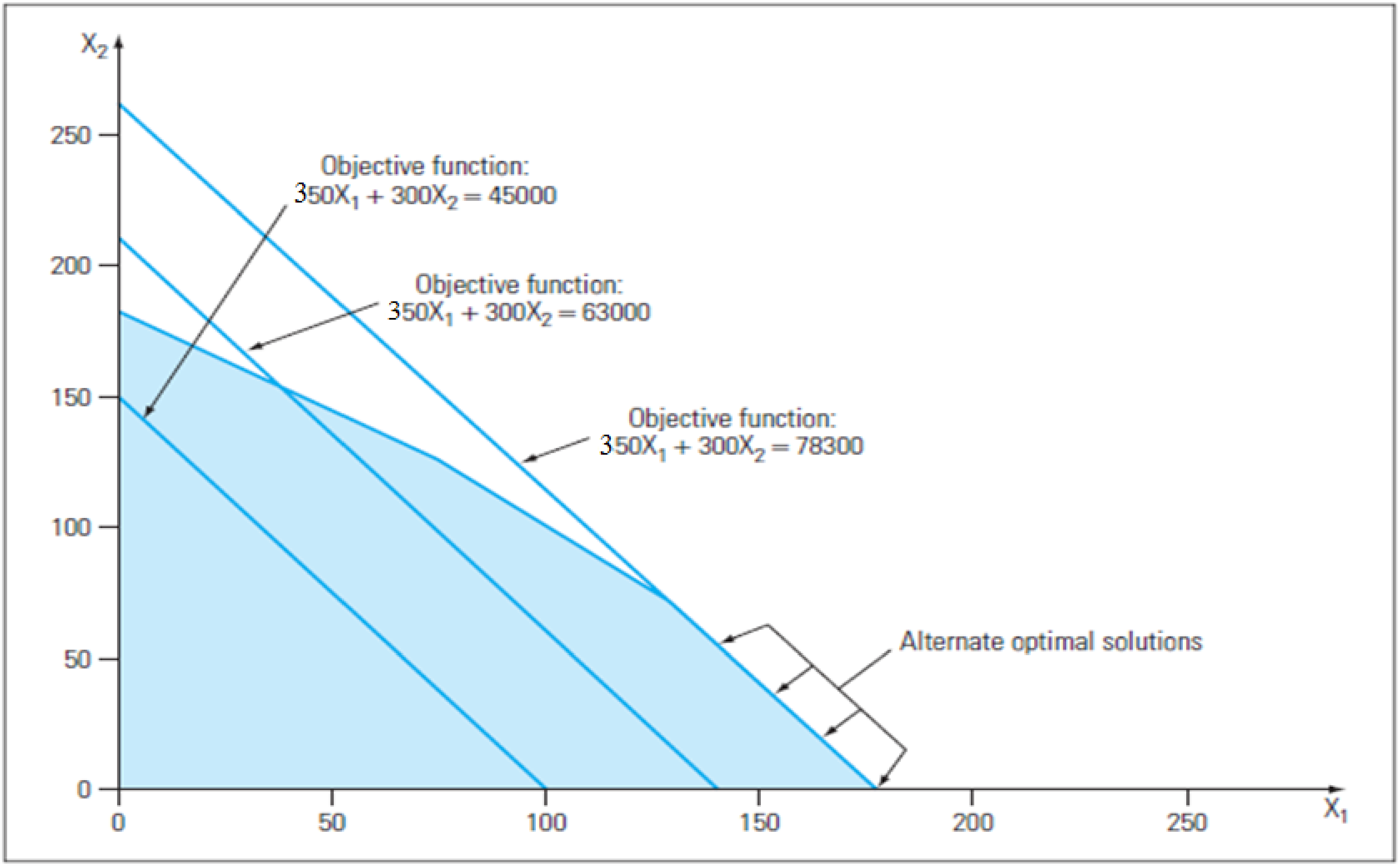
$$X1 \geq 0; X2 \geq 0$$

$$
\begin{array}{llcl}
\text{MAX:} & 350X_1 + 300X_2 & & \\
\text{Subject to:} & 1X_1 + 1X_2 & \leq & 200 \\
 & 9X_1 + 6X_2 & \leq & 1{,}566 \\
 & 12X_1 + 16X_2 & \leq & 2{,}880 \\
 & 1X_1 & \geq & 0 \\
 & 1X_2 & \geq & 0
\end{array}
$$



X_2
250
200
150
100
50
0
(0, 180)
Boundary line of
labor constraint
Boundary line of
pump constraint
Feasible Region
Boundary line of
tubing constraint:
$12X_1 + 16X_2 = 2880$
(240, 0)
0
50
100
150
200
250
X_1

CSE 7213c


X_2
250
200
150
100
50
0
Objective function:
$350X_1 + 300X_2 = 45000$
Objective function:
$350X_1 + 300X_2 = 63000$
Objective function:
$350X_1 + 300X_2 = 78300$
Alternate optimal solutions
0
50
100
150
200
250
X_1

