Def. 1. Mówimy, że macierz symetryczna S jest *dodatnio półokreślona*, gdy dla dowolnego wektora x zachodzi $x^T Sx \geq 0$, a *dodatnio określona*, gdy dla dowolnego niezerowego x zachodzi $x^T Sx > 0$.

Uwaga 1. Wyrażenie $x^T Sx$ nazywane jest *formą kwadratową* o macierzy S. Pojęcia dodatniej określoności i półokreśloności przenosimy z macierzy S na jej formę kwadratową.

Twierdzenie 1. *Macierz symetryczna jest dodatnio określona wtedy i tylko wtedy, gdy ma dodatnie wartości własne.*

Twierdzenie 2. (x, y) *jest iloczynem skalarnym w* $\mathbb{R}^n$ *wtedy i tylko wtedy, gdy* $(x, y) = x^T Sy$ *dla pewnej dodatnio określonej symetrycznej macierzy* S.

Kryterium Sylvestera

Twierdzenie 3 (Sylvestera)**.** *Macierz symetryczna* S *stopnia* n *jest dodatnio określona wtedy i tylko wtedy, gdy wszystkie jej początkowe minory główne* $M_1, M_2, ..., M_n$ *są dodatnie.*

Przykład 1. Sprawdzić, które z macierzy symetrycznych są dodatnio określone:

$$\begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 3 \\ 3 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 4 \\ 3 & 4 & 15 \end{bmatrix}, \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

Fakt 4. *W macierzu* U *rozkładu* $A = LU$ *symetrycznej macierzy dodatnio określonej* A *elementy główne* $[U]_{ii}$ *są dodatnie,* $[U]_{11} = M_1$ *i* $[U]_{ii} = \dfrac{M_i}{M_{i-1}}$ *dla* $i > 1$.

Rozkład Choleskiego $S = A^T A$, gdzie A jest trójkątna górna

1. W rozkładzie $S = LU$ dodatnio określonej macierzy symetrycznej S elementy główne macierzy U są dodatnie.
2. Tworzymy z nich macierz diagonalną D zachowując kolejność.
3. Z symetrii macierzy S wynika $S = LU = LDL^T$.
4. Wyznaczamy taką macierz $\sqrt{D}$, że $\sqrt{D} \cdot \sqrt{D} = D$ (pierwiastkując wyrazy macierzy D) i stąd $S = L\sqrt{D}\sqrt{D}L^T$.
5. Przyjmując $A = \sqrt{D}L^T$ dostajemy rozkład Choleskiego
$$S = A^T A$$
z pierwiastkami elementów głównych macierzy U na głównej przekątnej macierzy A.

Przykład 2. Wyznaczyć rozkłady Choleskiego macierzy $\begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}$ i $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 10 & 7 \\ 1 & 7 & 9 \end{bmatrix}$.

Fakt 5. *Dla dowolnej macierzy A (niekoniecznie kwadratowej):*

1. *macierz A^TA jest dodatnio półokreśloną macierzą symetryczną,*
2. $r(A^TA) = r(AA^T) = rA = rA^T$,
3. *niezerowe wartości własne macierzy AA^T i A^TA są takie same.*

Uwaga 2. Nie zawsze dla macierzy rzędu r rząd ich iloczynu jest równy r tak, jak w przypadku A^TA.

Fakt 6. *Jeśli A jest $(m \times r)$ macierzą rzędu r i B jest $(r \times n)$ macierzą rzędu r, to AB ma również rząd r.*

SVD - Singular Value Decomposition

Dla dowolnej macierzy $A \in \mathbf{M}_{m\times n}$ rzędu r szukamy takich liczb $\sigma_1, \sigma_2, ..., \sigma_r$ spełniających warunek $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$ oraz takich ortonormalnych baz $\{v_1, v_2, ..., v_n\}$ przestrzeni $\mathbb{R}^n$ i $\{u_1, u_2, ..., u_m\}$ przestrzeni $\mathbb{R}^m$, że

$$Av_1 = \sigma_1 u_1, ..., Av_r = \sigma_r u_r, \qquad Av_{r+1} = 0, ..., Av_n = 0.$$

Warunek zapisuje się macierzowo:

$$AV = U\Sigma$$

$$A \begin{bmatrix} v_1 & v_2 & ... & v_n \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & ... & u_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right]$$

