Numerical Linear Algebra

Charles J. Geyer
School of Statistics
University of Minnesota

Stat 8054 Lecture Notes

Numerical Linear Algebra is about

- solving linear equations
- matrix factorizations
- eigenvalues and eigenvectors

Before We Begin

One thing you never want to do is matrix inversion.

Think of it as solving linear equations instead. Not

$$\widehat{\beta} = (X^T X)^{-1} X^T y$$

Rather, $\widehat{\beta}$ is the solution of

$$X^T X \beta = X^T y$$

Gaussian Elimination

Gaussian elimination is how you were taught to solve systems of linear equations in high school.

- Solve one equation for one variable in terms of the rest.
- Plug that solution into the rest of the equations.
- The rest of the equations are now a system of equations with one less variable.
- Repeat.

Gaussian Elimination (cont.)

Eventual result

$$
\begin{aligned}
x_1 &= a_{12}x_2 + \cdots + a_{1n}x_n + b_1 \\
x_2 &= a_{23}x_3 + \cdots + a_{2n}x_n + b_2 \\
&\vdots \\
x_n &= b_n
\end{aligned}
$$

Now go from bottom up, plugging each solution into the one above.

LU Decomposition

The mathematical abstraction corresponding to Gaussian elimination is LU decomposition

$$M = LU$$

where L is lower triangular and U is upper triangular.

Triangular matrices are easy to invert or to solve, essentially the plug-in on the preceding page, more formally called *back substitution*.

Pivoting

LU decomposition can be very unstable (have large errors due to inexactness of computer arithmetic) if done without pivoting (carefully choosing the order in which variables are eliminated).

A *permutation matrix* P is an identity matrix whose rows have been permuted. It contains exactly one 1 in each row and each column.

Pivoting means we factor PM, which is M with rows permuted (row pivoting), or MP, which is M with columns permuted (column pivoting).

Pivoting (cont.)

Permutation matrices are trivial to invert.

LU decomposition with pivoting gives factorizations

$$M = P^{-1}LU$$
$$M = LUP^{-1}$$

Gram-Schmidt

Given a set of vectors v_1, ..., v_n compute an orthonormal basis by repeating the following.

- Orthogonalize

$$w_k = v_k - \sum_{i=1}^{k-1} u_i u_i^T v_k$$

- Normalize

$$u_k = \frac{w_k}{\sqrt{w_k^T w_k}}$$

QR Decomposition

The mathematical abstraction corresponding to Gram-Schmidt is QR decomposition

$$M = QR$$

where Q is orthogonal and R is right (upper) triangular.

Triangular matrices are easy to invert or to solve (back substitution). Orthogonal matrices too ($Q^T = Q^{-1}$).

QR also done with pivoting.

QR Decomposition (cont.)

$M = QR$ makes sense when the row dimension of M is at least as large as the column dimension.

In the *full* QR factorization Q is square and R the same shape as M.

In the *reduced* QR factorization R is square and Q the same shape as M.

QR Decomposition: Application

Least Squares

Normal equations for least squares are

$$M^T y = M^T M \beta$$

or

$$R^T Q^T y = R^T Q^T Q R \beta = R^T R \beta$$

If we use the reduced QR factorization and R is full rank (which happens when M is full rank), then

$$R\beta = Q^T y$$

and solving to get least squares estimates is back substitution.

Eigenvalues and Eigenvectors

The words annoy language purists, being German-English hybrids (*eigen* is German for proper or innate, terms that don't mix languages are *eigenwert* and *eigenvektor*). Eigenvalues are also called characteristic values and proper values (and similarly for vectors), but the "eigen" terms are most widely used.

If M is a matrix, v a vector, λ a scalar, and

$$Mv = \lambda v,$$

then λ is an eigenvalue of M corresponding to the eigenvector v.

Spectral Decomposition

The mathematical abstraction corresponding to eigenvalues and eigenvectors of a symmetric matrix is the spectral decomposition

$$M = ODO^T$$

where O is orthogonal and D is diagonal.

Diagonal matrices are trivial to invert or to solve. Orthogonal matrices too ($O^T = O^{-1}$).

Spectral Decomposition (cont.)

