The Many Applications of Eigenvalues

David Bindel
7 Feb 2019
My Goals for Today

bindel@cornell.edu
Jones 317 (mostly until mid-May)

• Show how applied math happens (to me at Cornell).
• Convince you that eigenvalue problems are fun!
• Get you to talk to me, read slides, read papers, etc. (And maybe apply to Cornell for grad school!)
The Computational Science & Engineering Picture

- MEMS
- Fusion
- Networks
- Systems

- Linear algebra
- Approximation theory
- Symmetry + structure
- Optimization

- HPC / cloud
- Simulators
- Solvers
- Frameworks
My super power is turning everything you show me into an eigenvalue problem.

— Me (at every new grad student lunch)
Why Eigenvalue Problems?

Dynamics

Optimization

Data approximation

Densities and invariants
Why Eigenvalue Problems?

Dynamics: \( \frac{du}{dt} = Au \) or \( u(k + 1) = Au(k) \)

Optimization: minimize \( x^T Ax \) s.t. \( x^T x = 1 \)

Data approximation: minimize \( \|A - XY^T\|_F^2 \)

Invariants: \( \forall \) analytic \( f : \mathbb{C} \to \mathbb{C} \), compute \( \text{tr}(f(A)) \)

All these perspectives are connected!
Why Eigenvalue Problems?

Dynamics

Optimization

Data approximation

Densities and invariants
“On the beats in the vibrations of a revolving cylinder or bell”
by G. H. Bryan, 1890
Free vibrations in a rotating frame (simplified):

\[ \ddot{q} + 2\beta \Omega J \dot{q} + \omega_0^2 q = 0, \quad J \equiv \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \]

Eigenvalue problem: \((-\omega^2 I + 2i\omega\beta \Omega J + \omega_0^2) \ q = 0.\)

Solutions: \(\omega \approx \Omega_0 \pm \beta \Omega. \quad \Rightarrow \quad \text{beating } \propto \Omega!\)
This is a common picture:

- Symmetry leads to degenerate modes
- Perturbations split (some) degeneracies
A General Picture

\[ q_1(t) \approx \cos(-\beta \Omega t) q_0^0(t) - \sin(-\beta \Omega t) q_0^1(t) \]

\[ q_2(t) \approx \sin(-\beta \Omega t) q_0^0(t) + \cos(-\beta \Omega t) q_0^1(t) \]
Foucault in Solid State
A Small Application

Northrup-Grummond HRG
(developed c. 1965–early 1990s)
A Smaller Application (Cornell)
Perturbations split degenerate modes:

- **Coriolis forces** (good)
- **Imperfect fab** (bad, but physical)
- **Discretization error** (non-physical)
Basic framework:

- Represent geometry and imperfections in Fourier series
- Treat imperfections as perturbations
Analyzing Imperfections

Payoff:

- Quantitative: Fast and accurate "2.5D" simulations
- Qualitative: Selection rules identify "dangerous" imperfections
Yilmaz and Bindel
“Effects of Imperfections on Solid-Wave Gyroscope Dynamics”

Or talk to me about:

• Damping, radiation, and nonlinear eigenproblems in MEMS
• Nonlinear dynamics in MEMS (ongoing!)
Why Eigenvalue Problems?

Dynamics

Optimization

Data approximation

Densities and invariants
A basic model:

- A fixed *intrinsic* opinion $s_i$
- A variable *expressed* opinion $x_i$
- Equilibrium $x_i = \text{argmin}_{z_i} c_i(z_i)$, where

$$c_i(z_i) \equiv (s_i - z_i)^2 + \sum_{j \in N(i)} w_{ij} (z_i - x_j)^2$$

- Define a *social cost* $c(z) = \sum_i c_i(z_i)$
Methodology: Graph problem $\rightarrow$ linear algebra problem.