Liczby $\sigma_1, \sigma_2, ..., \sigma_r$ nazywamy *wartościami osobliwymi* (*szczególnymi*) (ang. *singular*), $u_1, u_2, ..., u_m$ - *lewostronnymi*, a $v_1, v_2, ..., v_n$ - *prawostronnymi wektorami osobliwymi (szczególnymi).* Macierze V i U są macierzami ortogonalnymi stopnia n i m odpowiednio. Macierz Σ ma ten sam wymiar co A (na ogół nie jest kwadratowa!) i jest uogólnieniem macierzy diagonalnej. Równość $AV = U\Sigma$ jest równoważna równości

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

którą nazywamy SVD *rozkładem według wartości osobliwych (szczególnych).*

Twierdzenie 7. *Dla dowolnej macierzy A istnieje SVD.*

Sposób wyznaczania SVD

1. Za prawostronne wektory osobliwe v_i dla $n \leq r$ przyjmujemy unormowane wektory własne macierzy $A^T A$ uporządkowane od największej wartości własnej λ_1 do najmniejszej niezerowej λ_r.
2. Wyznaczamy wartości osobliwe $\sigma_i = \sqrt{\lambda_i}$ dla $i = 1, 2, ..., r$.
3. Lewostronne wektory u_i są unormowanymi obrazami wektorów v_i w przekształceniu liniowym o macierzy A dla $i = 1, 2, ..., r$. Zachodzi przy tym $u_i = \dfrac{Av_i}{\sigma_i}$.
4. Dla $n > r$ i $m > r$ za pozostałe v_i i u_i bierzemy wektory z ortonormalnych baz przestrzeni zerowej macierzy A i odpowiednio A^T.

Przykład 3. Wyznaczyć SVD macierzy $A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix}$.

Rozkład na sumę macierzy rzędu 1

Wn. 1. *Jeśli $\sigma_1, \sigma_2, ..., \sigma_r$ są wartościami osobliwymi, $u_1, u_2, ..., u_m$ lewostronnymi, a $v_1, v_2, ..., v_n$ prawostronnymi wektorami osobliwymi rozkładu SVD macierzy A to:*

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + ... + \sigma_r u_r v_r^T.$$

Przykład 4. Wyznaczyć rozkład macierzy z przykładu 3 na sumę macierzy rzędu 1.

Zredukowana forma SVD

Jeśli

$$A = \begin{bmatrix} u_1 & u_2 & ... & u_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right] \begin{bmatrix} v_1 & v_2 & ... & v_n \end{bmatrix}^T,$$

to również

$$\begin{bmatrix} u_1 & u_2 & ... & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} v_1 & v_2 & ... & v_r \end{bmatrix}^T = U_r \Sigma_r V_r^T.$$

Ten drugi rozkład macierzy A nazywamy *zredukowaną formą SVD*. Macierze $U_r \in \mathrm{M}_{m\times r}$ i $V_r \in \mathrm{M}_{n\times r}$ nie są ortogonalne, natomiast macierz Σ_r jest macierzą diagonalną.

Uwagi dotyczące SVD

1. Jeśli $A = U\Sigma V^T$, to $A^T = V\Sigma^T U^T$ i można znaleźć SVD macierzy A^T zamiast A i wynik przetransponować. Warto tak zrobić, gdy A ma więcej kolumn niż wierszy.

2. Rozkład SVD istnieje dla dowolnej macierzy A podczas, gdy rozkład $A = P\Lambda P^{-1}$ tylko dla macierzy kwadratowej, która ma bazę wektorów własnych.

3. SVD dodatnio półokreślonej symetrycznej macierzy A pokrywa się z rokładem $A = Q\Lambda Q^T$, gdzie $U = V = Q$ i $\Sigma = \Lambda$.

4. Jeśli macierz symetryczna ma ujemną wartość własną λ_i, w rokładzie $A = Q\Lambda Q^T$, to wystarczy w SVD wziąć $\sigma_i = -\lambda_i$ i jeden z wektorów osobliwych u_i albo v_i zastąpić wektorem przeciwnym.

5. Jeśli $A = xy^T$ jest macierzą rzędu 1, to $u_1 = \dfrac{x}{||x||}$, $v_1 = \dfrac{y}{||y||}$ i $\sigma_1 = ||x|| \cdot ||y||$.