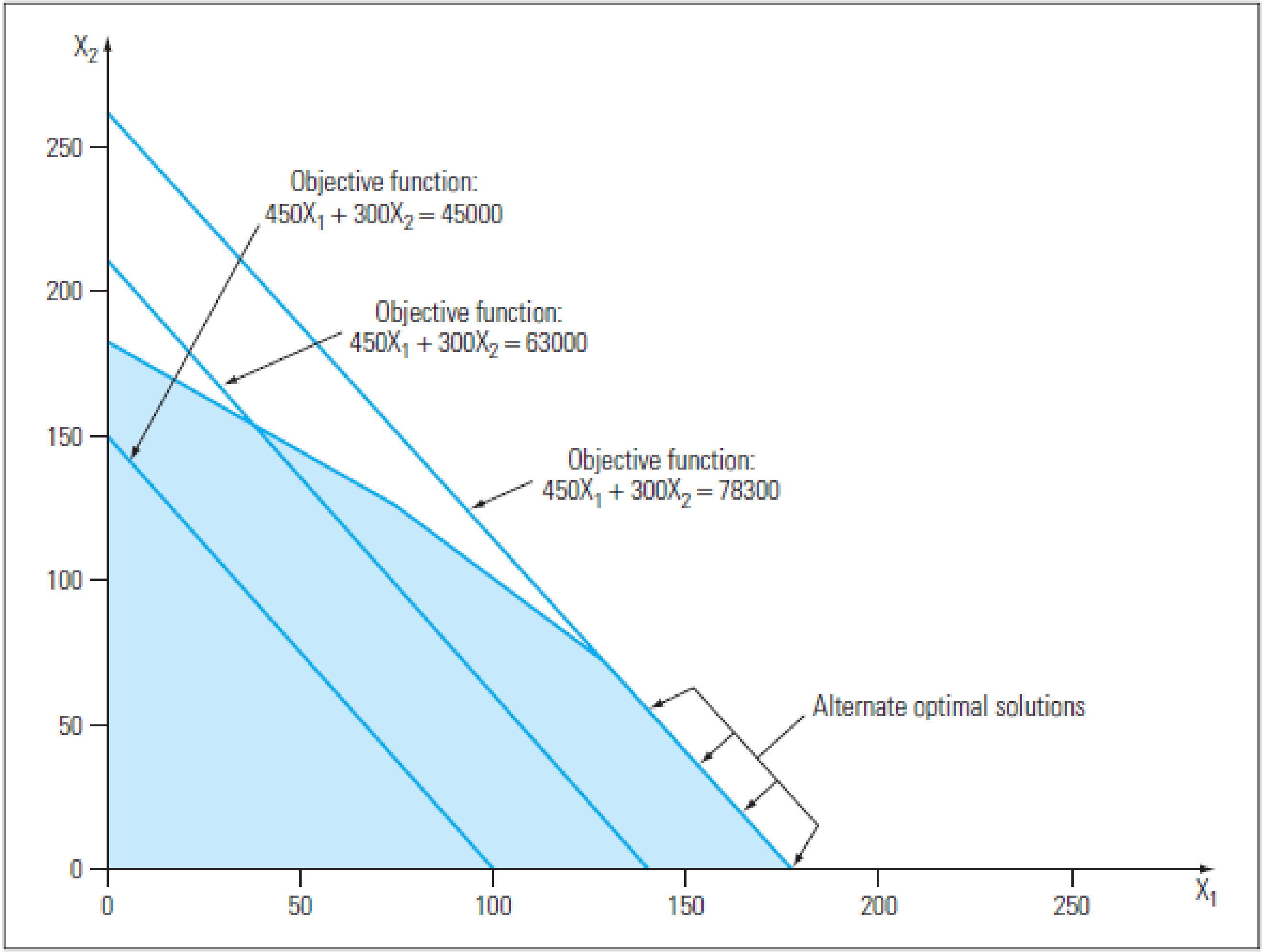
Corner points

- If an LP problem has a finite optimal solution, this solution always will occur at some corner point of the feasible region.
  - Identify all the corner points of the feasible region and calculate the objective function at each of them