Rewrite $M = ODO^T$ as

$$MO = OD$$

Make summations explicit

$$\sum_j m_{ij} o_{jk} = \sum_l o_{il} d_{lk} = o_{ik} d_{kk}$$

(since D is diagonal).

Write v_k for k-th column of O and λ_k for the k-th diagonal element of D and this becomes

$$Mv_k = \lambda_k v_k$$

Spectral Decomposition (cont.)

Conclusion: the spectral decomposition theorem asserts that every symmetric $n \times n$ matrix has n eigenvectors that make up an orthonormal basis.

Spectral Decomposition: Applications

Positive Definiteness

$M = ODO^T$ is positive semidefinite if and only if all the eigenvalues (diagonal elements of D) are nonnegative.

M is positive definite if and only if all its eigenvalues are positive.

Symmetric Square Roots

If $M = ODO^T$ is positive semidefinite, then

$$M^{1/2} = OD^{1/2}O^T$$

is a square root of M, where $D^{1/2}$ is the elementwise square root of D.

Spectral Decomposition: Applications (cont.)

Multivariate Normal Random Vectors

If

$$\text{var}(X) = M = ODO^T$$

then

$$OD^{1/2}Z$$

has the same distribution as X, where Z is multivariate standard normal.

$$\text{var}(OD^{1/2}Z) = OD^{1/2}E(ZZ^T)D^{1/2}O^T = ODO^T = M$$

QR Algorithm

Unlike LU and QR decompositions, the ODO^T decomposition cannot be computed in a finite number of steps. An iterative algorithm must be used.

The best algorithm (for symmetric matrices) first factors

$$M = Q_0 A_1 Q_0^T$$

where Q_0 is orthogonal and A_1 tridiagonal (which can be done in finite number of steps). Then repeats

$$\begin{aligned} A_k &= Q_k R_k \\ A_{k+1} &= R_k Q_k \\ &= Q_k^T Q_k R_k Q_k \\ &= Q_k^T A_{k-1} Q_k \end{aligned}$$

QR Algorithm (cont.)

If A_k converges to a diagonal matrix (which it does under some conditions), then that is the D in the spectral decomposition. The O in the spectral decomposition is the product of the Q_k.

Actually, *best* algorithm is a modification: QR with shifts, which we don't explain.

Cholesky Decomposition

The *Cholesky decomposition* of a positive semidefinite symmetric matrix is

$$A = LL^T$$

where L is lower triangular.

Like LU and QR this is a finite algorithm.

Cholesky Decomposition (cont.)

$$a_{ik} = \sum_{j=1}^{\min(i,k)} l_{ij} l_{kj}$$

Compute columns of L in order.

$$a_{11} = l_{11}^2$$
$$a_{1k} = l_{11} l_{k1}$$

Solve for first column of L. Now

$$a_{22} = l_{21}^2 + l_{22}^2$$
$$a_{2k} = l_{21} l_{k1} + l_{22} l_{k2}$$

Solve for second column of L (using the values of l_{1k} we solved for in the first step and $l_{12} = 0$). And so forth.

Cholesky Decomposition (cont.)

For the i-th column of L we have

$$a_{ii} = \sum_{j=1}^{i} l_{ij}^2$$

which we can solve to get l_{ii} in terms of A and columns of L before the i-th, and then for $k < i$ we have

$$a_{ik} = \sum_{j=1}^{k} l_{mj} l_{kj}$$

which we can solve to get l_{ik} in terms of A and columns of L before the i-th and l_{ii}. And $l_{ik} = 0$ for $k > i$ by definition of lower triangular.

Singular Value Decomposition (SVD)

For any matrix A, not necessarily square

$$A = UDV^T$$

where U and V are orthogonal (hence square) and D is diagonal (but not square).

Singular values of A are diagonal elements of D.

$$AA^T = UDD^TU^T$$
$$A^TA = VD^TDV^T$$

are spectral decompositions of AA^T and A^TA because DD^T and D^TD are diagonal.

Eigenvalues of AA^T and A^TA are squares of singular values of A.

Singular Value Decomposition (cont.)

Useful for "solving" ill-conditioned systems of equations.

$$Ax = b$$

$$UDV^T x = b$$

$$DV^T x = U^T b$$

Small elements of D are essentially divide-by-zero. At least you know what you're up against!

