ROYAL INSTITUTE OF TECHNOLOGY

Linear Algebra, part 2
Eigenvalues, eigenvectors and least squares solutions

Anna-Karin Tornberg

Mathematical Models, Analysis and Simulation
Fall semester, 2012

Main problem of linear algebra 2:

Given an $n \times n$ matrix $\mathbf{A}$, find eigenvector(s) and eigenvalue(s) $\mathbf{x}$ and λ such that

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}.$$

Rewrite the equation as $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$.
This equation can only have a non-trivial solution if the matrix $(\mathbf{A} - \lambda\mathbf{I})$ is singular.

The number λ is an eigenvalue of A *if and only if* $det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

$det(\mathbf{A} - \lambda\mathbf{I})$ is a polynomial in λ of degree n.
Will have n roots, $\lambda_1, \ldots, \lambda_n$.
Can have multiplicity larger than one, and can have complex eigenvalues, even if A is real.

Can compute eigenvalues by hand using this approach for small systems. Other approaches needed for computer algorithms for large systems. Not the focus here.

Facts about eigenvalues

The product of the n eigenvalues equals the determinant of A: $\det(A) = (\lambda_1)\dots(\lambda_n)$.

The sum of the eigenvalues equals the trace (sum of diagonal entries) of $\mathbf{A}$: $\lambda_1 + \lambda_2 + \cdots + \lambda_n = a_{11} + a_{22} + \cdots + a_{nn}$.

If $\mathbf{A}$ is triangular, then its eigenvalues lie along the main diagonal.

The eigenvalues of A^2 are $\lambda_1^2, \dots, \lambda_n^2$.
The eigenvalues of A^{-1} are $1/\lambda_1, \dots, 1/\lambda_n$.
Eigenvectors of A are also eigenvectors of A^2 and A^{-1} (and any function of A).

Eigenvalues of $\mathbf{A} + \mathbf{B}$ and $\mathbf{AB}$ are in general not known from eigenvalues of $\mathbf{A}$ and $\mathbf{B}$, except for the special case when $\mathbf{A}$ and $\mathbf{B}$ commute, i.e. when $\mathbf{AB} = \mathbf{BA}$.

Linear differential equations with constant coefficients

The simple equation

$$\frac{dy}{dt} = ay$$

has the general solution $y(t) = Ce^{at}$. (Initial condition det. C).

Consider a system ($\mathbf{u}$ is $n \times 1$ vector, $\mathbf{A}$ is $n \times n$):

$$\frac{d\mathbf{u}}{dt} = \mathbf{Au}.$$

Let $\mathbf{x}_i$ be an eigenvector of $\mathbf{A}$ with corresponding eigenvalue λ_i, and define $\mathbf{u}_i = e^{\lambda_i t}\mathbf{x}_i$. We then have

$$\mathbf{Au}_i = \mathbf{A}(e^{\lambda_i t}\mathbf{x}_i) = e^{\lambda_i t}\mathbf{Ax}_i = e^{\lambda_i t}\lambda_i\mathbf{x}_i = \lambda_i\mathbf{u}_i$$

which is equal to $\frac{d\mathbf{u}_i}{dt} = \frac{d}{dt}(e^{\lambda_i t}\mathbf{x}_i) = \lambda_i e^{\lambda_i t}\mathbf{x}_i = \lambda_i\mathbf{u}_i$.
The general solution is

$$\mathbf{u}(t) = C_1 e^{\lambda_1 t}\mathbf{x}_1 + \cdots + C_n e^{\lambda_n t}\mathbf{x}_n$$

Growth and decay

For the scalar equation, solution $u(0)e^{at}$ decays if $a < 0$, grows if $a > 0$. If a is complex, the real part of a determines the growth or decay.

For the system, the λ_is determine which modes that will grow and which that will decay.

Example 1 - A 2×2 system with real eigenvalues.

Example 2 - Rigid body rotation: complex eigenvalues.

[NOTES]

Diagonalization of a matrix

Suppose that the $n \times n$ matrix $\mathbf{A}$ has n linearly independent eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Form a matrix $\mathbf{S}$, whose columns are the eigenvectors of $\mathbf{A}$. Then, $\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{\Lambda}$ is diagonal. The diagonal entries of $\mathbf{\Lambda}$ are the eigenvalues $\lambda_1, \ldots, \lambda_n$.

If A has no repeated eigenvalues, i.e. all λ_is are distinct, then $\mathbf{A}$ has n linearly independent eigenvectors, and $\mathbf{A}$ is diagonalizable.

If one or more eigenvalues of $\mathbf{A}$ have multiplicity larger than one $\mathbf{A}$ might have or might not have a full set of linearly independent eigenvectors.