Corner point

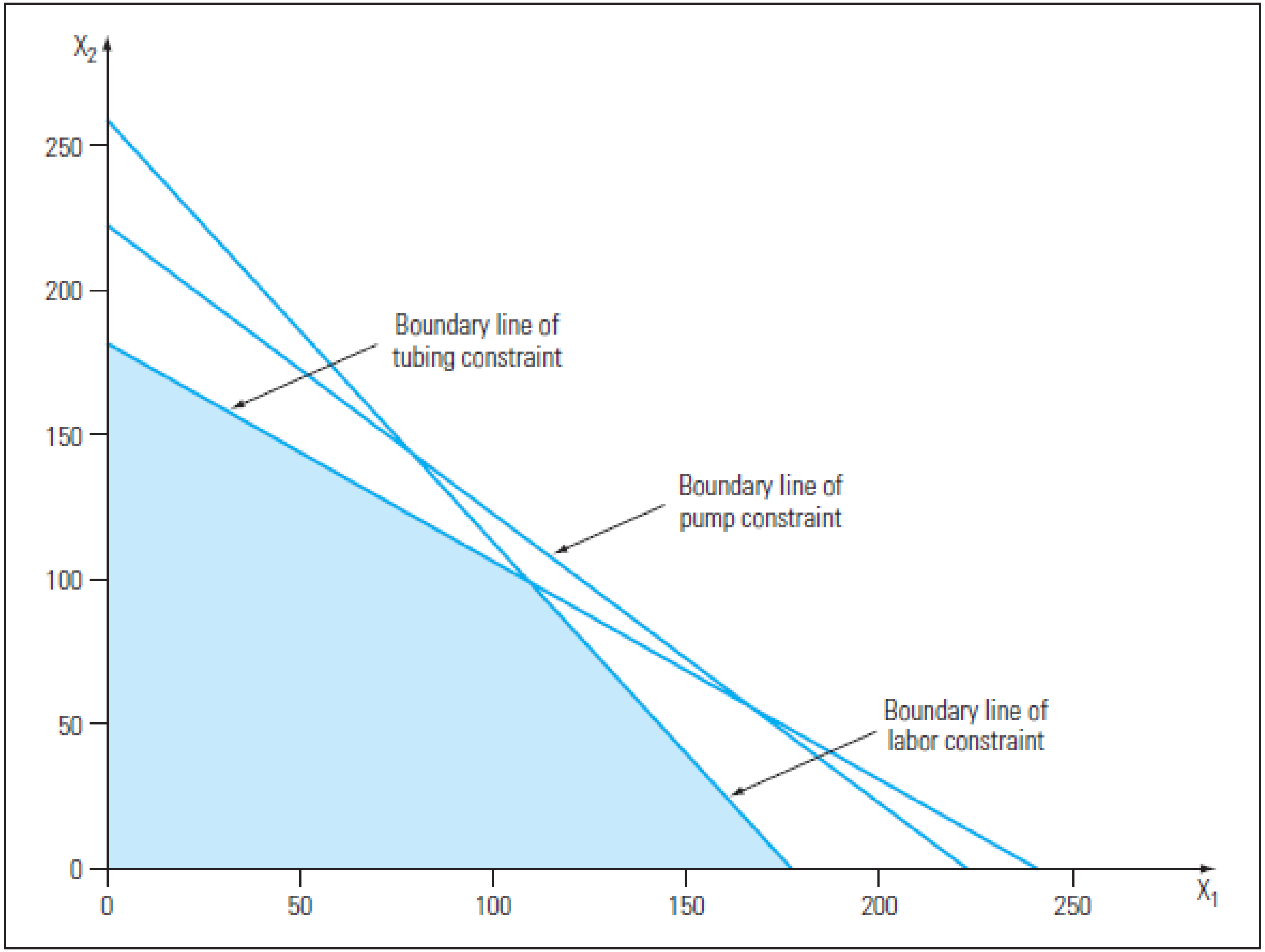
- The corner point with the largest objective function value is the optimal solution to the problem.

