

Computing the Singular Value Decomposition

Reading GV96 Chapter 8.3 and 8.6, 5.1 and 5.2

Recall we have $A \in \mathbb{C}^{m \times n}$ $m \geq n$, and unitary matrices $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ such that

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^H$$

where $\Sigma \in \mathbb{R}^{n \times n}$ is a diagonal matrix with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

There are many ways to compute all or part of this decomposition. We first relate it to a somewhat simpler problem for Hermitian or symmetric matrices.

DEFINITION: If $M \in \mathbb{C}^{n \times n}$ is an Hermitian matrix then there exists a unitary matrix $Q \in \mathbb{C}^{n \times n}$ such that

$$M = Q\Lambda Q^H$$

where Λ is a diagonal matrix with real entries.

DEFINITION: The columns of Q are the eigenvectors of M and the diagonal elements of Λ are the corresponding eigenvalues. We have

$$MQ = Q\Lambda$$

$$MQe_i = Q\Lambda e_i$$

$$Mq_i = q_i\lambda_i$$

$$(M - \lambda_i I)q_i = 0$$

$$\det(M - \lambda_i I) = 0$$

So we see that an eigenvalue is a shift that makes M singular, and that the eigenvectors are mutually orthogonal directions in \mathbb{C}^n along which the action of M is equivalent to scaling. Also, the eigenvalues are the roots of the determinant of $M - \lambda I$.

The eigenvectors are unique up to scaling and for repeated eigenvalues up to the choice of basis for the associated eigenspace.

LEMMA: The SVD can be related to two Hermitian eigendecompositions.

$$\begin{aligned} A &= U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^H \\ A^H A &= V \Sigma^2 V^H \\ A A^H &= U \begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} U^H \end{aligned}$$

So $A^H A$ and $A A^H$ have nonzero eigenvalues that are the square of the singular values of A . The eigenvectors are the right and left singular vectors respectively. Note the fact that the singular values are nonnegative is reflected in the fact that the Hermitian matrices are both positive semidefinite.

The Hermitian (symmetric) eigenproblem has a rich theory behind it and there are many ways to approximate it.

We will concentrate on the so-called symmetric QR algorithm for real symmetric matrices.

We start by assuming a simpler symmetric problem, i.e.,

DEFINITION: $T \in \mathbb{R}^{n \times n}$ is a symmetric tridiagonal matrix if and only if $T^T = T$ and $e_i^T T e_j = 0$ when $i > j + 1$. In other words, $e_i^T T e_j \neq 0$ only when $i = j$ (the main diagonal), $i = j + 1$ (the subdiagonal), or $i + 1 = j$ (the superdiagonal).

LEMMA: If $e_{i+1}^T T e_i = 0$ then the eigendecomposition can be split into two eigendecompositions since T and Q are both block diagonal.

$$T = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} = \begin{pmatrix} Q_1 \Lambda_1 Q_1^T & 0 \\ 0 & Q_2 \Lambda_2 Q_2^T \end{pmatrix}$$

So we may assume that the matrix T is **unreduced**, i.e., all of the elements in the super and subdiagonals are nonzero.

The eigendecomposition cannot be computed in a finite number of computations for an arbitrary matrix so we must produce an iteration that converges to the eigenvectors and eigenvalues.

We exploit the properties of a similarity transformation.

LEMMA: Any two matrices A and $M = ZAZ^{-1}$ are called **similar** and they have the same eigenvalues.

Proof:

$$\begin{aligned}Ax &= x\lambda \\ZAx &= Zx\lambda \\ZA(Z^{-1}Z)x &= Zx\lambda \\(ZAZ^{-1})(Zx) &= (Zx)\lambda \\M(Zx) &= (Zx)\lambda \\My &= y\lambda\end{aligned}$$

The symmetric QR algorithm is based on repeated QR factorizations of a symmetric tridiagonal matrix.

LEMMA: If $T = QR$ where T is symmetric tridiagonal, R is upper triangular and Q is orthogonal then R has only two nonzero super-diagonals and

$$\begin{aligned}T_+ &= RQ \\ &= Q^T(QR)Q \\ &= Q^T T Q\end{aligned}$$

is symmetric tridiagonal and has the same eigenvalues as T .

Since $Q^{-1} = Q^T$, we have similar symmetric tridiagonal matrices T and T_+ .

Given's Rotations

Given's rotations are 2×2 orthogonal matrices that can be used to introduce 0's into selected positions of matrices with structure. We will use them to compute the QR factorization of T .

DEFINITION: Let θ be an angle through which you wish to rotate in \mathbb{R}^2 in a counterclockwise direction. Then the matrix

$$G(\theta) = \begin{pmatrix} \gamma & \sigma \\ -\sigma & \gamma \end{pmatrix}$$

where $\gamma = \cos(\theta)$ and $\sigma = \sin(\theta)$. is the Given's rotation that rotates a vector $x \in \mathbb{R}^2$

$$\begin{aligned} y &= G^T(\theta)x \\ &= \begin{pmatrix} \gamma\xi_1 - \sigma\xi_2 \\ \sigma\xi_1 + \gamma\xi_2 \end{pmatrix} \end{aligned}$$

It is easy to verify that $G^T G = I_2$ and if we take

$$\begin{aligned}\gamma &= \xi_1 \rho^{-1} \\ \sigma &= -\xi_2 \rho^{-1} \\ \rho &= \|x\|_2\end{aligned}$$

then

$$\begin{aligned}y &= G^T(\theta)x \\ &= \begin{pmatrix} \gamma\xi_1 - \sigma\xi_2 \\ \sigma\xi_1 + \gamma\xi_2 \end{pmatrix} \\ &= \begin{pmatrix} \rho \\ 0 \end{pmatrix}\end{aligned}$$

See pp. 216 and 217 on how to compute the rotation reliably and how to represent it by a single scalar.

