8_6 Singular value decomposition

DIAGONALIZATION USING TWO ORTHOGONAL MATRICES

Diagonalizations

- $A=PDP^T$. A symmetric P orthogonal
- $A=PHP^T$ Hessenberg A non-symmetric
- $A=PSP^T$ Schur decomposition
- $A=PJP^{-1}$, A any J Jordan form, P invertible only. This is sensitive to round off errors.
- $A=USV^T$, U,V orthogonal, S diagonal with positive or zero entries in the diagonal.

Theorem 8.6.1 *If A is an $n \times n$ matrix of rank k, then A can be factored as*

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

where U and V are $n \times n$ orthogonal matrices and Σ is an $n \times n$ diagonal matrix whose main diagonal has k positive entries and $n - k$ zeros.

- proof: A^TA is symmetric.
  - $A^TA=VDV^T$ for D diagonal, V orthogonal.
  - The diagonal elements of D are eigenvalues of A^TA. The column vectors of V are eigenvectors of A^TA.
  - If x is an eigenvector of A^TA, then Ax.Ax=x.A^TAx=x.lx=l(x.x), l is nonnegative.
  - Rank A=rank A^TA=rank D. (Th. 7.5.8,8.2.3.)
  - We let V be arranged so that the corresponding eigenvalues are decreasing.
  - Thus l_1≥l_2≥…≥l_k>0, l_k+1=..=l_n=0.

- Consider $\{Av_1, Av_2, \ldots, Av_n\}$
- $Av_i . Av_j = v_i . A^T Av_j = v_i . l_j v_j = l_j (v_i . v_j) = 0$ for $i \neq j$ by the orthogonality of v_is.
- $||Av_i||^2 = Av_i . Av_i = v_i . A^T Av_j = v_i . l_i v_i = l_i (v_i . v_i) = l_i$.
- $||Av_i|| = \sqrt{l_i}$.
- $\{Av_1, \ldots, Av_k\}$ the basis of the column space of A. (col rank A=rank A=k)
- We normalize to obtain $u_1, \ldots, u_k$.
- $u_j = Av_j / ||Av_j|| = Av_j / \sqrt{l_j}$. $Av_j = \sqrt{l_j} u_j$
- Extend to an orthonormal basis $u_1, \ldots, u_n$.
- Let $U = [u_1, .., u_k, u_{k+1}, \ldots, u_n]$

- Let S be the diagonal matrix with diagonal entries $\sqrt{l_1}, \sqrt{l_2}, .., \sqrt{l_k}, 0, .., 0$.
- Then $US = [\sqrt{l_1} u_1, \sqrt{l_2} u_2, \dots, \sqrt{l_k}, 0, .., 0]$ $= [Av_1, Av_2, .., Av_k, Av_{k+1}, .., Av_n] = AV$.
- Thus, $A = USV^T$.

Theorem 8.6.2 (***Singular Value Decomposition of a Square Matrix***) *If A is an $n \times n$ matrix of rank k, then A has a singular value decomposition $A = U\Sigma V^T$ in which:*

(*a*) $V = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n]$ *orthogonally diagonalizes* A^TA.

(*b*) *The nonzero diagonal entries of* Σ *are*

$$\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \ldots, \sigma_k = \sqrt{\lambda_k}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ *are the nonzero eigenvalues of* A^TA *corresponding to the column vectors of* V.

(*c*) *The column vectors of* V *are ordered so that* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$.

(*d*) $\mathbf{u}_i = \dfrac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \dfrac{1}{\sigma_i}A\mathbf{v}_i \qquad (i = 1, 2, \ldots, k)$

(*e*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *is an orthonormal basis for* $\text{col}(A)$.

(*f*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ *is an extension of* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *to an orthonormal basis for* R^n.

- Example 1.
- Singular value decomposition of symmetric matrices.
  - A symmetric.
  - $A=PDP^T$.
  - D may have negative eigenvalues.
  - Let S be the diagonal matrix with the absolute values of the diagonal entries of D arranged the right way.
  - Then $A=PSV^T$. We put some negative signs to the columns of V.
- Example 2.

Polar decompositions

Theorem 8.6.3 (***Polar Decomposition***) *If A is an $n \times n$ matrix of rank k, then A can be factored as*

$$A = PQ \tag{9}$$

where P is an $n \times n$ positive semidefinite matrix of rank k, and Q is an $n \times n$ orthogonal matrix. Moreover, if A is invertible (rank n), then there is a factorization of form (9) in which P is positive definite.