Special cases: Alternate solutions



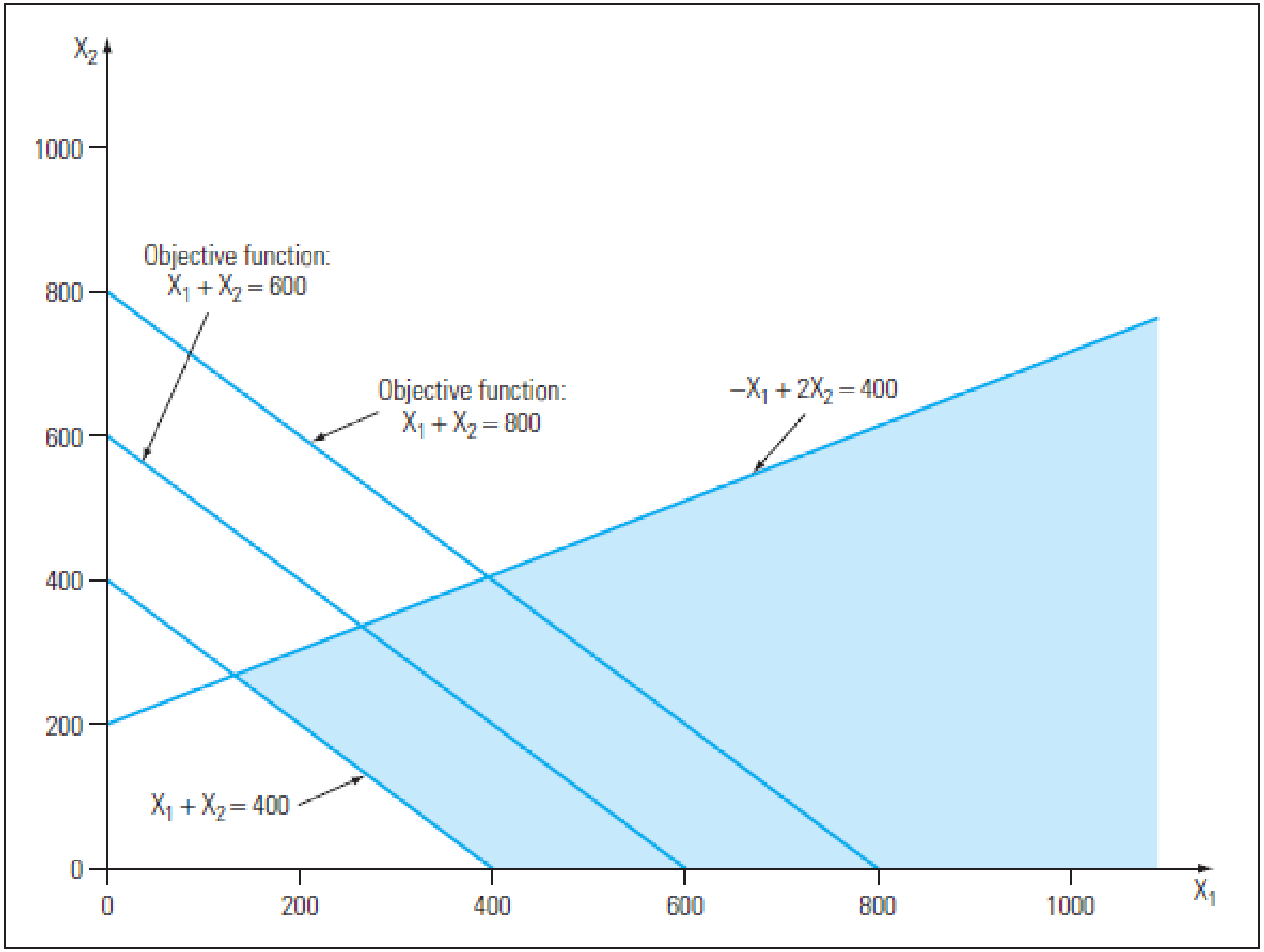
CSE 7213c

Redundancy



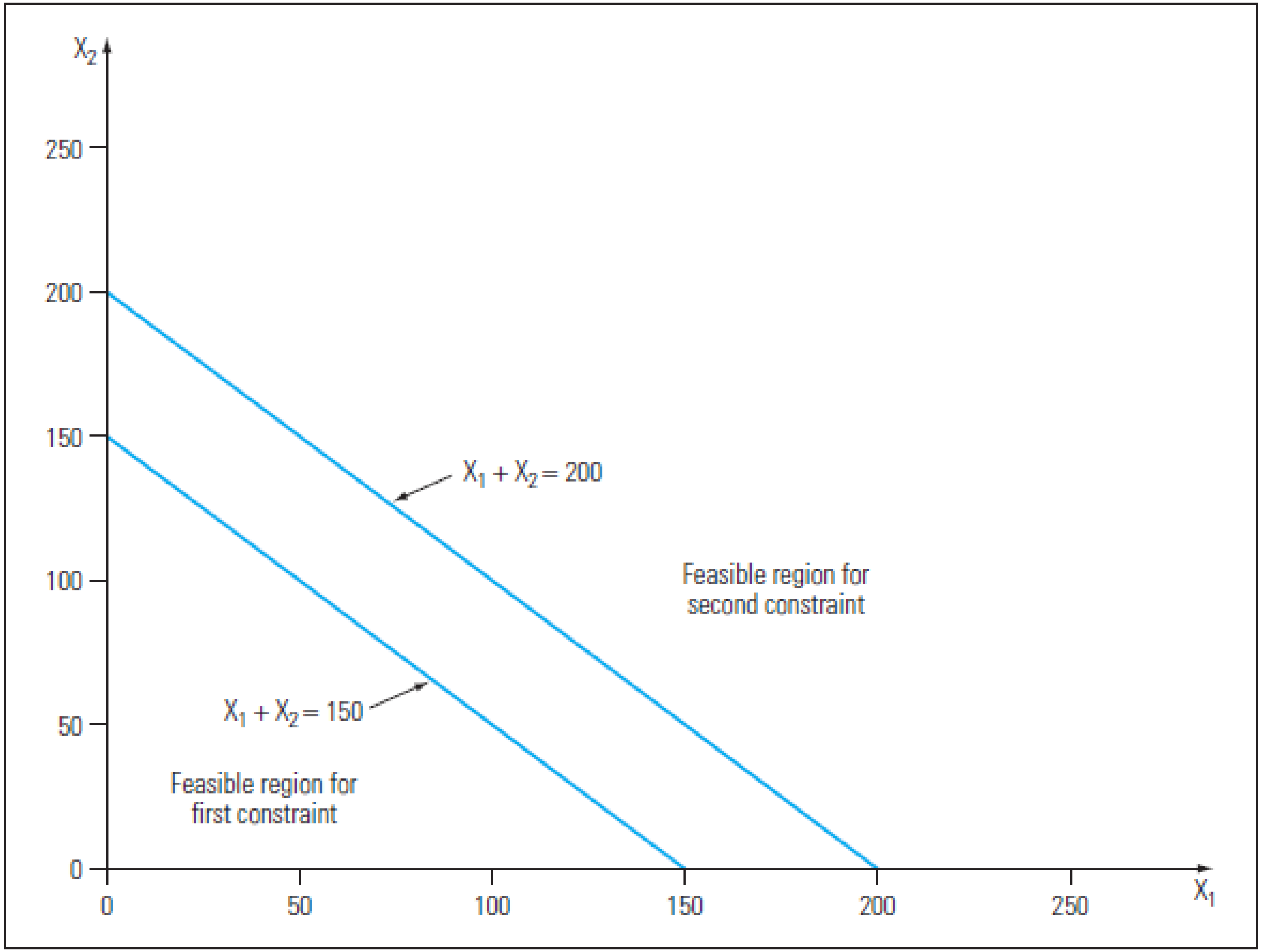
CSE 7213c

Unbounded



CSE 7213c

Not feasible



CSE 7213c

Duality and Sensitivity analysis

- http://optlab-server.sce.carleton.ca/POAnimations2007/Sensitivity.html

Sensitivity analysis: A simpler problem

Variable non-negativity: $x_1 \geq 0, \quad x_2 \geq 0$

Objective Function:

Maximize daily profit: $\text{MAX } z = 15x_1 + 10x_2$

Constraints:

