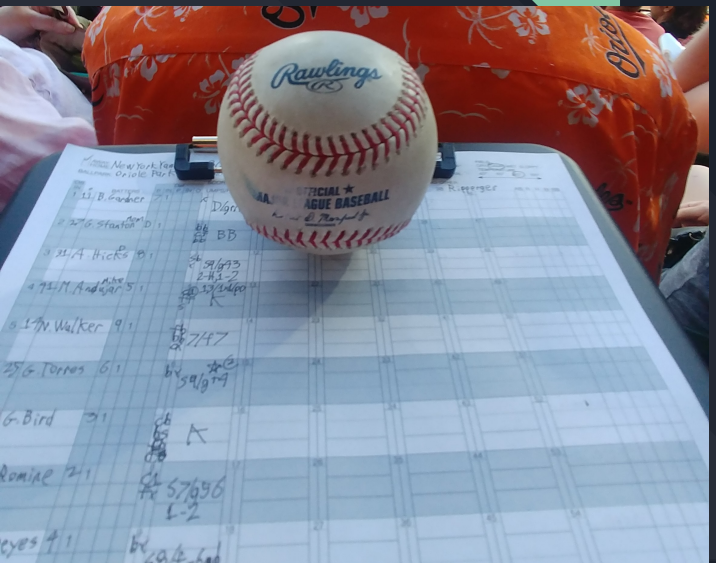
The Ultimate Baseball Road Trip and Algorithm



Michael Mountain
CLSAC 10/30/18


CitiFIELD
Nikon

Combinatorial Optimization

- Optimize a combination from a finite set of objects
- Define a set of constraints that a solution must satisfy
- Define criteria to compare one solution to another

Major League Baseball (MLB) schedule

- 30 teams each play 81 home games in a season, between April and September
- 30 x 81 = 2430 total games
- Find the most efficient driving route to attend one game at each stadium
- Primary optimization: Minimize total number of days (trip duration)
- Secondary: Minimize driving time (*travel* duration)

Constraints

Our problem space has 2,494 decision variables and approximately 120,000 constraints

https://harvardsportsanalysis.files.wordpress.com/2011/06/shortest_possible_baseball_road_trip1.pdf

$$
\begin{aligned}
&(1)\quad B \le S_{n,b} + (1 - X_{n,b}) * M \quad \forall b \in B,\ n \in N \\
&(2)\quad E \ge S_{n,b} + (1 - X_{n,b}) * M \quad \forall b \in B,\ n \in N \\
&(3)\quad \textstyle\sum_{n \in N} X_{n,b} = 1 \quad \forall b \in B \\
&(4)\quad X_{n,b} + X_{n',b'} \le 1 \quad \forall b, b' \in B,\ n, n' \in N \ \ such\,that\ |S_{n,b} - S_{n',b'}| \le T_{b,b'} + D \\
&(5)\quad \textstyle\sum_{n \in N} X_{n,b} * S_{n,b} \le (1 - F_b) * M + \sum_{n' \in N} X_{n',b'} * S_{n',b'} \quad \forall b, b' \in B \\
&(6)\quad (1 - L_b) * M + \textstyle\sum_{n \in N} X_{n,b} * S_{n,b} \le \sum_{n' \in N} X_{n',b'} * S_{n',b'} \quad \forall b, b' \in B \\
&(7)\quad \textstyle\sum_{b \in B} F_b = 1 \\
&(8)\quad \textstyle\sum_{b \in B} L_b = 1 \\
&(9)\quad P \ge T_{b,b'} - (2 - F_b - L_{b'}) * M \quad \forall b, b' \in B
\end{aligned}
$$

