

Optimization revisited



Prof. Noah Snavely
CS1114

<http://www.cs.cornell.edu/courses/cs1114/>



Cornell University
Computer Science

Administrivia

- Assignments:
 - A6 due tomorrow (demo slots available)
- Prelim 3 next Thursday
 - Review in class on Tuesday



Administrivia

- Final projects
 - Demo session: Tuesday, May 15, 1:30-3:30
 - For the demo session, prepare a 7-minute demo for your project
 - Sign-up slots available soon
 - Best demo will receive the “Best Demo” prize

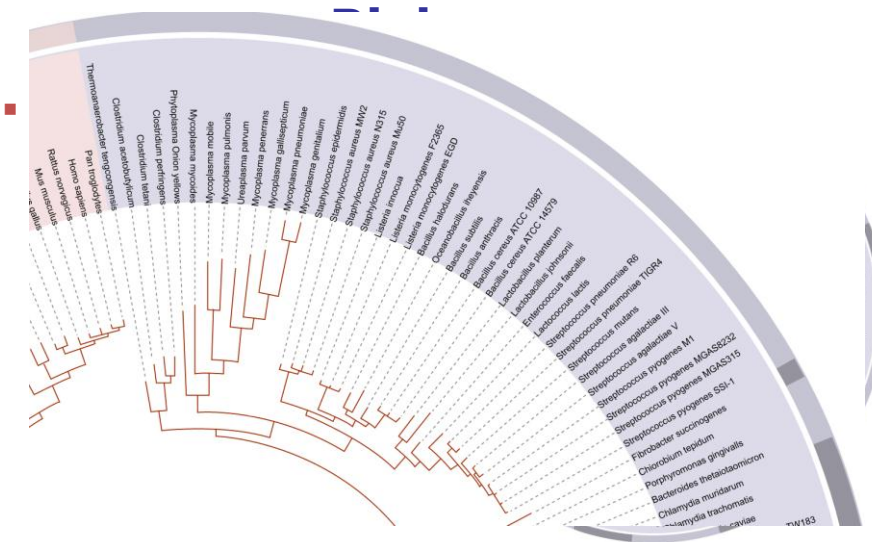
Optimization problems

- Very important concept for many different areas of science and engineering
- City planning
 - How do I figure out what bus routes to put in Ithaca?
 - How do I decide where to put the next hospital?