Mountain bike production limit: $x_1 \leq 2$

Racer production limit: $x_2 \leq 3$

Metal finishing machine production limit: $x_1 + x_2 \leq 4$

$$\begin{array}{llcll} \text{maximize} & c^T x & & \text{minimize} & (b+\Delta)^T y \\ \text{subject to} & Ax \leq b + \Delta & \text{and} & \text{subject to} & A^T y \geq c \\ & x \geq 0 & & & y \geq 0. \end{array}$$

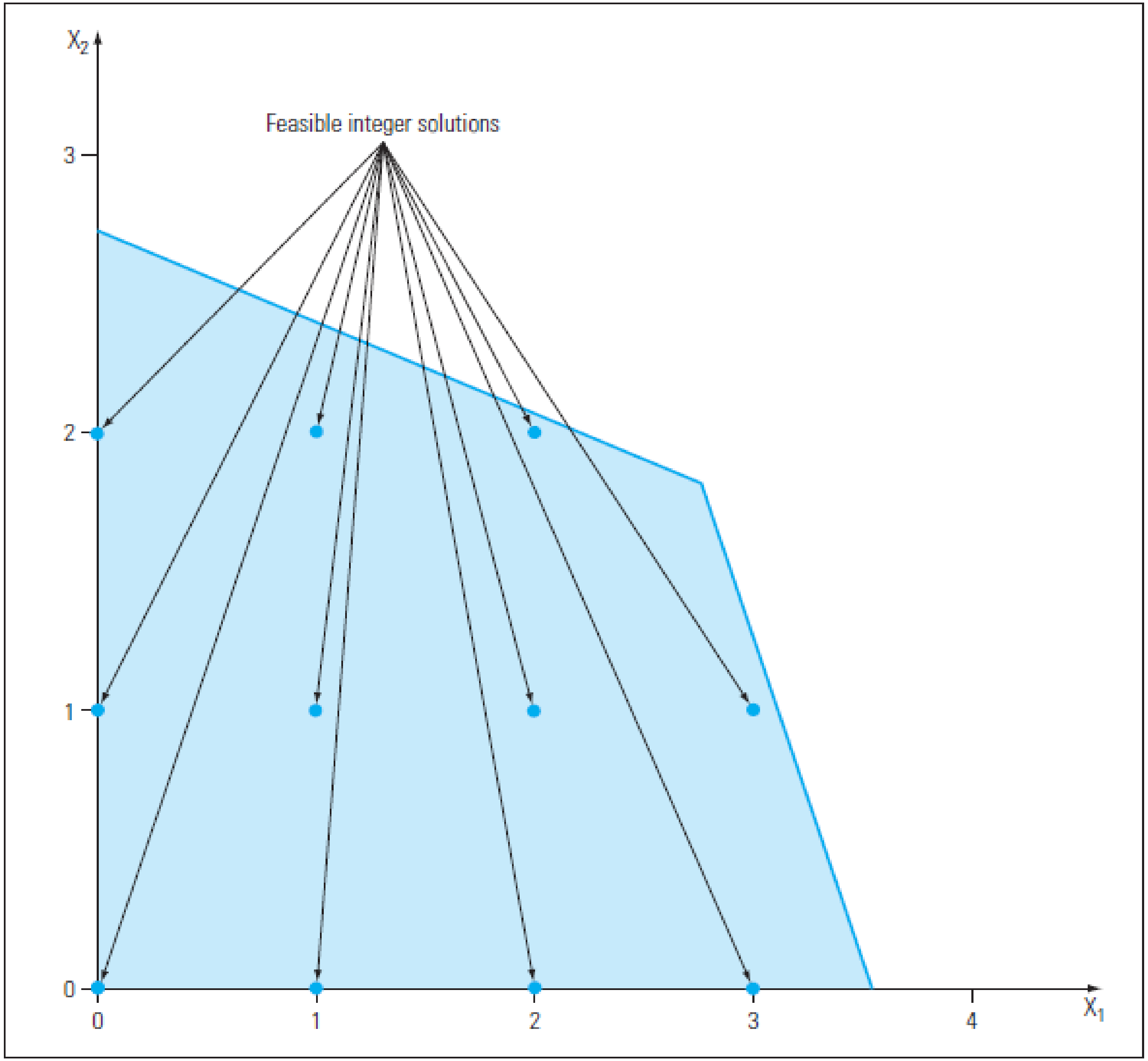
Linear programming problems are optimization problems in which the objective function and the constraints are all linear. In the primal problem, the objective function is a linear combination of *n* variables. There are *m* constraints, each of which places an upper bound on a linear combination of the *n* variables. The goal is to maximize the value of the objective function subject to the constraints. A *solution* is a vector (a list) of *n* values that achieves the maximum value for the objective function.

In the dual problem, the objective function is a linear combination of the *m* values that are the limits in the *m* constraints from the primal problem. There are *n* dual constraints, each of which places a lower bound on a linear combination of *m* dual variables.