- Proof: $A=USV^T=(USU^{T)(}UV^T) =PQ$
  - rank P=rankS=k.
  - A invertible -> k=n -> S positive definite -> P positive definite.
- Example 3.

Theorem 8.6.4 (***Singular Value Decomposition of a General Matrix***) *If A is an $m \times n$ matrix of rank k, then A can be factored as*

$$A = U\Sigma V^T = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k \mid \mathbf{u}_{k+1} \quad \cdots \quad \mathbf{u}_m] \left[\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \cdots & 0 & O_{k\times(n-k)} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \sigma_k & \\ \hline & O_{(m-k)\times k} & & & O_{(m-k)\times(n-k)} \end{array}\right] \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \\ \hline \mathbf{v}_{k+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \tag{12}$$

in which U, Σ, and V have sizes $m \times m$, $m \times n$, and $n \times n$, respectively, and in which:

(*a*) $V = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n]$ *orthogonally diagonalizes A^TA.*

(*b*) *The nonzero diagonal entries of Σ are $\sigma_1 = \sqrt{\lambda_1}$, $\sigma_2 = \sqrt{\lambda_2}, \ldots, \sigma_k = \sqrt{\lambda_k}$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the nonzero eigenvalues of A^TA corresponding to the column vectors of V.*

(*c*) *The column vectors of V are ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$.*

(*d*) $\mathbf{u}_i = \dfrac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \dfrac{1}{\sigma_i} A\mathbf{v}_i \qquad (i = 1, 2, \ldots, k)$

(*e*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *is an orthonormal basis for* $\operatorname{col}(A)$.

(*f*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ *is an extension of* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *to an orthonormal basis for R^m.*

- u_1,…,u_k, the left singular vectors of A.
- v_1,…,v_k, the right singular vectors of A.
- Example 4.

Singular value decompositions and the fundamental spaces

Theorem 8.6.5 *If A is an $m \times n$ matrix with rank k, and if $A = U\Sigma V^T$ is the singular value decomposition given in Formula* (12), *then:*

(*a*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *is an orthonormal basis for* $\text{col}(A)$.

(*b*) $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ *is an orthonormal basis for* $\text{col}(A)^\perp = \text{null}(A^T)$.

(*c*) $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ *is an orthonormal basis for* $\text{row}(A)$.

(*d*) $\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ *is an orthonormal basis for* $\text{row}(A)^\perp = \text{null}(A)$.

- Proof: (a) u_1,..,u_k. normalized from Av_is. Thus a basis of col(A).
- (b) $\text{col}(A)^c$ has basis u_k+1,..,u_n

- (d): v_1,..,v_n orthonormal set of eigenvectors of A^TA.
  - v_k+1, …, v_n corr to 0.
  - Thus v_k+1,..,v_n the orthonormal basis of null A^TA=nullA of dim n-k.
  - (d) proved.
- (c): v_1,..,v_k. are in null$(A)^c$=row(A).
  - row(A) has dimension k. Thus, v_1,..,v_k form an orthonormal basis of row(A).

Reduced singular value decompositions

- We can remove zero rows and zero columns from S.
- We also eliminate u_k+1,..,u_n, v^T_k+1,…v^T_n.
- $A=U\_1^{mxk}S\_1^{kxk}V\_1^{kxn}$.
- $A=s\_1u\_1v\_1^T+s\_2u\_2v\_1^T+\ldots+s\_ku\_k v\_k^T$.
- Example 5.

Data compression and image processing.

- We can omit small terms in $A=s_1u_1v_1^T+s_2u_2v_1^T+\ldots+s_ku_kv_k^T$.
- This decrease the amount one has to store and get approximate images.

Singular value decomposition from the transformation point of view.

- T_A:R^n->R^m
- Use basis B=[v_1,…,v_n] for R^n.
- B'= [u_1,..,u_n] for R^m.
- Then [T_A]_B,B'=S.
- Thus, in this coordinate, one collapses in v_k+1,..,v_n direction and multiply by s_1,..,s_k in u_1,…,u_k direction….

