Linear Algebra Review
Part 3: Spectral Properties

Edwin Olson
University of Michigan
The Three-Day Plan

- Geometry of Linear Algebra
  - Vectors, matrices, basic operations, lines, planes, homogeneous coordinates, transformations

- Solving Linear Systems
  - Gaussian Elimination, LU and Cholesky decomposition, over-determined systems, calculus and linear algebra, non-linear least squares, regression

- The Spectral Story
  - Eigensystems, singular value decomposition, principle component analysis, spectral clustering
Minimize distance?

Equation for hyperplane: $p \cdot \hat{n} = b$

- This makes sense too. Which one should we minimize?
  - Depends on the nature of the error.

$n$ is the unit normal to the line
$p$ is any point on the line
Regression

• Suppose points to-be-fit, pi, are zero-mean
  ▷ (Without loss of generality...)

• Can formulate fitting problem as:

\[
\begin{bmatrix}
p_1^T \\
p_2^T \\
p_3^T \\
p_4^T \\
\vdots \\
\end{bmatrix}
\begin{bmatrix}
n_x \\
n_y \\
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0 \\
\end{bmatrix}
\]

\[Ax = b = 0\]
Over-determined system!

- Use our trick from before to reduce the over-determined system to a rectangular system.

\[ A^T A x = A^T 0 = 0 \]

- What value of \( x \) minimizes the error?

\[ x = 0 \]

- But that’s dumb and not helpful. What if we constrain \( ||x||^2 = 1 \) and minimize \( ||A^T A x||^2 \)
A wacky matrix

• Consider a matrix composed of two orthonormal matrices U & V and a diagonal matrix S:

  U          S          V^T

• For what unit vector \( x \) would \( USV^T x \) be the longest? The shortest?
Checkpoint

• How are the singular values of $A$ and $\text{inv}(A)$ related?

•
Singular Value Decomposition

- For any matrix $A$, an SVD factorization exists:

$$A = USV^T$$

- $U$ is an $M \times M$ orthonormal basis
- $V$ is an $N \times N$ orthonormal basis
- $S$ is diagonal ($M \times N$)

- If $A$ is symmetric, $U=V$
Singular Value Decomposition

- Rotate, scale, rotate

\[ A = USV^T \]

- Compute (by inspection) SVDs for matrices below:
Back to regression

• Now we can solve our regression problem

\[
\begin{bmatrix}
p_1^T \\
p_2^T \\
p_3^T \\
p_4^T \\
\vdots \\
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0 \\
\vdots \\
\end{bmatrix}
\]

Where is our solution?

\[
\begin{align*}
\text{minimize} & \quad \| A^T A x \|^2 \\
\text{such that} & \quad \| x \|^2 = 1
\end{align*}
\]
Regression: Results
Some definitions

• Semi-Positive Definite
  ▶ All singular values $\geq 0$

• Semi-Negative Definite
  ▶ All singular values $\leq 0$

• Condition number. (Bigger or smaller is better?)
  ▶ $\lambda_{max}/\lambda_{min}$
Checkpoint

• Show that the product of two SPD matrices is SPD.
• Show that the product of two SND matrices is SPD.
• Suppose you’ve factored A into USV’. What is inv(A)?
• What’s the condition number of a singular matrix?
• Show that the condition number of A’A is worse than A. (How much worse?)
Orthonormal Bases

• Any vector can be written as a linear combination of an orthonormal basis.

• The behavior of that vector can be understood as a mixture of the behavior of those bases.
The power method

• How to compute the dominant singular vector?

• How fast will it converge?

• How to compute the second most dominant singular vector?

• Give a power-method like way of computing the minimum eigenvector of A.
PCA

• The major axis of a covariance matrix
Spectral Clustering and Random Walks

• Let matrix $A$ represent a state transition probability matrix.

• Suppose we start at $t=0$ in a random state. What is the distribution of states we’re likely to be in at $t=\infty$?

  - $x = \text{random}(n, 1)$
  - $x' = Ax$
  - $\ldots x'''' = A^5x$