Nash equilibrium: $\quad (L + I)x = s$
Social optimum: $\quad (A + I)y = s$
Cost at equilibrium: $\quad c(x) = s^T Cs$
Optimal social cost: $\quad c(y) = s^T Bs$

Price of anarchy is a ratio of quadratics:

$$\text{PoA}(s) = \frac{c(x)}{c(y)} = \frac{s^T Cs}{s^T Bs}$$
Given

$$\text{PoA}(s) = \frac{s^T Cs}{s^T Bs}$$

Maximize by setting gradient to zero:

$$\nabla_s \text{PoA}(s) = \frac{2}{s^T Bs} [Cs - \text{PoA}(s) Bs] = 0$$

Find worst case through a generalized eigenvalue problem:

$$Cs_\star = \lambda Bs_\star$$
Sigal Oren: Jon Kleinberg and I are working on this problem, he suggested you might have some insight [explains]. So why is PoA always bounded by 9/8 for symmetric networks?

DB: OK

• PoA is a generalized eigenvalue.
• Matrices are $B = p(L)$ and $C = q(L)$
• Eigs are $p(\mu)/q(\mu)$ for $\mu$ an eig of $L$
• $p(\mu)/q(\mu)$ has a max of 9/8 for $\mu \geq 0$.

SO: Great, thanks! [Exit office]

— Ten minutes pass —

SO (knocks): So what about nonsymmetric networks?
Bindel, Kleinberg, Oren
“How Bad is Forming Your Own Opinion?”

Or talk to me about:

- Similar bounds for 3D image reconstruction!
- Spectral methods for community detection
- Fast parameterized PageRank computations
Why Eigenvalue Problems?

Dynamics

Optimization

Data approximation

Densities and invariants

\[ \begin{align*}
    & a \\
    & Z_1 \rightarrow X_1 \\
    & Z_2 \rightarrow X_2 \( n_m(n_m - 1) \) \\
    & 1 \leq m \leq M
\end{align*} \]
Old idea: Latent Semantic Indexing

- Documents as a word count vectors ("bag of words")
- Reweight to account for frequency (tf-idf)
- Compute *singular value decomposition* and truncate
  - Gives best rank $k$ approximation to $T$
- Cluster words/docs via $U_k$ and $V_k$
  - Rows for similar documents are similar
  - "Blurs out" related terms (car/automobile)
- But hard to interpret rows of $U_k$ / cols of $V_k$
  - May have negative entries, not normalized
A generative model for documents:

- Topics are distributions over words
- Documents involve distribution over topics
- Generate document by picking topic, then word from topic

Goal: Jointly determine topic and document distributions.
Beyond LDA


- Work with word co-occurrence statistics (topics only)
- Assume anchor words for each topic
- Much faster than MCMC-based LDA training (NLA-based)
- Provable guarantees with enough data from model
But — this is not how we write documents!

- Co-occurrence may not behave as model predicts
- Result: sometimes funky topics for real data
Idea: Enforce co-occurrence structure under model

- Should represent probability (non-negative, sums to 1)
- Should be low rank and positive semi-definite

Algorithm: Alternating projections
Alternating projections

- Alternate PSD-rank-$k$ and normalized matrix projections
- PSD-rank-$k$ projection by partial eigendecomposition
- Can compute fast using only matrix-vector products
- Run inference on the resulting matrix
Lee, Bindel, and Mimno,
“Robust Spectral Inference for Joint Stochastic Matrix Factorization,” NIPS 2015

- Still some ongoing work in this direction!
- Moontae Lee is now faculty at the UIC business school
Why Eigenvalue Problems?