Biology

DNA sequence alignment

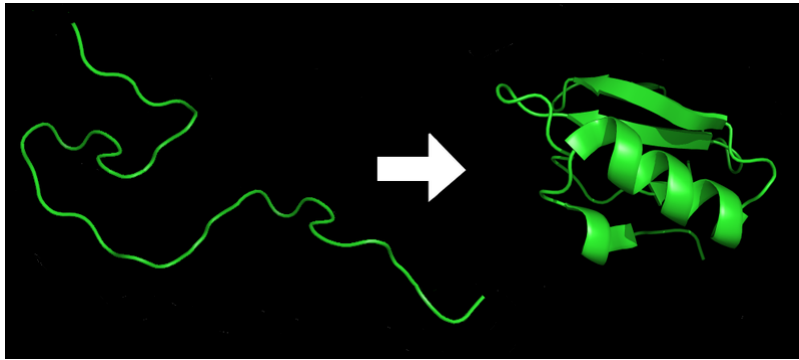
Q5E940_BOVIN	-----MPEEDRATWNSYELKELI	LLDDPKCFVYADVVGKQKQ	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	76
RLAO_HUMAN	-----MPEEDRATWNSYELKELI	LLDDPKCFVYADVVGKQKQ	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	76
RLAO_MOUSE	-----MPEEDRATWNSYELKELI	LLDDPKCFVYADVVGKQKQ	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	76
RLAO_RAT	-----MPEEDRATWNSYELKELI	LLDDPKCFVYADVVGKQKQ	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	76
RLAO_CHICK	-----MPEEDRATWNSYELKELI	LLDDPKCFVYADVVGKQKQ	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	76
RLAO_BANXY	-----MPEEDRATWNSYELKELI	LLDDPKCFVYADVVGKQKQ	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	76
Q7ZUG3_BRARE	-----MPEEDRATWNSYELKELI	LLDDPKCFVYADVVGKQKQ	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	76
RLAO_ICTPU	-----MPEEDRATWNSYELKELI	LLDDPKCFVYADVVGKQKQ	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	76
RLAO_DROME	-----MPEEDRATWNSYELKELI	LLDDPKCFVYADVVGKQKQ	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	76
RLAO_DICED	-----MSAG-SKREKLFSEKATLFTT	DDKMTVABADVVGKQKQ	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	75
Q54LPO_DICED	-----MSAG-SKREKLFSEKATLFTT	DDKMTVABADVVGKQKQ	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	75
RLAO_PLAFL	-----MACKLQKQKQMTYKESLZQCK	KLITVHFDV	QIDMSIRKQ	AVVLMGKTMHMKAIKIGL	ENH	---P	76
RLAO_SULAC	-----HTGLAVTTKZAKKVDVFAE	DTKLTITLIANTIGPADK	HEIKKIRKQ	ADIKVTEHLF	IALKNG	---D	79
RLAO_SULTO	-----HLMHVIYQKKEKWELEEFEL	ELAEHTLIIANTIGPADK	HEIKKIRKQ	ADIKVTEHLF	IALKNG	---D	80
RLAO_SUGO	-----MKRLALAKQKQKQKWELEEF	ELAEHTLIIANTIGPADK	HEIKKIRKQ	ADIKVTEHLF	IALKNG	---D	80
RLAO_AERPE	-----HVSIVYQMYKREKPEKNTLM	RELELFRKRVFLADITG	QVYVYVSKKQKQ	YHMKAKKELI	EMKNG	---L	86
RLAO_PYRAB	-----HMKSEKRYRYTQDANVYK	TEHLELQKRVYFLQIMSE	ELIIEVYKRYE	QVTEHTEHLF	QKNTTQK	---D	85
RLAO_MEFAC	-----MSEEMHTEHQQKDEENKEL	LQSKVYFVYVTEGLATK	NKQIRKQ	ADIKVTEHLF	IALKNG	---D	78
RLAO_METHA	-----MSEEMHTEHQQKDEENKEL	LQSKVYFVYVTEGLATK	NKQIRKQ	ADIKVTEHLF	IALKNG	---D	78
RLAO_MCFU	-----MSEVRES--SSEVAVRARE	EKHEIKKIRKQ	ADIKVTEHLF	IALKNG	---D	75	
RLAO_METKA	-----HAYKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---P	88
RLAO_METHI	-----MAHVAWKKKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---P	74
RLAO_METHJ	-----MTDAKSEKZAPWKEE	VWALKELKSNVIAL	EDHBEVAV	QVTEHTEHLF	IALKNG	---D	82
RLAO_PYRJA	-----METKVAHVAWKEE	VEKTEKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---P	81
RLAO_PYRAB	-----MAHVAWKKKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---P	77
RLAO_PYRHO	-----MAHVAWKKKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---P	77
RLAO_PYRPU	-----MAHVAWKKKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---P	77
RLAO_PYRKO	-----MAHVAWKKKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---P	76
RLAO_HALMA	-----MSESEKRTETEFKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---D	79
RLAO_HALYO	-----MSESEKRTETEFKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---D	79
RLAO_HALSA	-----MSESEKRTETEFKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---D	79
RLAO_THAC	-----MKEVSOQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---E	72
RLAO_THRYO	-----MKEVSOQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---E	72
RLAO_PICTO	-----MTEKQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	KQKQKQKQKQKQKQKQ	---N	72



Optimal Phylogenetic tree

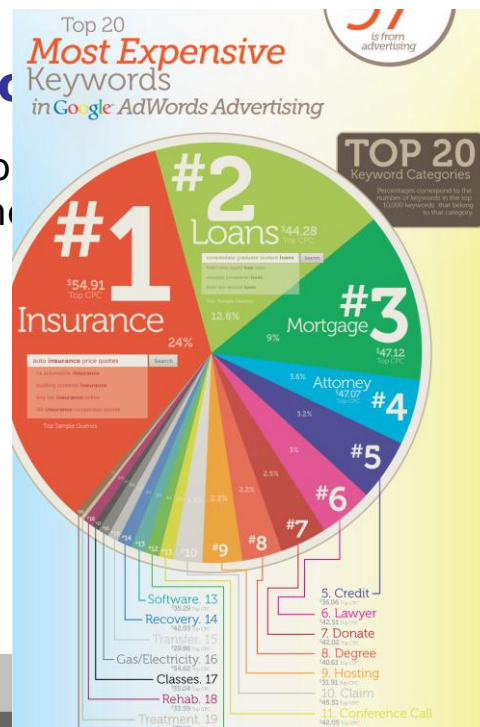
Biology

- Protein folding



Economic

- How much does Google keyword advertisement



Games

- What is the optimal strategy for a given game?
 - Rock / paper / scissors?
 - Chess?
 - Is there a winning strategy for white?

Computer Vision

Robotics



Optimization problems

- We've seen several optimization-type problems in this class:
 - k -means
 - k -centers
 - Fitting a line to a set of points
 - Coloring a graph with the minimum number of colors / fitting animals into the minimum set of cages
 - Finding the minimum spanning tree of a graph
 - Others?