Domain-Specific Heuristics

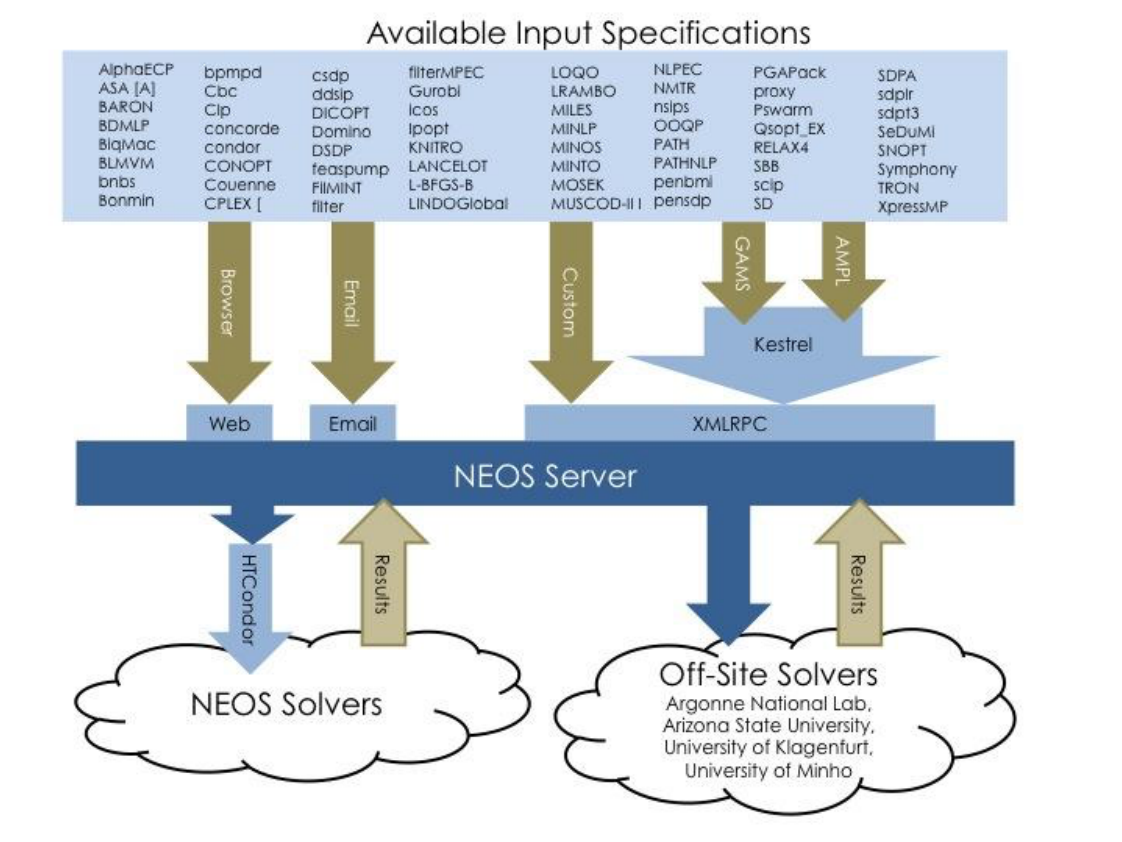
- Isolated West Coast stadiums
- Reasonable upper bound
- All-Star Break (early July)

Creating Subsets of the Problem Space

1. Identify a date range where the seven West Coast stadiums can be visited in no more than nine days
2. Select all games from other stadiums that are scheduled within 26 days of either end of the West Coast date range
3. Remove any games which are on a different side of the All-Star Break from the West Coast date range
4. Repeat steps 1-3 with a different West Coast date range

The larger subsets will have about 700 decision variables and 15,000 constraints (~20x reduction in computational complexity)

NEOS (Network-Enabled Optimization System)



GUB cover cuts applied: 8
Clique cuts applied: 73
Cover cuts applied: 28
Implied bound cuts applied: 39
Flow cuts applied: 27
Mixed integer rounding cuts applied: 151
Zero-half cuts applied: 40
Lift and project cuts applied: 61
Gomory fractional cuts applied: 5

