CMPSCI 791BB: Advanced ML

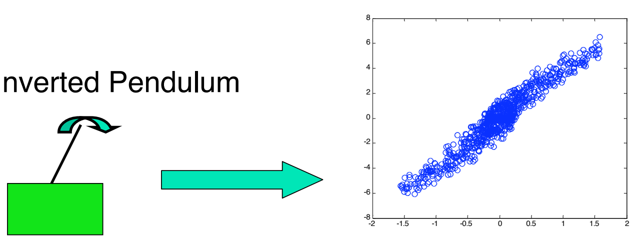
Sridhar Mahadevan

Overview of lecture

- *Laplacian* framework for machine learning
  - What is it?
  - Why is it interesting and exciting?
  - Some ongoing research
- Course structure and reading lists
- Background preparation
- Discussion

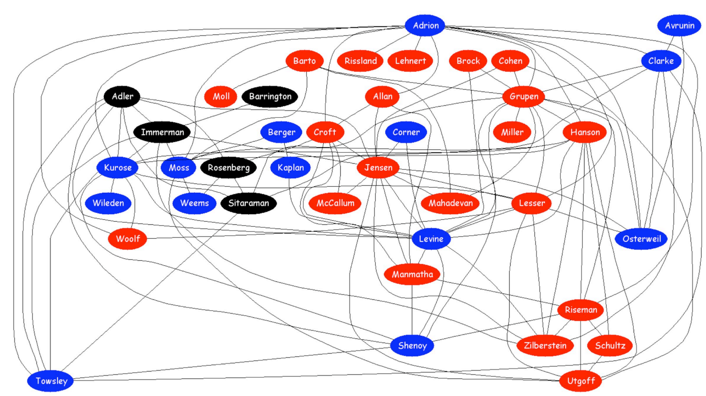
Control Learning

Inverted Pendulum



Points on state space manifold

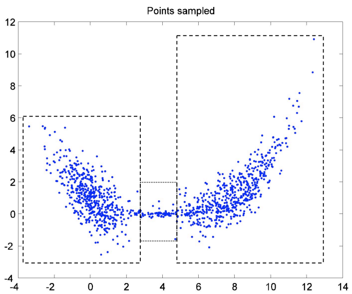
Graph Clustering



"Social Networks"

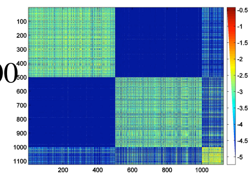
Random Walk Analysis

(Maggioni and Mahadevan, 2006)



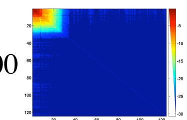
Random walk in a continuous two-room environment

1000x1000



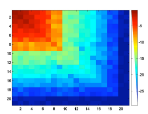
T

100 x 100



T^4

20 x 20



T^6

Many Applications of Laplacian Learning

- Fast inference in graphical models
- Stability of complex molecules (e.g, Bucky Ball)
- Developmental learning in humans and other animals
- Information retrieval and text analysis
- Robot motion planning and workspace modeling
- Factory optimization
- Game playing (checkers, chess, Go)
- Routing in overlay networks
- Sensor networks

Laplace Operator

- The Laplace operator has been termed *the most beautiful object* in all of math and physics
  - It is a motif in every major differential equation that describes the physics of our universe (heat equation, Maxwell's equations, Schrodinger's equation)
  - It also is a central conceptual abstraction that ties together disparate fields of mathematics: analysis, geometry, topology, number theory
- In this course, we will explore the properties of the Laplace operator in depth
  - *Spectral clustering* using the graph Laplacian
  - *Representation learning* using the eigenspace of the graph Laplacian

The Laplace Operator in Physics

- Suppose we measure the temperature each point (x,y,z) in this room. This associates a scalar value f(x,y,z) with every point. We can map this scalar into a vector by measuring its gradient.

$$\nabla f = \frac{\partial f}{\partial x} i + \frac{\partial f}{\partial y} j + \frac{\partial f}{\partial z} k$$

- Over time, we might want to measure the process of *heat diffusion* (maybe its correlated with how hard you're thinking!)

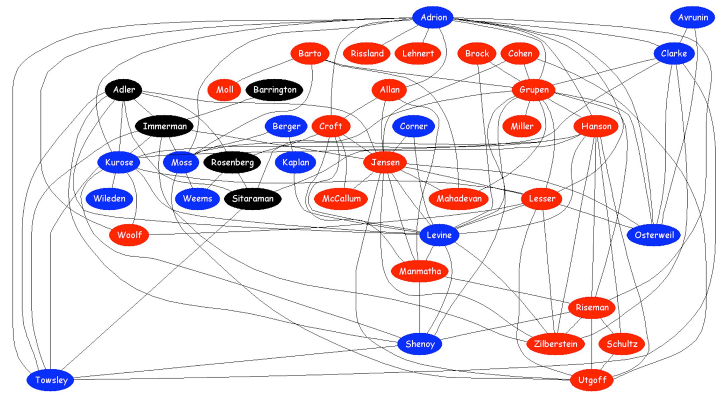
The Laplace Operator in Physics

- To measure diffusion, we need to map the vector back into a scalar value using the divergence

$$\langle \nabla, f \rangle = \frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} + \frac{\partial f_z}{\partial z}$$

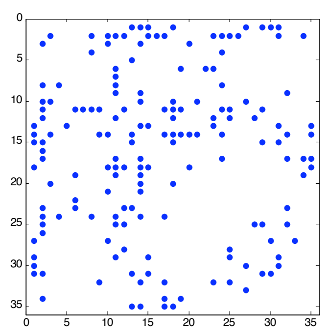
- The Laplacian is the divergence of the gradient

Faculty Collaboration Graph: Spectral Analysis via "Heat" Diffusion!