Norma spektralna macierzy

Def. 2. *Normą spektralną* macierzy $A \in \mathbf{M}_{m\times n}$ nazywamy

$$||A|| := \max\{\frac{||Ax||}{||x||} : x \in E^n\}.$$

Lemat 8. *Niech $\lambda_1 > 0$ będzie największą wartością własną dodatnio półokreślonej macierzy symetrycznej S. Wówczas największą wartością wyrażenia $\dfrac{x^T Sx}{x^T x}$ jest λ_1.*

Wn. 2. *Norma spektralna macierzy A jest jest równa jej największej wartości osobliwej σ_1.*

Twierdzenie 9 (Eckarta-Younga)**.** *Niech $A, B \in \mathbf{M}_{m\times n}$,*

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + ... + \sigma_r u_r v_r^T,$$

$$A_k = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + ... + \sigma_k u_k v_k^T$$

dla każdego $k \in \{1, 2, ..., r\}$ i $rB \leq k$. Wówczas

$$||A - B|| \geq ||A - A_k|| = \sigma_{k+1}.$$

Zastosowania do redukcji zbioru danych

Twierdzenie Eckarta-Younga określa błąd przybliżenia dowolnej macierzy (wysokiego) rzędu r najlepszą z możliwych macierzy (niskiego) rzędu k. Macierzą tą jest suma pierwszych k składników, które odpowiadają największym wartościom osobliwym, a błąd wynosi σ_{k+1}. Pakiet danych zawartych w r-wymiarowej przestrzeni rzutujemy na najlepszą (minimalizującą błąd) podprzestrzeń k-wymiarową.

Przykład 5. Typowe czarno-białe zdjęcie w zapisie cyfrowym to macierz 2064×3088. Oryginalny obraz wymaga zapamiętania $2064 \times 3088 = 6373632$ liczb. Jeśli ograniczymy się np. do sześciu pierwszych składowych, to wystarczy zapamiętać sześć wartości osobliwych oraz odpowiadających im lewych i prawych wektorów osobliwych. Daje to w sumie $6 \times (1 + 2064 + 3088) = 30918$ liczb do zapamiętania, czyli mniej niż $0,5\%$ oryginalnego rozmiaru zdjęcia.

Wprowadzenie do PCA

Przykład 6.

1. Na płaszczyźnie dane są punkty $(-4,-1)$, $(-1,-2)$, $(2,0)$, $(3,3)$. Znaleźć taką prostą $y = ax$, dla której suma kwadratów odległości od tych punktów będzie najmniejsza.
2. Dla danych wektorów $\begin{bmatrix} -4 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ znaleźć najbliższą im podprzestrzeń jednowymiarową.
3. Znaleźć najlepsze przybliżenie macierzy $A = \begin{bmatrix} -4 & -1 & 2 & 3 \\ -1 & -2 & 0 & 3 \end{bmatrix}$ macierzą rzędu 1.

Odp. Otrzymana metodą SVD podprzestrzeń $lin \begin{bmatrix} 5 \\ 3 \end{bmatrix}$ generowana przez pierwszy wektor osobliwy u_1 rozkładu SVD macierzy A (czyli prosta $y = 0,6x$) spełnia rządany warunek.

Uwaga 3. Znaleziona metodą SVD prosta nie pokrywa się z prostą regresji znalezioną metodą najmniejszych kwadratów, która w tym samym przykładzie ma równanie $y = 0,5x$. Zastosowanie SVD w tym problemie nazywane bywa *ortogonalną metodą najmniejszych kwadratów* (ang. *perpendicular least squares*).

Uwaga 4. Dane w przykładzie są uśrednione (tzn. średnie każdego wiersza macierzy A są równe 0. Przy tym założeniu symetryczna macierz $S = \frac{1}{n}AA^T$ jest tzw. *macierzą kowariancji*. Jej wyrazy głównej przekątnej $[S]_{11}$ i $[S]_{22}$ są miarami rozrzutu (wariancjami) danych poszczególnych wierszy, a wyraz drugiej przekątnej $[S]_{12} = [S]_{21}$ to tzw. *kowariancja*, która jest miarą zależności liniowej pomiędzy danymi z pierwszego i drugiego wiersza.

- Jeśli każdy wektor a_i danych (kolumnę macierzy A) przedstawimy w postaci kombinacji wektorów bazy ortonormalnej $\{u_1, u_2\}$, $a_i = \alpha_i u_1 + \beta_i u_2$, to
$$\sum_{i=n}^{4} ||a_i||^2 = \sum_{i=n}^{4} \alpha_i^2 + \sum_{i=n}^{4} \beta_i^2.$$

Ponieważ lewa strona jest ustalona (przez macierz danych), więc maksymalizując pierwszą sumę po prawej stronie jednocześnie minimalizujemu drugą.

- Metoda SVD daje maksymalną wariancję danych na prostej i minimalną sumę kwadratów odchyleń od prostej.

- Metoda SVD wykorzystywana jest w *analizie głównych składowych* (ang. *Principal Component Analisis*) (skrót PCA). PCA wykorzystuje metody algebry liniowej, statystyki i analizy matematycznej do redukcji wymiaru przestrzeni danych.