

Things You Should Know (from Linear Algebra)

Jonathan Harel

April 11, 2009

Contents

1	Spaces	1
2	Eigenvalues and Eigenvectors	1
2.1	Diagonalization	2
3	Singular Value Decomposition (SVD)	3
4	Symmetric Matrices	3
5	Positive Definite Matrices	4
6	Principal Component Analysis (PCA)	5

1 Spaces

Matrices have $\dim(\text{row-space}) = \dim(\text{col-space})$. For example, a random 2×60 matrix has a row-space spanning a planar subspace in R^{60} (with high probability).

Let A be an $n \times n$ matrix. Then

$$Ax = 0$$

has solutions $x \neq 0$ ($x \in R^n$) iff A is not full rank.

2 Eigenvalues and Eigenvectors

Defined only on $n \times n$ matrices.

Eigenvalues and eigenvectors are the solutions to

$$Ax = \lambda x \tag{1}$$

This is tricky to solve, since we don't know x or λ .

However, eigenvectors must be solutions to:

$$(A - I\lambda)x = 0.$$

that is, eigenvector e_i with eigenvalue λ_i is in the null-space of $A - I\lambda_i$. We can, in principle, solve the eigensystem by first obtaining the eigenvalues through

$$\det(A - \lambda I) = 0 \tag{2}$$

which has n unknowns and n equations, and then solving for the vectors through (1).

An aside on solving for eigensystems Actually solving the characteristic polynomial (2) is expensive and unreliable. Instead, iterative methods are used which exploit that the number and signs of the pivots are the same as the number and signs of the eigenvalues (e.g., N positive pivots, M negative pivots $\Rightarrow M$ positive eigenvalues, N negative eigenvalues).

2.1 Diagonalization

Suppose the column vectors of S are the eigenvectors of A (and that they fully span the row-space of A , more on this below). Then,

$$AS = S\Lambda,$$

where Λ is the diagonal matrix containing the eigenvalues of A . Therefore,

$$A = S\Lambda S^{-1}.$$

This is called the eigen, or spectral, decomposition, or diagonalization of A .

A matrix without a full set of eigenvectors has a "Jordan Block" of size 2 or greater. Such a matrix is improbable (e.g., sampled from the set of square matrices whose entries are i.i.d. normal), can be full rank or not, and intuitively corresponds to a linear transformation which cannot be fully decoupled to scaling along linearly independent axes. A shearing transformation is such a transformation.

Note 1: The diagonalization of A can be used to easily compute A^n through:

$$A^n = S\Lambda^n S^{-1}.$$

Note 2: You can form another matrix M with the same eigenvalues, by choosing an invertible matrix T of the same dimensions as A and setting

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

Then,

$$M = TAT^{-1} = TSAS^{-1}T^{-1} = (TS)\Lambda(TS)^{-1}$$

so M has the same eigenvalues as A , and its eigenvectors are transformed by T . (In general, $(AB)^{-1}AB = I \Rightarrow (AB)^{-1}A = IB^{-1} \Rightarrow (AB)^{-1} = B^{-1}A^{-1}$).

Two fun facts about eigenvalues. The trace Tr

$$\text{Tr}(A) = \sum_i \lambda_i = \text{sum of diagonal elements of } A.$$

Also,

$$\prod_i \lambda_i = \det(A)$$

Meaning of eigenvalue with multiplicity $m > 1$. Eigenvalue with multiplicity m has eigenvectors which span an m dimensional subspace. Identity matrix has eigenvalue 1 with multiplicity n – the eigenvectors span R^n .

3 Singular Value Decomposition (SVD)

This is analogous to spectral/eigen decomposition for nonsquare matrices (though, even for square matrices, it is only exactly equivalent for symmetric matrices). Let A be an $m \times n$ matrix, then it can be written:

$$A = U\Sigma V^T$$

where Σ is diagonal, and U and V are orthonormal ("unitary" in general), U is $m \times m$, Σ is $m \times n$, and V^T is $n \times n$. Remark: the values in Σ are the singular values, and the columns of U and rows of V^T are the left and right singular vectors.

The Eckhart-Young theorem states that the matrix of rank $r \leq n$ which minimizes the absolute-square ("Frobenius norm") error between \tilde{A} and A is

$$\tilde{A} = U\tilde{\Sigma}V^T$$

where $\tilde{\Sigma}$ is Σ , except with the last $n - r$ diagonal entries set to zero.

4 Symmetric Matrices

Symmetric matrices:

- (1) have real eigenvalues
- (2) have an orthogonal set of eigenvectors
- (3) have singular values equal to their eigenvalues

Observations: due to (2), a symmetric $n \times n$ matrix S can be written

$$S = Q\Lambda Q^T$$

since a matrix (with a full set of eigenvectors) can be written $S = Q\Lambda Q^{-1}$, but due to orthonormality of Q , $QQ^T = I$, thus $Q^{-1} = Q^T$.

Proof that the eigenvectors must be orthonormal (probably): Assume S is symmetric. The eigendecomposition is $S = Q\Lambda Q^{-1}$. Suppose that $Q^{-1} \neq kQ^T$ for any scalar k , then S is not symmetric, violating our assumption (this is probably true, not entirely sure – it seems any other Q^{-1} would give rise to a non-symmetric matrix).