Singular Value Decomposition: Application

Solution Path for Ridge Regression

Minimize

$$(y - M\beta)^T(y - M\beta) + \lambda\beta^T\beta$$

Strictly convex quadratic function so solution satisfies

$$2M^T(y - M\beta) + 2\lambda\beta = 0$$

or

$$M^T y = (M^T M + \lambda I)\beta$$

If $M = UDV^T$, then

$$M^T = (VD^T DV^T + \lambda I)\beta = V(D^T D + \lambda I)V^T\beta$$
$$\beta = V(D^T D + \lambda I)^{-1}V^T M^T y$$

Summary of Old Stuff

Methods for non-square complex matrices: SVD and QR.

Methods for square complex matrices: LU, eigenvalue-eigenvector (latter does not work for some matrices).

Methods for square positive semidefinite real matrices: LL^T.

A Really Good Reference

Trefethen and Bau

Numerical Linear Algebra

SIAM, 1997

Operation Counts

Taken from Trefethen and Bau in flops (floating point operations)

- LU is $\frac{2}{3}m^3$ flops.
- LL^T is $\frac{1}{3}m^3$ flops.
- QR is $2mn^2$ flops.

Operation Counts (cont.)

Eigendecomposition and SVD require an infinite number of operations.

Eigendecomposition proceeds in two stages. The finite stage, reduction to Hessenberg form takes $\frac{10}{3}m^3$ flops. The infinite stage takes $O(m)$ per iteration for symmetric matrices and $O(m^2)$ per iteration for non-symmetric matrices. How many iterations are done depends on the particular matrix.

SVD is not completely analyzed in Trefethen and Bau but is similar.

Break

Preceding stuff is the basics.

Now for some fancier stuff.

Normal Operators

- A matrix is *Hermetian* if $M = M^*$.
- A matrix is *unitary* if $U^*U = UU^* = I$ (the identity matrix).
- A matrix is *normal* if $MM^* = M^*M$.

Any normal matrix has a spectral decomposition

$$M = UDU^*$$

where D is diagonal and U is unitary.

If M is Hermetian, the eigenvalues are real, otherwise complex.

Non-Normal Operators

A non-normal matrix may (or may not) have eigenvalues and eigenvectors.

If its eigenvectors form a vector basis, then there is something sort of like the spectral decomposition but much less useful.

$$M = VDV^{-1}$$

where D is diagonal and the columns of V are the eigenvectors ($MV = VD$ is again the eigenvector equation in matrix notation).

But V is not orthogonal or unitary.

Matrix Norms

A *norm* for a vector space is a function $\|\cdot\|$ satisfying the following axioms

$$\|x\| \geq 0, \qquad \text{for all vectors } x$$
$$\|x\| = 0 \text{ if and only if } x = 0$$
$$\|cx\| = |c| \cdot \|x\|, \qquad \text{for all vectors } x \text{ and all scalars } c$$
$$\|x + y\| \leq \|x\| + \|y\|, \qquad \text{for all vectors } x \text{ and } y$$

Examples are the L^p norms

$$\|x\|_p = \left(\sum_{1 \leq i \leq d} |x_i|^p \right)^{1/p}$$

and

$$\|x\|_\infty = \max_{1 \leq i \leq d} |x_i|$$

Matrix Norms (cont.)

Matrices of the same dimension form a vector space (they can be added and multiplied by scalars).

Particularly useful matrix norms are related the fact that square matrices A represent linear operators $x \mapsto Ax$.

If $\|\cdot\|$ denotes a norm for the vector space V on which the matrix acts (if A is $d \times d$, then it acts on d-dimensional vectors), then

$$\|A\|_{\text{matrix}} = \max_{\substack{x \in V \\ x \neq 0}} \frac{\|Ax\|}{\|x\|}$$

is called the *operator norm* for matrices, corresponding to the vector norm $\|\cdot\|$.

Matrix Norms (cont.)

Since the 2-norm

$$\|x\|_2 = \sqrt{\sum_{1 \le i \le d} x_i^2}$$

is widely used for vectors, the operator 2-norm

$$\|A\|_{2\mathrm{M}} = \max_{\substack{x \in V \\ x \neq 0}} \frac{\|Ax\|_2}{\|x\|_2}$$

is widely used for matrices.