INTEGER AND BINARY PROGRAMMING

- Suppose, for example, that Blue Ridge Hot Tubs has only 1,520 hours of labor and 2,650 feet of tubing available during its next production cycle. The company might be interested in solving the following ILP problem

CSE 7213c


X_2
Feasible integer solutions
3
2
1
0
0
1
2
3
4
X_1

$$
\begin{array}{lrcrcrl}
\text{MAX:} & 350X_1 & + & 300X_2 & & & \}\ \text{profit} \\
\text{Subject to:} & 1X_1 & + & 1X_2 & \leq & 200 & \}\ \text{pump constraint} \\
 & 9X_1 & + & 6X_2 & \leq & 1{,}520 & \}\ \text{labor constraint} \\
 & 12X_1 & + & 16X_2 & \leq & 2{,}650 & \}\ \text{tubing constraint} \\
 & \multicolumn{5}{l}{X_1, X_2 \geq 0} & \}\ \text{nonnegativity conditions} \\
 & \multicolumn{5}{l}{X_1, X_2 \text{ must be integers}} & \}\ \text{integrality conditions}
\end{array}
$$

Capital budget allocation

- In his position as vice president of research and development (R&D) for CRT Technologies, Mark Schwartz is responsible for evaluating and choosing which R&D projects to support. The company received 18 R&D proposals from its scientists and engineers, and identified six projects as being consistent with the company’s mission.

CSE 7213c

- However, the company does not have the funds available to undertake all six projects. Mark must determine which of the projects to select. The funding requirements for each project are summarized in the following table along with the NPV the company expects each project to generate.

Project	Expected NPV (in $1,000s)	Capital (in $1,000s) Required in Year 1	Year 2	Year 3	Year 4	Year 5
1	$141	$ 75	$25	$20	$15	$10
2	$187	$ 90	$35	$ 0	$ 0	$30
3	$121	$ 60	$15	$15	$15	$15
4	$ 83	$ 30	$20	$10	$ 5	$ 5
5	$265	$100	$25	$20	$20	$20
6	$127	$ 50	$20	$10	$30	$40

CSE 7213c

- The company currently has $250,000 available to invest in new projects. It has budgeted $75,000 for continued support for these projects in year 2 and $50,000 per year for years 3, 4, and 5. Surplus funds in any year are re-appropriated for other uses within the company and may not be carried over to future years.

$$
\begin{array}{ll}
\text{MAX:} & 141X_1 + 187X_2 + 121X_3 + 83X_4 + 265X_5 + 127X_6 \\
\text{Subject to:} & 75X_1 + 90X_2 + 60X_3 + 30X_4 + 100X_5 + 50X_6 \leq 250 \\
& 25X_1 + 35X_2 + 15X_3 + 20X_4 + 25X_5 + 20X_6 \leq 75 \\
& 20X_1 + 0X_2 + 15X_3 + 10X_4 + 20X_5 + 10X_6 \leq 50 \\
& 15X_1 + 0X_2 + 15X_3 + 5X_4 + 20X_5 + 30X_6 \leq 50 \\
& 10X_1 + 30X_2 + 15X_3 + 5X_4 + 20X_5 + 40X_6 \leq 50 \\
& \text{All } X_i \text{ must be binary}
\end{array}
$$

Assignment problem

- *Air-Express is an express shipping service that guarantees overnight delivery of packages anywhere in the continental United States. The company has various operations centers, called hubs, at airports in major cities across the country. Packages are received at hubs from other locations and then shipped to intermediate hubs or to their final destinations.*

CSE 7213c

- *The manager of the Air-Express hub in Baltimore, Maryland, is concerned about labor costs at the hub and is interested in determining the most effective way to schedule workers. The hub operates seven days a week, and the number of packages it handles each day varies from one day to the next. Using historical data on the average number of packages received each day, the manager estimates the number of workers needed to handle the packages as:*

Day of Week	Workers Required
Sunday	18
Monday	27
Tuesday	22
Wednesday	26
Thursday	25
Friday	21
Saturday	19

- The package handlers working for Air-Express are unionized and are guaranteed a five-day work week with two consecutive days off. The base wage for the handlers is $655 per week. Because most workers prefer to have Saturday or Sunday off, the union has negotiated bonuses of $25 per day for its members who work on these days. The possible shifts and salaries for package handlers are:

Shift	Days Off	Wage
1	Sunday and Monday	$680
2	Monday and Tuesday	$705
3	Tuesday and Wednesday	$705
4	Wednesday and Thursday	$705
5	Thursday and Friday	$705
6	Friday and Saturday	$680
7	Saturday and Sunday	$655

CSE 7213c

- The manager wants to keep the total wage expense for the hub as low as possible. With this in mind, how many package handlers should be assigned to each shift if the manager wants to have a sufficient number of workers available each day?

Decision variables

- X1 _ the number of workers assigned to shift 1
- X2 _ the number of workers assigned to shift 2
- X3 _ the number of workers assigned to shift 3
- X4 _ the number of workers assigned to shift 4
- X5 _ the number of workers assigned to shift 5
- X6 _ the number of workers assigned to shift 6
- X7 _ the number of workers assigned to shift 7

Objective and constraints

The LP model for the Air-Express scheduling problem is summarized as:

MIN: $680X_1 + 705X_2 + 705X_3 + 705X_4 + 705X_5 + 680X_6 + 655X_7$ } total wage expense

Subject to:

$0X_1 + 1X_2 + 1X_3 + 1X_4 + 1X_5 + 1X_6 + 0X_7 \geq 18$ } workers required on Sunday

$0X_1 + 0X_2 + 1X_3 + 1X_4 + 1X_5 + 1X_6 + 1X_7 \geq 27$ } workers required on Monday

$1X_1 + 0X_2 + 0X_3 + 1X_4 + 1X_5 + 1X_6 + 1X_7 \geq 22$ } workers required on Tuesday

