From Bells of Frost to Opinion’s Cost: The Many Applications of Eigenvalues

David Bindel
25 Oct 2019
Stitch Fix is using something called eigenvector decomposition, a concept from quantum mechanics, to tease apart the overlapping “notes” in an individual’s style. Using physics, the team can better understand the complexities of the clients’ style minds.
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
- Optimization
- Data approximation
- Densities and invariants
The fact of harmony between Heaven and Earth and Man does not come from physical union, from a direct action, it comes from a tuning on the same note producing vibrations in unison.

— Tong Tshung-chu

A thousand valleys’ rustling pines resound.
My heart was cleansed, as if in flowing water.
In bells of frost I heard the resonance die.

— Li Bai (translated by Vikram Seth)
Eigenvalues and Dynamics
A Case Study: Musical Microspheres

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

\[
\ddot{q} + 2\beta\Omega J\dot{q} + \omega_0^2 q = 0,
\]

\[
J \equiv \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}
\]

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

Solutions: \(\omega \approx \omega_0 \pm \beta\Omega \). \(\Rightarrow\) beating \(\propto \Omega\)!
This is a common picture:

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

\[
\begin{bmatrix}
q_1(t) \\
q_2(t)
\end{bmatrix} \approx \begin{bmatrix}
\cos(-\beta\Omega t) & -\sin(-\beta\Omega t) \\
\sin(-\beta\Omega t) & \cos(-\beta\Omega t)
\end{bmatrix} \begin{bmatrix}
q_1^0(t) \\
q_2^0(t)
\end{bmatrix}.
\]
Foucault in Solid State
A Small Application

Northrup-Grummond HRG
(developed c. 1965–early 1990s)
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
Payoff:

- Quantitative: Fast and accurate “2.5D” simulations
- Qualitative: Selection rules for “dangerous” imperfections
Why Eigenvalue Problems?

- Dynamics
- Optimization
- Data approximation
- Densities and invariants
The DeGroot Model (Dynamics)

- *Opinions* are numbers between $-1$ and $1$
- People like to agree with others
- Update opinions by averaging over neighbors
The DeGroot Model (Dynamics)
The DeGroot Model (Dynamics)
The DeGroot Model (Dynamics)
The DeGroot Model (Dynamics)
The DeGroot Model (Dynamics)
The DeGroot Model (Dynamics)
The DeGroot Model (Dynamics)
• *Opinions* are numbers between $-1$ and $1$
• People like to agree with others *and their “core beliefs”*
• Update opinions by averaging over neighbors
The Modified DeGroot Model (Dynamics)
The Modified DeGroot Model (Dynamics)
The Modified DeGroot Model (Dynamics)
The Modified DeGroot Model (Dynamics)
The Modified DeGroot Model (Dynamics)
The Modified DeGroot Model (Dynamics)
Eigenvalues and Optimization

\[ ax^2 + 2bxy + cy^2 = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1 \]
Define an “unhappiness” (cost), e.g.

\[ c_c(z) = (z - b_C)^2 + 2(z - x_A)^2 + (z - x_P)^2 \]

where \( x_i \) is an “expressed opinion” and \( s_i \) is a “core belief.”

To minimize cost, take a (weighted) average of opinions

\[ x_c^{\text{new}} = (s_C + 2x_A + x_P)/5 \]
Nash equilibrium: All players try to minimize their own cost.
The Price of Anarchy

Social cost

\[ c(x) = c_A(x) + c_B(x) + c_C(x) + c_D(x) + c_P(x) \]

Questions:

- What is the price of anarchy?
  
  \[ \text{PoA}(s) = \frac{c(x^{Nash})}{c(x^{optimal})} \]

- What is the worst case price of anarchy?
Methodology: Graph problem $\mapsto$ linear algebra problem.

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

Price of anarchy is a ratio of quadratics:

$$\frac{c(x)}{c(y)} = \frac{b^T C b}{b^T B b}$$

Worst case solves a generalized eigenvalue problem

$$Cs = \lambda Bs, \quad \text{PoA}(s) = \lambda$$
• 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?
Why Eigenvalue Problems?