Adjacency Matrix of Faculty Graph

- The adjacency matrix A of a graph G = (V, E) is one of several operators that we will study in depth
- Other operators:
  - Combinatorial Graph Laplacian = L = D - A
  - Normalized Laplacian = $D^{-1/2}$ (D-A) $D^{-1/2}$
  - Random Walk = $D^{-1/2}$ A
  - Diffusion Process
- For each operator, we will study its spectrum and eigenspace



Combining Graph Theory and Linear Algebra

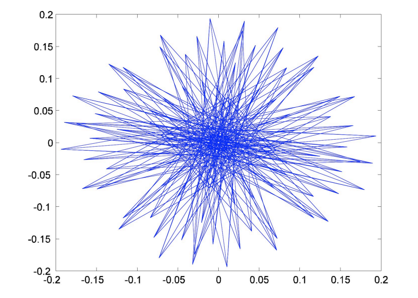
- Given a graph G = (V, E), we can study the space of functions on the graph: V ! R
- From basic linear algebra, we know that since the adjacency matrix A is symmetric, we can use the spectral theorem

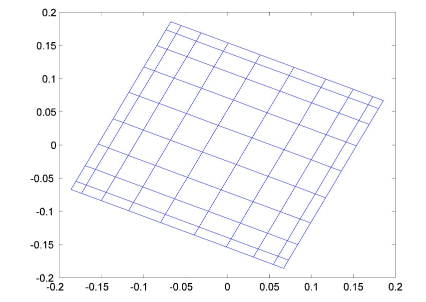
$$A = V \, \Lambda \, V^T$$

- V is a matrix of eigenvectors, Λ is a diagonal matrix of eigenvalues
- Eigenvectors satisfy the following property:

$$A \, x = \lambda \, x$$

Adjacency vs. Laplacian Embedding

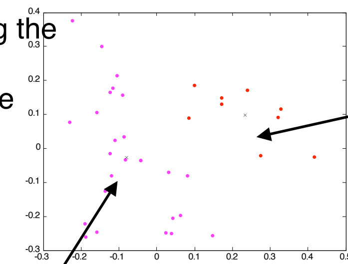




Embedding of a 10x10 grid using the 2nd and 3rd eigenvectors

Spectral Clustering using Graph Laplacian

Embedding using the 2nd and 3rd eigenvector of the graph Laplacian



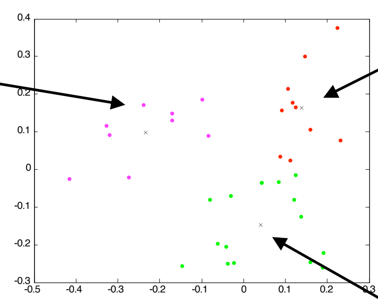
Cluster: 1
Adler
Barrington
Immerman
Kurose
Rosenberg
Shenoy
Sitaraman
Towsley
Weems

cluster: 2
Adrion Allan Avrunin Barto Brock Clarke Cohen
Croft Grupen Hanson Jensen Lehnert Lesser Levine Mahadevan
Manmatha McCallum Moll Moss Osterweil Riseman
Rissland Schultz Utgoff Woolf Zilberstein

UMASS UNIVERSITY OF MASSACHUSETTS, AMHERST • DEPARTMENT OF COMPUTER SCIENCE

Spectral Clustering using the Graph Laplacian

Cluster: 2
Adler
Barrington
Immerman
Kurose
Rosenberg
Shenoy
Sitaraman
Towsley
Weems



Cluster: 1
Barto
Brock
Grupen
Hanson
Mahadevan
Moll
Moss
Riseman
Schultz
Utgoff

Cluster: 3
Adrion
Allan
Avrunin
Clarke
Cohen
Croft
Jensen
Lehnert
Lesser
Levine
Manmatha
McCallum
Osterweil
Rissland
Woolf
Zilberstein

Embedding using the 2nd and 3rd eigenvector of the graph Laplacian

UMASS UNIVERSITY OF MASSACHUSETTS, AMHERST • DEPARTMENT OF COMPUTER SCIENCE

Spectral Clustering using Graph Laplacian

Cluster: 1
Adrion
Avrunin
Brock
Clarke
Cohen
Lehnert
Lesser
Osterweil
Rissland
Utgoff
Woolf