8_7 Pseudo-inverse

- $A=U\_1S\_1V\_1^T$. mxk, kxk,nxn.
- If A is an invertible nxn-matrix, then S_1 is nxn and so U_1,V_1 are nxn.
- $A^{-1}= V\_1S\_1^{-1}U\_1^T$.
- Suppose A is not nxn or invertible, then k<n.
- We define pseudo-inverse
  $A^{+}=V\_1S\_1^{-1}U\_1^T$ eqn. (2)

- Example 1.

Theorem 8.7.1 *If A is an $m \times n$ matrix with full column rank, then*

$$A^{+} = (A^{T}A)^{-1}A^{T} \tag{3}$$

- Proof: A=U_1S_1V_1^T.
  - A^{T}A=(V_1S_1^{T}U_1^T)(U_1S_1V_1^T) =V_1S_1^{2}V_1^T.
  - A full rank -> A^{T}A invertible. V nxn-matrix.
  - (A^{T}A)$^{-1}$= V_1S_1^{-2}V_1^T.
  - (A^{T}A)$^{-1}$A^T= V_1S_1^{-2}V_1^T(V_1S_1^{T}U_1^T)
  - =V_1S_1^{-1}U_1^T = A$^+$

Properties of the pseudo-inverses.

Theorem 8.7.2 *If A^+ is the pseudoinverse of an $m \times n$ matrix A, then:*

(*a*) $AA^+A = A$

(*b*) $A^+AA^+ = A^+$

(*c*) $(AA^+)^T = AA^+$

(*d*) $(A^+A)^T = A^+A$

(*e*) $(A^T)^+ = (A^+)^T$

(*f*) $A^{++} = A$

- Proof: computations using (2) and
  - $V\_1^T V\_1 = I$ (kxk-matrix)
  - $U^T U = I$ (kxk-matrix.)

Theorem 8.7.3 *If $A^+ = V_1\Sigma_1^{-1}U_1^T$ is the pseudoinverse of an $m \times n$ matrix A of rank k, and if the column vectors of U_1 and V_1 are $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, respectively, then:*

(a) $A^+\mathbf{y}$ *is in* row(A) *for every vector* $\mathbf{y}$ *in* R^m.

(b) $A^+\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{v}_i \qquad (i = 1, 2, \ldots, k)$

(c) $A^+\mathbf{y} = \mathbf{0}$ *for every vector* $\mathbf{y}$ *in* null(A^T).

(d) AA^+ *is the orthogonal projection of* R^m *onto* col(A).

(e) A^+A *is the orthogonal projection of* R^n *onto* row(A).

Proof: (d) AA$^+$= (U_1S_1V_1^T)V_1S_1^{-1}U_1^T
= U_1U_1^T = proj_span{u_1,…,u_k} = proj_col(A)
(Theorem 8.6.5(a).
(e)) A$^+$A= V_1S_1^{-1}U_1^T (U_1S_1V_1^T)=V_1V_1^T
= proj_span{v_1,..,v_k}=proj_row(A) (Theorem 8.6.5 (c))

Pseudo-inverses and the least squares

- If A has full column rank, then A^TA is invertible and Ax=b has the unique least squares solution
- $x=(A^TA)^{-1}A^Tb=A^+b$. (Theorem 7.8.3)
- If A does not have a full rank, by Theorem 7.8.3, there is a unique one in the row space of A. (minimum norm one.)

Theorem 8.7.4 *If A is an $m \times n$ matrix, and $\mathbf{b}$ is any vector in R^m, then*

$$\mathbf{x} = A^{+}\mathbf{b}$$

is the least squares solution of $A\mathbf{x} = \mathbf{b}$ that has minimum norm.

Proof: $x=A^{+}b = V\_1S\_1U\_1^{T}b$
Thus, $(A^{T}A)A^{+}b= V\_1S\_1^{2}V\_1^{T}V\_1S\_1^{-1}U\_1^{T}b$
$= V\_1S\_1^{2}S\_1^{-1}U\_1^{T} b =V\_1S\_1U\_1^{T}b=A^{T}b$.
Thus x satisfies the least squares equation (10) p.395.

By Theorem 7.8.3, if x is in the row space of A, we are done.
Theorem 8.7.3 implies that x is in row(A).

Condition numbers

- If some eigenvalues of A is zero or close to zero, then Ax=b is said to be ill conditioned.
- If the system is ill conditioned, then errors can become large…. $A^T A$ has too many problems.