- Dynamics
- Optimization
- Data approximation
- Densities and invariants
Eigenvalues and Data Approximation

Distance

\[
\begin{bmatrix}
0 & 0.895 & 0.001 & 0.073 & 0.592 \\
0.895 & 0 & 0.913 & 0.457 & 0.034 \\
0.001 & 0.913 & 0 & 0.080 & 0.600 \\
0.073 & 0.457 & 0.080 & 0 & 0.249 \\
0.592 & 0.034 & 0.609 & 0.249 & 0.0
\end{bmatrix}
\]
Eigenvalues and Data Approximation

\[ T \approx U_k \Sigma_k V_k \]
Case Study: Spectral Text Analysis and Topic Models
“Bag of Words” and the Vector Space Model

Trump
Apprentice
Fired

Paris Accord
Climate
Green Deal

Trump
Apprentice
Fired
Paris
Climate
Green Deal
Congress

<table>
<thead>
<tr>
<th></th>
<th>Trump</th>
<th>Apprentice</th>
<th>Fired</th>
<th>Paris</th>
<th>Climate</th>
<th>Green Deal</th>
<th>Congress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trump</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Apprentice</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fired</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Paris</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Climate</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Green Deal</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Congress</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Old idea: Latent Semantic Indexing

- Documents as a word count vectors (“bag of words”)
- Reweight to account for frequency (tf-idf)
- Decompose to determine term/document similarity
Decomposing Data in Different Domains

<table>
<thead>
<tr>
<th>Domain</th>
<th>Object ($a_{i,j}$)</th>
<th>Cluster ($w_{i,k}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document</td>
<td>Word</td>
<td>Topic</td>
</tr>
<tr>
<td>Image</td>
<td>Pixel</td>
<td>Segment</td>
</tr>
<tr>
<td>Network</td>
<td>User</td>
<td>Community</td>
</tr>
<tr>
<td>Legislature</td>
<td>Member</td>
<td>Party/Group</td>
</tr>
<tr>
<td>Playlist</td>
<td>Song</td>
<td>Genre</td>
</tr>
<tr>
<td>Chemical spectra</td>
<td>Mixture</td>
<td>Molecules</td>
</tr>
</tbody>
</table>

$A \approx WH$
Latent Dirichlet Allocation (LDA) is a generative model:

- For each topic, choose word distribution \( \phi_k \sim \text{Dir}(\beta) \)
- For each doc, choose topic distribution \( \theta_m \sim \text{Dir}(\alpha) \)
- For word \( n \) in document \( m \)
  - Choose topic \( z_{m,n} \sim \text{Cat}(\theta_m) \)
  - Choose word \( w_{m,n} \sim \text{Cat}(\phi_{z_{m,n}}) \)

How does LDA compare to spectral inference methods?
<table>
<thead>
<tr>
<th>Arora et al. 2013 (Baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td>neuron layer hidden recognition signal cell noise</td>
</tr>
<tr>
<td>neuron layer hidden cell signal representation noise</td>
</tr>
<tr>
<td>neuron layer cell hidden signal noise dynamic</td>
</tr>
<tr>
<td>neuron layer cell hidden control signal noise</td>
</tr>
<tr>
<td>neuron layer hidden cell signal recognition noise</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probabilistic LDA (Gibbs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>neuron cell visual signal response field activity</td>
</tr>
<tr>
<td>control action policy optimal reinforcement dynamic robot</td>
</tr>
<tr>
<td>recognition image object feature word speech features</td>
</tr>
<tr>
<td>hidden net layer dynamic neuron recurrent noise</td>
</tr>
<tr>
<td>gaussian approximation matrix bound component variables</td>
</tr>
</tbody>
</table>
Not a conventional NeurIPS author.
Rectification by Alternating Projections
### How Well Does It Work?

<table>
<thead>
<tr>
<th>Lee et al. 2015 (AP)</th>
<th>Probabilistic LDA (Gibbs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>neuron circuit cell synaptic signal layer activity</td>
<td>neuron cell visual signal response field activity</td>
</tr>
<tr>
<td>control action dynamic optimal policy controller reinforcement</td>
<td>control action policy optimal reinforcement dynamic robot</td>
</tr>
<tr>
<td>recognition layer hidden word speech image net</td>
<td>recognition image object feature word speech features</td>
</tr>
<tr>
<td>cell field visual direction image motion object orientation</td>
<td>hidden net layer dynamic neuron recurrent noise</td>
</tr>
<tr>
<td>gaussian noise hidden approximation matrix bound examples</td>
<td>gaussianian approximation matrix bound component variables</td>
</tr>
</tbody>
</table>
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
Kernel polynomial method (see Weisse, Rev. Modern Phys.):

What’s different in the graph case?
What Do You Hear?
Why Eigenvalue Problems?

Dynamics

Optimization

Data approximation

Densities and invariants
Why Eigenvalue Problems?

Our journey this hour:

- Clothing styles
- Tuning forks
- Foucault’s pendulum and gyroscopes
- Game theory and opinion models
- Modeling topics in document collections
- Hearing the shape of a drum or a graph

... with only a brief mention of quantum mechanics.
Stitch Fix is using something called the eigenvector decomposition, a fundamental concept from linear algebra used in physics, data analysis, and many other settings. With this technique, they tease apart the overlapping “factors” in an individual’s style. Using linear algebra, the team can better understand the complexities of the clients’ styles.
Why not?
— David Bindel