Cluster: 2
Adler
Barrington
Immerman
Kurose
Rosenberg
Shenoy
Sitaraman
Towsley
Weems

Cluster: 3
Allan
Croft
Jensen
Levine
Manmatha
McCallum
Zilberstein

Cluster: 4
Barto
Grupen
Hanson
Mahadevan
Moll
Moss
Riseman
Schultz

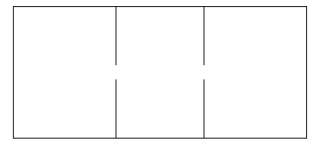
UMASS UNIVERSITY OF MASSACHUSETTS, AMHERST • DEPARTMENT OF COMPUTER SCIENCE

Spectral Clustering using Graph Laplacian

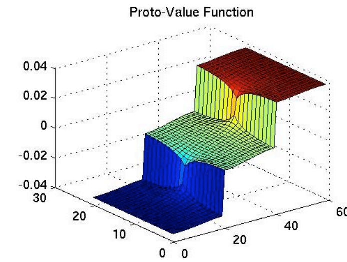
Cluster: 1	Cluster: 2	Cluster: 3	Cluster: 4	Cluster: 5	Cluster: 6
Adrion	Adler	Barto	Avrunin	Allan	Barrington
Brock	Rosenberg	Grupen	Clarke	Croft	Immerman
Cohen	Sitaraman	Hanson	Lesser	Jensen	Kurose
Lehnert	Weems	Mahadevan	Osterweil	Levine	Shenoy
Rissland		Moll		Manmatha	Towsley
Utgoff		Moss		McCallum	
Woolf		Riseman			
Zilberstein		Schultz			

Proto-Value Functions: Laplacian Approach to Agent Learning

(Mahadevan, AAAI, ICML, NIPS, UAI 2005)



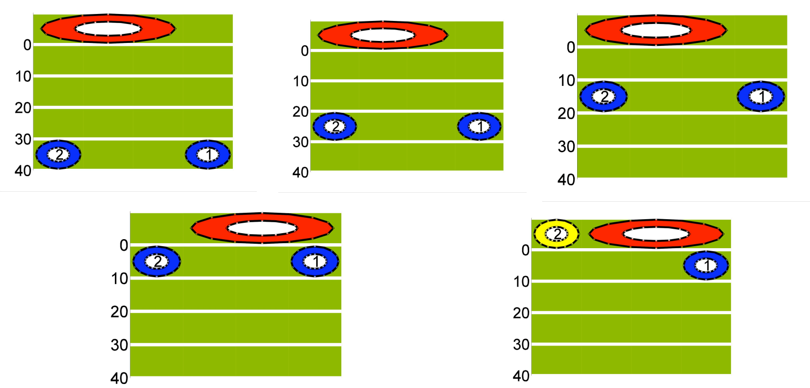
Spatial Environment



2nd Eigenvector of Graph Laplacian

Proto-Value Functions for Factored Environments

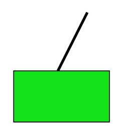
(Mahadevan, 2006)



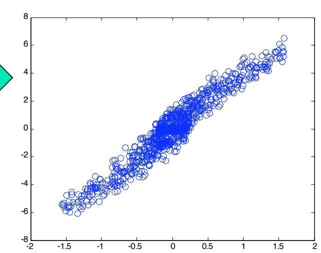
"Blocker" Domain

Harmonic Analysis: Fourier vs. Wavelets

(Coifman and Maggioni, ACHA 2006)



Inverted pendulum



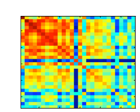
Samples from random walk on manifold in state space

Fourier

$$\Delta\phi = \lambda\phi$$

Eigenfunctions of Laplacian on state space: global basis functions

Wavelets



Dilations of diffusion operator: local multiscale basis functions

Challenges

- How do you analyze very large graphs?
  - By sampling techniques and exploiting graph regularities
- How can you exploit the regularity of a graph?
  - Representation theory of groups
- Can Laplacian learning be done in an incremental fashion?
  - Yes, there are distributed algorithms that scale nicely
- How well does Laplacian learning work?
  - Take the class and we'll find out!

Background for this Course

- Linear algebra: basic knowledge of matrices, eigenvalues, eigenvectors.
  - I recommend Strang's book, but any book will do.
- Basic knowledge of graph theory: undirected graphs, cycles, spanning trees, adjacency matrix representation
- Access to MATLAB (Windows, Linux, Mac)
- What I will cover:
  - Introduction to spectral graph theory
  - Representation theory of finite groups
  - Harmonic analysis: fourier and wavelets on groups

Course Workload

- MATLAB programming assignments (~ 2-3)
  - Example assignment: implement spectral clustering on faculty collaboration graph (or analyze grad student social behavior!)
- Weekly Reading:
  - For next class: Read Chapters 1 and 2 of Spielman's lecture notes on spectral graph theory (use Google to find these).
- Final Project:
  - Implement a learning system that uses Laplacian learning in a domain of your choice (maybe, we could do a "Blocker" competition!)
- Above all, I advocate self-learning! Try implementing stuff!