Dynamics

Optimization

Data approximation

Densities and invariants
“You mean, if you had perfect pitch could you find the shape of a drum.” — Mark Kac (quoting Lipmann Bers) 
American Math Monthly, 1966
Spectra define a *generalized function* (a *density*):

\[
\text{tr}(f(H)) = \int f(\lambda) \mu(\lambda) \, dx = \sum_{j=k}^{N} f(\lambda_k)
\]

where \( f \) is an analytic test function. Smooth to get a picture: a *spectral histogram* or *kernel density estimate*.
A Bestiary of Matrices

- Adjacency matrix: $A$
- Laplacian matrix: $L = D - A$
- Unsigned Laplacian: $L = D + A$
- Random walk matrix: $P = AD^{-1}$ (or $D^{-1}A$)
- Normalized adjacency: $\bar{A} = D^{-1/2}AD^{-1/2}$
- Normalized Laplacian: $\bar{L} = I - \bar{A} = D^{-1/2}LD^{-1/2}$
- Modularity matrix: $B = A - \frac{dd^T}{2n}$
- Motif adjacency: $W = A^2 \odot A$

All have examples of co-spectral graphs

... through spectrum uniquely identifies quantum graphs
Consider

\[ \text{tr}(\exp(\alpha A)) = \sum_{k=1}^{\infty} \frac{\alpha^k}{k!} \cdot (\# \text{ closed random walks of length } k). \]

- Global measure of connectivity in a graph.
- Can clearly be computed via DoS.
- Generalizes to other weights.
DoS information equivalent to looking at the *heat kernel trace*:

\[ h(s) = \text{tr}(\exp(-sH)) = \mathcal{L}[\mu](s) \]

Use \( H = LD^{-1} \) (continuous time random walk generator) \( \implies \)

\[ h(s)/N = P(\text{self-return after time } s \text{ from uniform start}). \]
DoS information equivalent to looking at the power moments:

$$\text{tr}(H^j).$$

Natural interpretation for matrices associated with graphs:

- $A$: number of length $k$ cycles.
- $\bar{A}$ or $P$: return probability for $k$-step random walk (times $N$).
- $L$: ??
Local DoS \( \nu_k(x) \): symmetric case with \( H = Q \Lambda Q^T \),

\[
\int f(x) \nu_k(x) \, dx = f(H)_{kk} = e_k^T Q f(\Lambda) Q^T e_k
\]

\[
\nu_k(x) = \sum_{j=1}^{n} q_{kj}^2 \delta(x - \lambda_j)
\]

DoS is sum of local densities of states:

\[
\mu(x) = \sum_{k=1}^{n} \nu_k(x)
\]
Can compute common *centrality measures* with LDoS

- Estrada centrality: $\exp(\gamma A)_{kk}$
- Resolvent centrality: $[(I - \gamma \bar{A})^{-1}]_{kk}$

Some motifs associated with localized eigenvectors:

- Chief example: Null vectors of $\bar{A}$ supported on leaves.
- Use LDoS + topology to find motifs?

What else?
Reconstruct graph from *fully resolved* LDoS at all nodes?

- Assume $H = QQ^T$
- No multiple eigenvalues $\implies$ know $|Q|$ and $\Lambda$
- Can we recover signs in $Q$?

Feels a little like phase retrieval...
Computing the (L)DoS?

- **Kernel Polynomial Method (KPM) from physics**
  - Expand density of $H$ in a (dual) Chebyshev series
  - Coefficients look like $d_j = \text{tr}(T_j(H))$
  - Use stochastic trace estimation for fast traces
  - Filtering to kill Gibbs oscillations

- Other related methods (e.g. Golub-Meurant GQL)
- Got into this by knowing KPM and a chat with David Gleich!
- Some additional tricks for graph case
- Not enough time for details – let’s look at pictures!
Internet topology
Internet topology (local)
Marvel characters (local)
Marvel comics
Yeast
Yeast (local)
Enron emails (SNAP)
US power grid (Pajek)
$N = 326186, \ nnz = 1615400, \ 80 \ s \ (1000 \ moments, \ 10 \ probes)$
$N = 1139905$, $nnz = 113891327$, 2093 s (1000 moments, 10 probes)
What Do You Hear?
For more...

http://www.cs.cornell.edu/bindel
bindel@cornell.edu
Jones 317 (mostly until mid-May)