Pseudospectra

The set $\sigma(A)$ of eigenvalues of a matrix A is called its *spectrum*.

The set $\sigma_\epsilon(A)$ of eigenvalues of $A + E$, where $\|E\| < \epsilon$, is called its ϵ-*pseudospectrum* (this depends on the norm used).

For a normal operator, $\sigma_\epsilon(A)$ is just the set of points less than ϵ away from $\sigma(A)$ when the operator 2-norm is used.

For a non-normal operator, $\sigma_\epsilon(A)$ can be arbitrarily larger than $\sigma(A)$.

Eigenvalues don't tell you much about non-normal matrices.

Bauer-Ficke Theorem

B_ϵ is open ball of radius ϵ in complex plane.

$$\kappa(A) = \frac{\|V\|_{2\mathsf{M}}}{\|V^{-1}\|_{2\mathsf{M}}}$$

is 2-norm condition number of A.

$$\sigma(A) + B_\epsilon \subset \sigma_\epsilon(A) \subset \sigma(A) + B_{\epsilon \cdot \kappa(A)}$$

Outer bound humongous if A ill-conditioned, i. e., $\kappa(A)$ large. (Holds for all invertible and diagonalizable A).

Schur Decomposition

The eigenvectors of a square matrix need not form a vector basis, in which case there is no $M = VDV^{-1}$ decomposition, in which case we say M is not *diagonalizable*.

But there is always a *Schur decomposition*

$$M = QUQ^*$$

where Q is unitary and U is upper triangular.

Change of Basis

Any invertible matrix V can be thought of as a change of basis.

Its rows (resp. columns) are linearly independent and make up a basis.

Its action on vectors is to transform from old basis to new. If

$$y = Vx,$$

then

$$x = \sum_i x_i e_i$$

where e_i are vectors of the standard basis (rows of the identity matrix) and

$$y = \sum_i y_i v_i$$

where v_i are vectors of the new basis (rows of V).

Similarity Transformations

If A is any square matrix (not necessarily invertible), then

$$y = Ax$$

is a linear transformation represented in the standard basis. In the new basis

$$x_{\text{new}} = Vx$$
$$y_{\text{new}} = Vy$$

and

$$y_{\text{new}} = Vy = VAx = VAV^{-1}x_{\text{new}}$$

so

$$VAV^{-1}$$

represents the same abstract linear transformation in the new basis.

Similarity Transformations (cont.)

$A \mapsto VAV^{-1}$ is called a *similarity transformation*.

A and VAV^{-1} are *similar* matrices.

Similarity Transformations do not Change Spectra

$$Ax = \lambda x$$

$$y = Vx$$

$$AV^{-1}y = \lambda V^{-1}y$$

$$VAV^{-1}y = \lambda y$$

Eigenvalues of Triangular Matrices

λ is eigenvalue of A if and only if $A-\lambda I$ is singular, which happens if and only if $\det(A - \lambda I) = 0$.

If A is upper triangular with elements a_{ii}, then

$$\det(A - \lambda I) = \prod_{i=1}^{n} (a_{ii} - \lambda)$$

Hence every diagonal element of a triangular matrix is an eigenvalue, and vice versa.

Eigenvalues of Triangular Matrices (cont.)

$$A = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

Only eigenvalue is 2, and only one eigenvector.

Solve $(A - 2I)x = 0$ for eigenvector(s) x.

$$(A - 2I)x = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_3 \\ x_4 \\ x_5 \\ 0 \end{pmatrix} = 0$$

Eigenvalues of Triangular Matrices (cont.)

$$\begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

has only one eigenvector

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Corresponding eigenvalue is 2.

New Stuff

Among the methods we studied LU, QR, and LL^T are finite. The spectral and singular value decompositions are not finite (take an infinite number of iterations to get infinite precision, which, of course, the computer doesn't have).

For large matrices, the finite methods don't pay. They do too much work. Good approximations can be calculated faster.

These methods are

	general matrices	symmetric matrices
Eigenvalue-Eigenvector	Arnoldi iteration	Lanczos iteration
Solving Linear Equations	GMRES	CG, MinRes