Connection to singular value decomposition: We can always write $M = U\Sigma V^T$, where U and V are unitary, but for a symmetric matrix S , $S = S^T$, therefore¹

$$S = U\Sigma V^T = S^T = V\Sigma U^T$$

which is satisfied if and only if $V = U$, therefore

$$S = U\Sigma U^T$$

which is precisely the eigendecomposition of S , therefore $\Sigma = \Lambda$, that is, for a symmetric matrix S , the singular values are the eigenvalues.

More generally, suppose M , with singular value decomposition $U\Sigma V^T$, is any (possibly non-square) matrix. Then

$$MM^T = U\Sigma V^T V\Sigma U^T$$

but $VV^T = I$, since V is orthonormal (or in general, $VV^* = I$ for unitary V), thus

$$MM^T = U\Sigma^2 U^T = U\Lambda U^T$$

That is, $\Lambda = \Sigma^2$: the eigenvalues Λ of the symmetric matrix MM^T are the squared singular values Σ of M . Furthermore, the eigenvectors of MM^T are simply the left singular vectors of M .

5 Positive Definite Matrices

Positive definite matrices are square, symmetric, have all positive eigenvalues, and all n subdeterminants are also positive (subdeterminant is determinant of a square upper left corner of matrix).

In particular, M is positive-definite iff $z^T M z > 0 \forall z \in R^n - 0$, and M is symmetric. Also, M is positive-definite iff all its eigenvalues are positive. Also M is positive definite iff all upper left submatrices of M have positive determinants.

Note that

$$z^T M z = \sum_{ij} z_i z_j M_{ij}$$

¹Recall $(ABC)^T = C^T B^T A^T$

6 Principal Component Analysis (PCA)

Suppose we have an $m \times n$ matrix A , in which we interpret the COLUMNS to be sample vectors. For any of the m components of each sample vector, we can define the sample variance (we assume zero-mean data), and we wish to find directions in R^m which have maximal variance over the set. The idea is that a few such directions span a subspace of R^m which captures most of the variance in the data. We can then project a new vector onto this subspace as a means of reducing dimensionality.

If the rows of A sum to zero, then the covariance matrix is an $m \times m$ matrix C such that:

$$\begin{aligned} C &= E[a_i a_j] \\ &= \text{const} \cdot AA^T \end{aligned}$$

where a_i and a_j are the i^{th} and j^{th} components ($i, j \in \{1, 2, \dots, m\}$) from the same randomly selected sample vector, and $\text{const} = \frac{1}{n}$. We will drop this leading constant from here on as it does not play a role in the analysis. The variance of the data in A projected onto an arbitrary direction v turns out to be:

$$\text{var}(v) = v^T C v = \sum_{i,j} v_i v_j C_{ij}. \quad (3)$$

Since variance must be positive, then $v^T C v > 0$ for all real nonzero v , and because C is also symmetric, C must be positive semi-definite².

Fun fact: It turns out that all covariance matrices are positive semi-definite, and all positive semi-definite matrices are covariance matrices.

According to optimization theory, the variance quantity (3) is maximized (subject to $v^T v = 1$) when

$$\begin{aligned} \text{grad}(v^T C v) &= \lambda \cdot \text{grad}(v^T v) \text{ [grad wrt components of } v\text{]} \\ \Rightarrow C v &= \lambda v \end{aligned}$$

Therefore, the direction with maximum variance is an eigenvector of C . We note that the variance of an eigenvector is

$$e^T C e = e^T \lambda e = \lambda.$$

So the line along the principal eigenvector of C is the one-dimensional subspace in which the data (projected onto such a subspace) in A have the greatest variance. The direction orthogonal to that eigenvector with the largest variance among all possible orthogonal directions is given by the second eigenvector, etc.. The eigenvectors are orthogonal because C is symmetric.

²positive semi-definite instead of positive-definite because you could have zero variance (=0 eigenvalue) along even the most variant directions, if you have degenerate data.

Thus, if

$$C = U\Lambda U^T$$

is the eigendecomposition of C , then

$$A = U\sqrt{\Lambda}V^T$$

is the singular value decomposition of A , as demonstrated in the previous section.

In matlab, if you have an $m \times n$ matrix A with rows summing to zero, where the columns index sample vectors, and the rows index their components, then

$$\begin{aligned} [u \ s \ v] &= \text{svd}(A) \\ \Rightarrow u(:, i) &\text{ is the } i^{\text{th}} \text{ most variant direction (columns of } U) \\ \Rightarrow A_r &= u(:, 1:r)' \cdot A \text{ is the dimensionally reduced data for } r < m. \end{aligned}$$

You can verify that

- (1) $\text{var}(u' \cdot A, 1, 2)$ is decreasing
- (2) $\text{sqrt}(\text{sum}((u' \cdot A).^2, 2))$ is the same as $\text{diag}(s)$
- (3) $\text{diag}(s).^2$ is n times the variance along the principal components
- (4) $\text{diag}(s)$ is the same as $\text{flipud}(\text{sqrt}(\text{eig}(A \cdot A')))$

Also, in matlab

$$v = u(:, 1:r) \cdot u(:, 1:r)' \cdot A$$

is the original data, but with only the components along the reduced axes retained, and the rest truncated (the dimensionality-reduced data in the original coordinate space). You can verify that the scatter plot

$$\text{plot}(v(:, :), A(:, :), ' .')$$

converges to the line $y=x$ as r increases to m .