$1X_1 + 1X_2 + 0X_3 + 0X_4 + 1X_5 + 1X_6 + 1X_7 \geq 26$ } workers required on Wednesday

$1X_1 + 1X_2 + 1X_3 + 0X_4 + 0X_5 + 1X_6 + 1X_7 \geq 25$ } workers required on Thursday

$1X_1 + 1X_2 + 1X_3 + 1X_4 + 0X_5 + 0X_6 + 1X_7 \geq 21$ } workers required on Friday

$1X_1 + 1X_2 + 1X_3 + 1X_4 + 1X_5 + 0X_6 + 0X_7 \geq 19$ } workers required on Saturday

$X_1, X_2, X_3, X_4, X_5, X_6, X_7 \geq 0$

All X_i must be integers

Stock selection problem

- Brian Givens is a financial analyst for Retirement Planning Services, Inc. who specializes in designing retirement income portfolios for retirees using corporate bonds.

- He has just completed a consultation with a client who expects to have $750,000 in liquid assets to invest when she retires next month. Brian and his client agreed to consider upcoming bond issues from the following six companies:

Company	Return	Years to Maturity	Rating
Acme Chemical	8.65%	11	1-Excellent
DynaStar	9.50%	10	3-Good
Eagle Vision	10.00%	6	4-Fair
MicroModeling	8.75%	10	1-Excellent
OptiPro	9.25%	7	3-Good
Sabre Systems	9.00%	13	2-Very Good

- The column labeled “Return” in this table represents the expected annual yield on each bond, the column labeled “Years to Maturity” indicates the length of time over which the bonds will be payable, and the column labeled “Rating” indicates an independent underwriter’s assessment of the quality or risk associated with each issue.

- Brian believes that all of the companies are relatively safe investments. However, to protect his client’s income, Brian and his client agreed that no more than 25% of her money should be invested in any one investment and at least half of her money should be invested in long-term bonds that mature in ten or more years.

CSE 7213c

- Also, even though DynaStar, Eagle Vision, and OptiPro offer the highest returns, it was agreed that no more than 35% of the money should be invested in these bonds because they also represent the highest risks (i.e., they were rated lower than “very good”).

CSE 7213c

Decision variables

- X1 = amount of money to invest in Acme Chemical
- X2 = amount of money to invest in DynaStar
- X3 = amount of money to invest in Eagle Vision
- X4 = amount of money to invest in MicroModeling
- X5 = amount of money to invest in OptiPro
- X6 = amount of money to invest in Sabre Systems

Objective function

- Maximize the investment income. Because each dollar invested in Acme Chemical (X1) earns 8.65% annually, and so on, the objective function for the problem is expressed as
  - MAX: .0865X1 + .095X2 + .10X3 + .0875X4 + .0925X5 + .09X6

Subject to:

$X_1 \le 187{,}500$	} 25% restriction per investment
$X_2 \le 187{,}500$	} 25% restriction per investment
$X_3 \le 187{,}500$	} 25% restriction per investment
$X_4 \le 187{,}500$	} 25% restriction per investment
$X_5 \le 187{,}500$	} 25% restriction per investment
$X_6 \le 187{,}500$	} 25% restriction per investment
$X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 750{,}000$	} total amount invested
$X_1 + X_2 + X_4 + X_6 \ge 375{,}000$	} long-term investment
$X_2 + X_3 + X_5 \le 262{,}500$	} higher-risk investment
$X_1, X_2, X_3, X_4, X_5, X_6 \ge 0$	} nonnegativity conditions

Product mix problems

- National Petroleum produces two types of unleaded gasoline: regular and premium. It sells these at Rs. 600 and 800 per barrel. These are blended from their internal domestic oil and foreign oil and must meet the following constraints

	Maximum vapor pressure	Minimum octane rating	Maximum demand (barrels/wk)	Minimum deliverables (barrels/wk)
Regular	23	88	100,000	50,000
Premium	23	93	20,000	5000

	vapor pressure	Octane rating	Inventory(barrels)	Cost(barrels)
Domestic	27	87	40,000	400
Foreign	15	98	60,000	500

Decision variables

- How do I blend regular and premium such that
  - Profit is maximized
  - Constraints are met
  - Let us say we use d1 barrels of domestic for regular oil and d2 barrels for premium oil. Similarly, we use f1 barrels of foreign for regular and f2 barrels for premium.

Objective function

- Domestic oil consumed = d1+d2
- Foreign oil consumed: f1+f2
- Regular blended: d1+f1
- Premium blended: d2+f2

- Costs
  - The cost of domestic oil: (d1+d2)400;
  - The cost of foreign oil: 500(f1+f2)
- Price
  - The price of regular oil: 600(d1+f1);
  - The price of premium oil: 800(d2+f2)

- The profit: total price - total cost = 600(d1+f1) + 800(d2+f2) – [400(d1+d2)+500(f1+f2)]

200d1+100f1+400d2+300f2

- Amount of domestic oil consumed: d1+d2

- Inventory available: 40,000 barrels
  - $d1+d2 \leq 40{,}000$
  - $f1+f2 \leq 60{,}000$

CSE 7213c

- Amount of regular produced: d1+f1
- The maximum demand: 100,000 and minimum deliverables = 50000 barrels

$$50,000 \leq d1+f1 \leq 100,000$$

$$5000 \leq d2+f2 \leq 20,000$$

- Vapor pressure is based on the weight fractions

- Vapor pressure of d1+f1 of regular = (weight fraction of domestic)* vapor pressure of domestic + (weight fraction of foreign)*vapor pressure of foreign