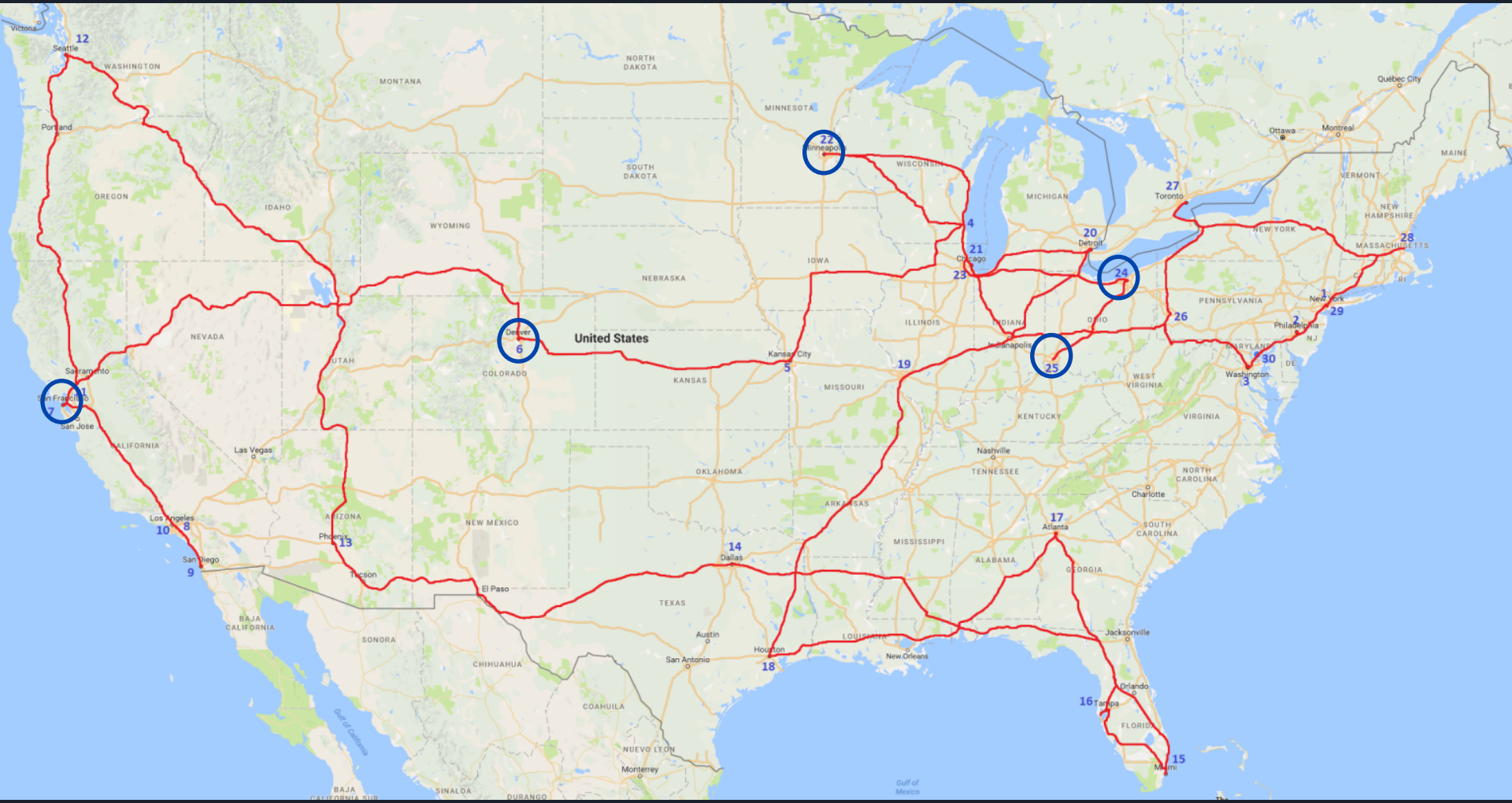
Root node processing (before b&c):
  Real time = 12.75 sec. (2154.94 ticks)
Parallel b&c, 4 threads:
  Real time = 178.28 sec. (36540.97 ticks)
  Sync time (average) = 27.16 sec.
  Wait time (average) = 0.01 sec.

Total (root+branch&cut) = 191.03 sec. (38695.90 ticks)

Solution pool: 16 solutions saved.

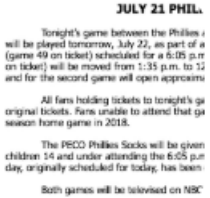
MIP - Integer optimal, tolerance (0.0001/1e-06): Objective = 4.9269000000e+04
Current MIP best bound = 4.9266243190e+04 (gap = 2.75681, 0.01%)
Solution time = 191.03 sec. Iterations = 844453 Nodes = 11742 (1)
Deterministic time = 38695.91 ticks (202.56 ticks/sec)

```
CPLEX> Incumbent solution
Variable Name           Solution Value
endTime                  340715.000000
startTime                291635.000000
timeToStart                 189.000000
NYYJul21                      1.000000
PHIJul21                      1.000000
WASJul22                      1.000000
MILJul23                      1.000000
KCRJul24                      1.000000
COLJul25                      1.000000
SFGJul27                      1.000000
SDPJul28                      1.000000
LAAJul29                      1.000000
OAKJul30                      1.000000
SEAAug01                      1.000000
LADAug03                      1.000000
ARIAug04                      1.000000
TEXAug06                      1.000000
MIAAug08                      1.000000
TBRAug09                      1.000000
ATLAug10                      1.000000
HOUAug12                      1.000000
STLAug13                      1.000000
DETAug14                      1.000000
CHCAug15                      1.000000
MINAug16                      1.000000
CWSAug17                      1.000000
CLEAug18                      1.000000
CINAug19                      1.000000
PITAug20                      1.000000
TORAug21                      1.000000
NYMAug22                      1.000000
BOSAug23                      1.000000
BALAug24                      1.000000
isFirstNYY                    1.000000
isLastBAL                     1.000000
All other variables in the range 1-538 are 0.
CPLEX>
```