If $\mathbf{A}$ is symmetric, all eigenvalues are real and it has a full set of *orthonormal* eigenvectors. Denote by $\mathbf{Q}$ the orthonormal matrix whose columns are eigenvectors of $\mathbf{A}$. (Defined on p 54 in Strang).
We have that $\mathbf{Q}^{-1} = \mathbf{Q}^T$ and $\mathbf{Q}^T\mathbf{A}\mathbf{Q} = \mathbf{\Lambda}$.

[WORKSHEET]

Vector and Matrix norms, Quadratic forms

Euclidean norm of $\mathbf{x}$: $\|\mathbf{x}\|_2 = \left(\sum_{k=1}^{n} |x_k|^2\right)^{1/2} = \sqrt{\mathbf{x}^T\mathbf{x}}$.

2-norm of matrix defined as : $\|\mathbf{A}\|_2 = \max \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \max \sqrt{\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}}$.

Consider the Rayleigh quotient: $R_{\mathbf{K}}(\mathbf{x}) = (\mathbf{x}^T\mathbf{K}\mathbf{x})/(\mathbf{x}^T\mathbf{x})$.

Differentiating $R_{\mathbf{K}}(\mathbf{x})$ with respect to x_k (DO IT!), can show that $\frac{\partial R}{\partial x_k} = 0,\ k = 1, \dots, n$ if and only if $\mathbf{K}\mathbf{x} = R_{\mathbf{K}}(\mathbf{x})\mathbf{x}$.

Theorem: If $\mathbf{K}\mathbf{x}^* = \lambda\mathbf{x}^*$, then $\mathbf{x}^*$ is a stationary point of $R_{\mathbf{K}}(\mathbf{x})$ and $\lambda = R_{\mathbf{K}}(\mathbf{x}^*)$.
Hence, the Rayleigh quotient is maximized by the largest eigenvalue of K.

Denote the largest eigenvalue of $\mathbf{A}^T\mathbf{A}$ by λ_M. We get $\|\mathbf{A}\|_2 = \sqrt{\lambda_M}$.
This is well defined, since $\mathbf{A}^T\mathbf{A}$ is SPD and all eigenvalues are positive.

Positive definite matrices and minimum principles

A symmetric matrix $\mathbf{A}$ is positive definite (SPD, introduced last time) if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all non-zeros vectors $\mathbf{x}$.
We know that a matrix is SPD if it is symmetric and all pivots are positive, or equivalently, that all eigenvalues are positive.

The solution $\mathbf{x}$ to $\mathbf{A}\mathbf{x} = \mathbf{b}$ can also be viewed as a solution to the following minimization problem:

If $\mathbf{A}$ is positive definite, then the quadratic $P(x) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b}$ is minimized at the point where $\mathbf{A}\mathbf{x} = \mathbf{b}$. The minimum value is $P(\mathbf{A}^{-1}\mathbf{b}) = -\frac{1}{2}\mathbf{b}^T\mathbf{A}^{-1}\mathbf{b}$.

Proof: [NOTES]

Example by calculus [NOTES]

Least squares solution

Consider again $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is $m \times n$ with $m > n$.
($\mathbf{x}$ is $n \times 1$, $\mathbf{b}$ is $m \times 1$.

Example: Find the line $y = c + dt$ that passes through four given points $(t_1, y_1), \ldots (t_4, y_4)$. If the four points all fall on a straight line, this overdetermined system has a solution. Otherwise, we want to find the straight line that "best" fit the data points.
[NOTES]

Normal equations

The vector $\mathbf{x}$ that minimizes $\|\mathbf{Ax} - \mathbf{b}\|^2$ is the solution to the *normal equations*

$$\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}.$$

This vector $\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$ is the *least squares solution* to $\mathbf{Ax} = \mathbf{b}$.

Proof: $\|\mathbf{Ax} - \mathbf{b}\|^2 = (A\mathbf{x} - \mathbf{b})^T(A\mathbf{x} - \mathbf{b}) = ((\mathbf{Ax})^T - \mathbf{b}^T)(A\mathbf{x} - \mathbf{b}) =$
$= \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} - \mathbf{b}^T\mathbf{Ax} + \mathbf{b}^T\mathbf{b}.$

$\mathbf{b}^T\mathbf{Ax}$ is a scalar, and can be transposed: $\mathbf{b}^T\mathbf{Ax} = (\mathbf{b}^T\mathbf{Ax})^T = \mathbf{x}^T\mathbf{A}^T\mathbf{b}$.
The term $\mathbf{b}^T\mathbf{b}$ is constant, and does not affect the minimization. So, the form to minimize (scaled by a factor of 2, which is arbitrary) is

$$P(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{A}^T\mathbf{b}.$$

According to earlier thm, this quadratic form is minimized for $\mathbf{x}$ which is a solution to $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$.

[NOTES CONTD.]