- Vapor pressure of regular =

$$\frac{d1}{d1+f1}(Vapor\ pressure\ of\ domestic) + \frac{f1}{d1+f1}(vapor\ pressure\ of\ foreign)$$

$$\frac{\boldsymbol{d1}}{\boldsymbol{d1+f1}}(\mathbf{27}) + \frac{\boldsymbol{f1}}{\boldsymbol{d1+f1}}(\mathbf{15}) \leq \mathbf{23},$$

27d1+15f1 = 23d1+23f1 -> 4d1-8f1 $\leq$0

- $\frac{d2}{d2+f2}(27) + \frac{f2}{d2+f2}(15) = 23$

$27d2+15f2 \leq 23d2+23f2 \rightarrow 4d2-8f2 \leq 0$

- Extending the same logic to octane rating

$$\frac{d1}{d1+f1}(87)+\frac{f1}{d1+f1}(98)\geq 88$$

$$-d1+10f1\geq 0$$

- $\frac{d2}{d2+f2}(87)+\frac{f2}{d2+f2}(98)=93$

$$-6d2+5f2\geq 0$$

- Hidden constraints: d1, d2, f1, f2 ≥ 0

Assignment

- A 400-meter medley relay involves four different swimmers, who successively swim 100 meters of the backstroke, breaststroke, butterfly and freestyle. A coach has six very fast swimmers whose expected times (in seconds) in the individual events are given in following

Assignment

	Event 1 (backstroke)	Event 2 (breaststroke)	Event 3 (butterfly)	Event 4 (freestyle)
Swimmer 1	65	73	63	57
Swimmer 2	67	70	65	58
Swimmer 3	68	72	69	55
Swimmer 4	67	75	70	59
Swimmer 5	71	69	75	57
Swimmer 6	69	71	66	59

Transportation Problem

- Tropicsun currently has 275,000 bags of citrus at Mt. Dora, 400,000 bags at Eustis, and 300,000 bags at Clermont. Tropicsun has citrus processing plants in Ocala, Orlando, and Leesburg with processing capacities to handle 200,000, 600,000, and 225,000 bags, respectively.

- Tropicsun contracts with a local trucking company to transport its fruit from the groves to the processing plants. The trucking company charges a flat rate for every mile that each bushel of fruit must be transported. Each mile a bushel of fruit travels is known as a bushel-mile. The following table summarizes the distances (in miles) between the groves and processing plants:

CSE 7213c

	Distances (in miles) Between Groves and Plants		
Grove	**Ocala**	**Orlando**	**Leesburg**
Mt. Dora	21	50	40
Eustis	35	30	22
Clermont	55	20	25

MIN: $21X_{14} + 50X_{15} + 40X_{16} + 35X_{24} + 30X_{25} + 22X_{26} + 55X_{34} + 20X_{35} + 25X_{36}$ } total distance fruit is shipped (in bushel-miles)

Subject to:

$X_{14} + X_{24} + X_{34} \leq 200{,}000$ } capacity restriction for Ocala

$X_{15} + X_{25} + X_{35} \leq 600{,}000$ } capacity restriction for Orlando

$X_{16} + X_{26} + X_{36} \leq 225{,}000$ } capacity restriction for Leesburg

$X_{14} + X_{15} + X_{16} = 275{,}000$ } supply available at Mt. Dora

$X_{24} + X_{25} + X_{26} = 400{,}000$ } supply available at Eustis

$X_{34} + X_{35} + X_{36} = 300{,}000$ } supply available at Clermont

$X_{ij} \geq 0$, for all i and j } nonnegativity conditions

Method 2

You start with the cost matrix as above but add dummy source or receiver to ensure that demand = supply

	Ocala	Orlando	Leesburg	Supply available
Mt. Dora	21	50	40	275000
Eustis	35	30	22	400000
Clermont	55	20	25	300000
Capacities	200000	600000	225000	

Production scheduling

- An industrial firm must plan for each of the four seasons over the next year. The company’s production capacities and the expected demands (all in units) are as follows:

	Spring	Summer	Fall	Winter
Demand	250	100	400	500
Regular Capacity	200	300	350	---
Overtime Capacity	100	50	100	150

- Regular production costs for the firm are $7.00 per unit. The unit cost of overtime varies seasonally being $8.00 in spring and fall, $9.00 in summer and $10.00 in winter.

- The company has 200 units of inventory on January 1, but as it plans to discontinue the product at the end of the year, it wants no inventory after the winter season. Units produced on regular shifts are not available for shipment during the season of production; generally, they are sold during the following season.

- Those that are not are added to inventory and carried forward at a cost of $0.70 per unit per season. In contrast, units produced on overtime shifts must be shipped in the same season as produced. Determine a production schedule that meets all demands at minimum total cost.

Costs						Supply
From/To	Spring	Summer	Fall	Winter	Dummy	
RegSpr	10000	7	7.7	8.4	0	200
RegSum	10000	10000	7	7.7	0	300
RegFall	10000	10000	10000	7	0	350
Initial	0	0.7	1.4	2.1	10000	200
OTSpr	8	10000	10000	10000	0	100
OTSum	10000	9	10000	10000	0	50
OTFall	10000	10000	8	10000	0	100
OTWinter	10000	10000	10000	10	0	150
Demand	250	100	400	500	200	