United States
Seattle
Portland
Sacramento
San Francisco
San Jose
Los Angeles
San Diego
Las Vegas
Phoenix
Tucson
El Paso
Denver
Kansas City
Minneapolis
Chicago
Detroit
Toronto
Indianapolis
Dallas
Houston
Austin
San Antonio
New Orleans
Atlanta
Nashville
Charlotte
Jacksonville
Orlando
Tampa
Miami
Washington
Philadelphia
New York
Montreal
Ottawa
Québec City
Monterrey
WASHINGTON
OREGON
CALIFORNIA
NEVADA
IDAHO
MONTANA
WYOMING
UTAH
COLORADO
ARIZONA
NEW MEXICO
NORTH DAKOTA
SOUTH DAKOTA
NEBRASKA
KANSAS
OKLAHOMA
TEXAS
MINNESOTA
IOWA
MISSOURI
ARKANSAS
LOUISIANA
WISCONSIN
ILLINOIS
MICHIGAN
INDIANA
OHIO
KENTUCKY
TENNESSEE
MISSISSIPPI
ALABAMA
GEORGIA
FLORIDA
SOUTH CAROLINA
NORTH CAROLINA
VIRGINIA
WEST VIRGINIA
PENNSYLVANIA
NEW YORK
MARYLAND
DE
NJ
VERMONT
NEW HAMPSHIRE
MASSACHUSETTS
RI
MAINE
Baja California
Sonora
Chihuahua
Coahuila
Nuevo León
Sinaloa
Durango
Gulf of California
Gulf of Mexico

Baseball Road Trip
@MLBRoadTrip

NOOOOOOOOOOOO


JULY 21 PHIL.

Tonight's game between the Phillies
will be played tomorrow, July 22, as part of
(game 49 on ticket) scheduled for a 6:05 p.
on ticket) will be moved from 1:35 p.m. to
and for the second game will open approxim

All fans holding tickets to tonight's
original tickets. Fans unable to attend that
season home game in 2018.

The PECO Phillies Socks will be given
children 14 and under attending the 6:05
day, originally scheduled for today, has been

Both games will be televised on NBC

Philadelphia Phillies @Phillies
Tonight's #Phillies game against the San Diego Padres has been rescheduled and will be played as part of a split-admission doubleheader tomorrow, July 22....


United States
Seattle
Olympia
Yakima
Kennewick
Spokane
Portland
Salem
Eugene
Bend
Medford
Redding
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NEVADA
CALIFORNIA
UTAH
COLORADO
ARIZONA
NEW MEXICO
TEXAS
OKLAHOMA
KANSAS
NEBRASKA
SOUTH DAKOTA
NORTH DAKOTA
MINNESOTA
IOWA
MISSOURI
ARKANSAS
LOUISIANA
MISSISSIPPI
ALABAMA
GEORGIA
FLORIDA
TENNESSEE
KENTUCKY
ILLINOIS
INDIANA
OHIO
MICHIGAN
WISCONSIN
PENNSYLVANIA
NEW YORK
NEW JERSEY
MARYLAND
DELAWARE
VIRGINIA
WEST VIRGINIA
NORTH CAROLINA
SOUTH CAROLINA
CONNECTICUT
MASSACHUSETTS
VERMONT
NEW HAMPSHIRE
MAINE
Boise
Twin Falls
Idaho Falls
Pocatello
Missoula
Helena
Butte
Bozeman
Great Falls
Billings
Miles City
Ogden
Salt Lake City
Sandy
Provo
Reno
Sacramento
San Francisco
San Jose
Fresno
Bakersfield
Los Angeles
Anaheim
San Diego
Tijuana
Mexicali
Las Vegas
St. George
Phoenix
Mesa
Tucson
Nogales
Fort Collins
Denver
Grand Junction
Colorado Springs
Pueblo
Santa Fe
Albuquerque
Las Cruces
El Paso
Amarillo
Lubbock
Midland
Odessa
Abilene
Fort Worth
Dallas
Austin
San Antonio
Houston
Galveston
Corpus Christi
Laredo
Norman
Tulsa
Wichita
Lawrence
Kansas City
Lincoln
Omaha
Sioux City
Sioux Falls
Rapid City
Bismarck
Dickinson
Fargo
Grand Forks
Minot
St Cloud
Minneapolis
Rochester
Des Moines
Cedar Rapids
Duluth
Bemidji
Madison
Oshkosh
Milwaukee
Rockford
Chicago
Naperville
Peoria
Springfield
Grand Rapids
Detroit
Ann Arbor
Toledo
Fort Wayne
Indianapolis
Cincinnati
Columbus
Cleveland
Pittsburgh
Philadelphia
New York
Toronto
Mississauga
Buffalo
London
Kingston
Ottawa
Montreal
Quebec City
Sherbrooke
Sudbury
Manitoulin Island
Thunder Bay
Winnipeg
Brandon
Weyburn
Estevan
Portland
St. Louis
Columbia
Springfield
Branson
Fayetteville
Evansville
Louisville
Lexington
Nashville
Knoxville
Asheville
Chattanooga
Huntsville
Memphis
Birmingham
Tuscaloosa
Montgomery
Columbus
Atlanta
Jackson
Shreveport
Beaumont
New Orleans
Mobile
Panama City Beach
Tallahassee
Jacksonville
Savannah
Charleston
Daytona Beach
Orlando
West Palm Beach
Miami
Nassau
Charlotte
Greensboro
Durham
Wilmington
Myrtle Beach
Richmond
Roanoke
Norfolk
Virginia Beach
Mark Twain National Forest
Gila National Forest
Death Valley National Park
Humboldt-Toiyabe National Forest
Yellowstone National Park
Shoshone National Forest
Salmon - Challis National Forest
Payette National Forest
Nez Perce – Clearwater National Forests
Flathead National Forest
Kootenai National Forest
COLVILLE RESERVATION
FORT PECK INDIAN RESERVATION
CROW RESERVATION
STANDING ROCK RESERVATION
CHEYENNE RIVER RESERVATION
PINE RIDGE RESERVATION
UINTAH AND OURAY RESERVATION
NAVAJO NATION RESERVATION
Superior National Forest
Lake Superior
Lake Michigan
Lake Huron
La Verendrye Wildlife Reserve
Algonquin Provincial Park
BAJA CALIFORNIA
SONORA
CHIHUAHUA
COAHUILA
Hermosillo
Ciudad Obregón
Navojoa
Chihuahua
Ciudad Cuauhtémoc
Delicias
Parral
Loreto
Reserva de la Biosfera El Vizcaíno
Gulf of California
Google

