



SPORTS SCHEDULING

An Introduction to Integer Optimization

15.071x – The Analytics Edge

The Impact of Sports Schedules



- Sports is a \$300 billion dollar industry
 - Twice as big as the automobile industry
 - Seven times as big as the movie industry
- TV networks are key to revenue for sports teams
 - \$513 million per year for English Premier League soccer
 - \$766 million per year for NBA
 - \$3 billion per year for NFL
- They pay to have a good schedule of sports games

Sports Schedules

- Good schedules are important for other reasons too
 - Extensive traveling causes player fatigue
 - Ticket sales are better on the weekends
 - Better to play division teams near the end of season
- All competitive sports require schedules
 - **Which pairs** of teams play each other and **when?**

The Traditional Way



- Until recently, schedules mostly constructed by hand
 - Time consuming: with 10 teams, there are over 1 trillion possible schedules (every team plays every other team)
 - Many constraints: television networks, teams, cities, . . .
- For Major League Baseball, a husband and wife team constructed the schedules for 24 years (1981-2005)
 - Used a giant wall of magnets to schedule 2430 games
- Very difficult to add new constraints

Some Interesting Constraints

- In 2008, the owners and TV networks were not the only ones who cared about the schedule
- President Barack Obama and Senator John McCain complained about the schedule
 - National conventions conflicted with game scheduling
- Then, the Pope complained about the schedule!
 - The Pope visited New York on April 20, 2008
 - Mass in Yankee stadium (the traditional location)
- Each of these constraints required a new schedule

An Analytics Approach



- In 1996, “The Sports Scheduling Group” was started
 - Doug Bureman, George Nemhauser, Michael Trick, and Kelly Easton
- They generate schedules using a computer
 - Have been scheduling college sports since 1999
 - Major League Baseball since 2005
- They use optimization
 - Can easily adapt when new constraints are added

Scheduling a Tournament

- Four teams
 - Atlanta (A) , Boston (B) , Chicago (C) , and Detroit (D)
- Two divisions
 - Atlanta and Boston
 - Chicago and Detroit
- During four weeks
 - Each team plays the other team in its division twice
 - Each team plays teams in other divisions once
- The team with the most wins from each division will play in the championship
- Teams prefer to play divisional games later

An Optimization Approach



- Objective
 - Maximize team preferences (divisional games later)
- Decisions
 - Which teams should play each other each week
- Constraints
 - Play other team in division twice
 - Play teams in other divisions once
 - Play exactly one team each week

Decision Variables

- We need to decide which teams will play each other each week
 - Define variables x_{ijk}
 - If team i plays team j in week k , $x_{ijk} = 1$
 - Otherwise, $x_{ijk} = 0$
- This is called a *binary decision variable*
 - Only takes values 0 or 1

$$x_{AC2} = 1$$

$$x_{AC1} = 0$$

$$x_{AC3} = 0$$

$$x_{AC4} = 0$$

Integer Optimization

- Decision variables can only take integer values
- Binary variables can be either 0 or 1
 - Where to build a new warehouse
 - Whether or not to invest in a stock
 - Assigning nurses to shifts
- Integer variables can be 0, 1, 2, 3, 4, 5, ...
 - The number of new machines to purchase
 - The number of workers to assign for a shift
 - The number of items to stock

The Formulation



- Objective
 - Maximize team preferences (divisional games later)
- Decisions
 - Which teams should play each other each week
- Constraints
 - Play other team in division twice
 - Play teams in other divisions once
 - Play exactly one team each week

The Formulation



- Objective
 - Maximize team preferences (divisional games later)
- Decisions
 - Binary variables x_{ijk}
- Constraints
 - Play other team in division twice
 - Play teams in other divisions once
 - Play exactly one team each week

The Formulation

- Objective
 - Maximize team preferences (divisional games later)
- Decisions
 - Binary variables x_{ijk}
- Constraints
 - $x_{AB1} + x_{AB2} + x_{AB3} + x_{AB4} = 2$
 - Play teams in other divisions once
 - Play exactly one team each week

Similar constraint for teams C and D

The Formulation

- Objective
 - Maximize team preferences (divisional games later)
- Decisions
 - Binary variables x_{ijk}
- Constraints
 - $x_{AB1} + x_{AB2} + x_{AB3} + x_{AB4} = 2$
 - $x_{AC1} + x_{AC2} + x_{AC3} + x_{AC4} = 1$
 - Play exactly one team each week

Similar constraint for teams C and D

Similar constraints for teams A and D, B and C, and B and D

The Formulation

- Objective
 - Maximize team preferences (divisional games later)
- Decisions
 - Binary variables x_{ijk}
- Constraints
 - $x_{AB1} + x_{AB2} + x_{AB3} + x_{AB4} = 2$
 - $x_{AC1} + x_{AC2} + x_{AC3} + x_{AC4} = 1$
 - $x_{AB1} + x_{AC1} + x_{AD1} = 1$

Similar constraint for teams C and D

Similar constraints for teams A and D, B and C, and B and D

Similar constraints for every team and week

The Formulation

- Objective
 - Maximize $x_{AB1} + 2x_{AB2} + 4x_{AB3} + 8x_{AB4}$
 $+x_{CD1} + 2x_{CD2} + 4x_{CD3} + 8x_{CD4}$
- Decisions
 - Binary variables x_{ijk}
- Constraints
 - $x_{AB1} + x_{AB2} + x_{AB3} + x_{AB4} = 2$
 - $x_{AC1} + x_{AC2} + x_{AC3} + x_{AC4} = 1$
 - $x_{AB1} + x_{AC1} + x_{AD1} = 1$

Similar constraint for
teams C and D

Similar constraints for
teams A and D, B and
C, and B and D

Similar constraints for
every team and week

Adding Logical Constraints

- Binary variables allow us to model logical constraints
- A and B can't play in weeks 3 and 4

$$x_{AB3} + x_{AB4} \leq 1$$

- If A and B play in week 4, they must also play in week 2

$$x_{AB2} \geq x_{AB4}$$

- C and D must play in week 1 or week 2 (or both)

$$x_{CD1} + x_{CD2} \geq 1$$

Solving Integer Optimization Problems

- We were able to solve our sports scheduling problem with 4 teams (24 variables, 22 basic constraints)
- The problem size increases rapidly
 - With 10 teams, 585 variables and 175 basic constraints
- For Major League Baseball
 - 100,000 variables
 - 200,000 constraints
 - This would be impossible in LibreOffice
- So how are integer models solved in practice?

Solving Integer Optimization Problems



- Reformulate the problem
 - The sports scheduling problem is solved by changing the formulation
 - Variables are sequences of games
 - Split into three problems that can each be solved separately
- Heuristics
 - Find good, but not necessarily optimal, decisions

Solving Integer Optimization Problems



- General purpose solvers
 - CPLEX, Gurobi, GLPK, COIN-OR
- In the past 20 years, the speed of integer optimization solvers has increased by a factor of 250,000
 - Doesn't include increasing speed of computers
- **A problem that can be solved in 1 second today took 7 years to solve 20 years ago!**

Solving the Sports Scheduling Problem



- When the Sports Scheduling Group started, integer optimization software was not useful
- Now, they can use powerful solvers to generate schedules
- Takes months to make the MLB schedule
 - Enormous list of constraints
 - Need to define priorities on constraints
 - Takes several iterations to get a good schedule

The Analytics Edge



- Optimization allows for the addition of new constraints or structure changes
 - Can easily generate a new schedule based on an updated requirement or request
- Now, all professional sports and most college sports schedules are constructed using optimization