Optimization problems

- Key elements:
 1. A set of valid solutions
 - Line fitting: all possible lines
 - Minimum spanning tree: all possible spanning trees of a graph
 - k -means: all possible sets of k means



Optimization problems

- Key elements:
 2. A way to measure how good/bad a solution is (an *objective function*):
 - Line fitting: sum of squared residuals
 - Minimum spanning tree: sum of edge weights
 - k -means: sum of squared distances from each input point to its assigned mean
 - Sometimes coming up with a good objective function is very difficult in itself (e.g., protein folding)



k-means objective function

- Find the centers that minimize the sum of squared distances to the points
- Objective function:

Given input points $x_1, x_2, x_3, \dots, x_n$, find the clusters C_1, C_2, \dots, C_k and the cluster centers $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_k$ that minimize

$$\sum_{j=1}^k \sum_{x_i \in C_j} |x_i - \bar{x}_j|^2$$



Optimization algorithms

- For any given optimization *problem*, we would like to come up with a (hopefully efficient) algorithm
 - That ideally finds the global minimum of the objective function



Optimization algorithms

- Example: minimum spanning tree (MST)
 - First algorithm: Borůvka's algorithm (1926)
 - For constructing an efficient electrical network for Moravia
 - Prim's algorithm
 - Kruskal's algorithm



Optimization algorithms

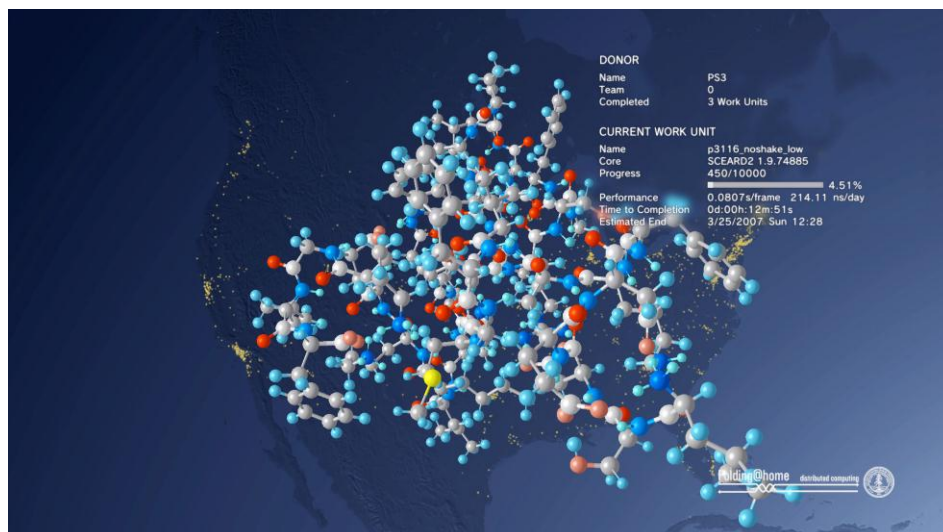
- Possible algorithm types:
 - Greedy algorithms (e.g. Prim and Kruskal)
 - Iterative algorithms (e.g. bubble sort, gradient descent)
 - Guess-and-check (e.g. RANSAC)
- Some algorithms return the global optimum, others just return a "good" answer



Another optimization problem

- Dense box
 - One statement: Given a set of 2D points, compute a good bounding box, that is not too big and contains most of the points
 - What objective function should we use?

Parallel optimization algorithms



Human optimization algorithms

The image shows a screenshot of a protein structure prediction game interface. On the left, a 3D protein structure is displayed in green and blue. A control panel at the bottom left includes buttons for 'Shake Sidechains', 'Wiggle Backbone', 'Clear Locks and Bands', 'Reset Puzzle', and 'Mouse Help'. A tooltip above the buttons reads 'Shake sidechains to improve the protein. Hotkey: S'. On the right, a leaderboard shows the following data:

Group Competition	
Group Name	Score
1 The Lone Folder	9308
2 Street Smarts	9367
3 Binon	9303
4 Berkeley	9255

Player Competition	
Player Name	Score
16 pson	9098
17 kathleen	9092
18 versat92	9091
19 daboares	9081
20 ccarico	9032
21 majoriegeen	9048
22 srickarson	9038

Below the game interface, a snippet from the journal 'nature' is shown. The article title is 'Predicting protein structures with a multiplayer online game' and the sub-headline is 'You did it!'. The authors listed are Seth Cooper, Firas Khatib, Adrien Treuille, James Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popovic, and Foldit players. A red circle highlights the authors' names. The journal logo and navigation links are also visible.

Questions?