Trip Stats

- 30 MLB games in 36 days
- 16,935 miles driven
- Every major US interstate (multiples of 5)
- 34 states + DC and Ontario
- 49 gas stops
- 574 players (77% of major league rosters)
- Four MLB debuts
- Orioles record: 9-23

- 2 guest appearances on FanGraphs podcast Effectively Wild (episodes 1169 and 1263)
- @MLBRoadTrip on Twitter with ballpark photos, stadium food selections and road trip playlists
- Interviews with WCBS radio in New York, Shepherd Express in Milwaukee, Fast Company and Technical.ly Baltimore
- Meetups with family, friends, and fans all over the country

Source Code

https://github.com/mountm/mlbroadtrip

Special Thanks

- Ben Blatt & Harvard Sports Analysis Collective
- neos-server.org
- Catalyte
- Effectively Wild (@EWPod)
- Interstate Highway System
- Assistant driver Dave Mountain (7/22, 8/2-8/10)

Source code at github.com/mountm/mlbroadtrip

Optimizing for Travel Distance with the Backtracking Approach

Backtracking approach

- Incrementally build a candidate solution
- Abandon your candidate (backtrack) as soon as you can confirm that the candidate is invalid (unable to be extended)
- Not a formal algorithm, more of a metaheuristic

Two Questions

- What does a partial candidate solution looks like?
- How do we know if a candidate solution is valid?

Backtracking pseudocode, where the entry point is bt(root(P))

```
procedure bt(c)
    if reject(P, c) then return
    if accept(P, c) then output(P, c)
    s ← first(P, c)
    while s ≠ Λ do
        bt(s)
        s ← next(P, s)
```

Λ is the *null candidate*

Let’s define our procedural parameters!

P - an array of games in the schedule, starting from the beginning point of our time-optimized trip, ending at the final allowed day, sorted by start time

root(P) - the first game in the time-optimized solution

reject(P, c) - verify that each game in the candidate *c* can be reached from the previous one and that no stadium is visited twice

accept(P, c) - verify that the length of the candidate array is 30

first(P, c) - From P, select the next game after the final game in *c*, and add it to *c* (or return Λ)

next(P, s) - From P, select the next game after the final game in *c* and replace the final game in *c*

output(P, c) - Calculate the total driving distance for *c*. If no previous solution was found, accept this as the best solution. Otherwise, keep the solution with the shortest distance needed.

Improvements to Consider

- The reject function should err on the side of caution
- But it should be as aggressive as possible
- Effective equivalence of partial candidates for earlier pruning
- Trading memory usage for execution time (e.g. parallelization)