EXAMPLE: $n = 5$

$$T = \begin{pmatrix} * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ 0 & * & * & * & 0 \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}$$

$$G^T(1,2)T = \begin{pmatrix} * & * & + & 0 & 0 \\ 0 & * & * & 0 & 0 \\ 0 & * & * & * & 0 \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}$$

$$G^T(2,3)G^T(1,2)T = \begin{pmatrix} * & * & + & 0 & 0 \\ 0 & * & * & + & 0 \\ 0 & 0 & * & * & 0 \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}$$

$$G^T(3,4)G^T(2,3)G^T(1,2)T = \begin{pmatrix} * & * & + & 0 & 0 \\ 0 & * & * & + & 0 \\ 0 & 0 & * & * & + \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}$$

$$G^T(4,5)G^T(3,4)G^T(2,3)G^T(1,2)T = \begin{pmatrix} * & * & + & 0 & 0 \\ 0 & * & * & + & 0 \\ 0 & 0 & * & * & + \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \end{pmatrix}$$

Each of the $G(i, i + 1)$ are the identity except for rows $i + 1$ and i which have the appropriate 2×2 rotation in the diagonal block.

$$G^T(2, 3) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & * & * & 0 & 0 \\ 0 & * & * & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

This is easily generalized to any two rows defining the plane of rotation, e.g., $G(i, j)$. See the text in Chapter 5 for details.

Shifts

The use of shifts does not change the structure of our basic step and it accelerates convergence.

LEMMA: Let $T - \mu I = QR$ where T is symmetric tridiagonal and $\mu \in \mathbb{R}$ then

$$\begin{aligned} T_+ &= RQ + \mu I \\ &= Q^T(QR)Q + \mu I \\ &= Q^T(T - \mu I)Q + \mu I \\ &= Q^T(T - \mu I)Q + \mu Q^T Q \\ &= Q^T(T - \mu I + \mu I)Q \\ &= Q^T T Q \end{aligned}$$

is symmetric tridiagonal and has the same eigenvalues as T .

If we knew an eigenvalue and used it as a shift on an unreduced T we have a very special form of QR .

LEMMA: Since T is unreduced the subdiagonal is nonzero and therefore independent of the shift μ the first $n-1$ columns of $T - \mu I$ are linearly independent and have a full column rank QR factorization $Q_{n-1}R_{n-1}$ and therefore if $\mu = \lambda$

$$T - \lambda I = \begin{pmatrix} Q_{n-1} & q_n \end{pmatrix} \begin{pmatrix} R_{n-1} & r \\ 0_{n-1}^T & 0 \end{pmatrix},$$

i.e., $e_n^T R e_n = 0$ to indicate an exact eigenvalue.

LEMMA: If λ is an eigenvalue of T then

$$\begin{aligned} QR &= T - \lambda I \\ T_+ &= RQ + \lambda I \\ T_+ e_n &= \lambda e_n \end{aligned}$$

So the eigenvalue appears in the (n, n) position of the symmetric tridiagonal matrix produced by reversing and shifting the QR factorization.

Explicit Shift Tridiagonal Symmetric QR

```
do until convergence to  $\lambda$  in  $(n, n)$ 
  choose a shift  $\mu$ 
  compute  $QR = T - \mu I$ 
   $T \leftarrow RQ + \mu I$ 
end do until
```

This identifies one of the eigenvalues and then the problem can be reduced by one, i.e., you work with the unreduced matrix defined by the first $n - 1$ rows and columns.

How to we choose the shift?

$$T = \begin{pmatrix} \alpha_1 & \beta_1 & & \cdots & 0 \\ \beta_1 & \alpha_2 & \beta_2 & \cdots & \vdots \\ & \cdots & \cdots & \cdots & \\ \vdots & & \beta_{n-2} & \alpha_{n-1} & \beta_{n-1} \\ 0 & \cdots & & \beta_{n-1} & \alpha_n \end{pmatrix}$$

Wilkinson shift takes μ to be the eigenvalue of

$$\begin{pmatrix} \alpha_{n-1} & \beta_{n-1} \\ \beta_{n-1} & \alpha_n \end{pmatrix}$$

This is known analytically as

$$\mu = \alpha_n + \delta - \text{sign}(\delta) \sqrt{\delta^2 + \beta_{n-1}^2}$$

where $\delta = \frac{\alpha_{n-1} - \alpha_n}{2}$

Convergence is cubic, i.e., the $(n, n - 1)$ element goes from ϵ to $O(\epsilon^3)$ on one step.

- The products of the Given's rotations can be accumulated if the eigenvectors are desired.
- If only selected eigenvectors are desired then inverse iteration can be used given an estimate of the eigenvalue, i.e.,

$$q_i^{(k)} \approx (T - \lambda I)^{-1} q_i^{(k-1)}$$

usually approximates q_i within one or two steps.

- The explicit shifting can be avoided by the Implicit Shift form.
- Then we must discuss how to apply this to the SVD problem.