Lec 7: Matrix Decomposition

Ailin Zhang

2022-09-25

Agenda

- Sweep Operator
- Matrix Decomposition (QR)
- Gram-Schmidt Process

Recap: Sweep operator

The sweep operator arises from a close analysis of elimination applied to the equation $\mathbf{X}^\top X\beta = \mathbf{X}^\top \mathbf{Y}$. For a symmetric matrix$A$, $B = Sweep(A, k)$ is obtained as follows:

1. Divide k^{th} row and k^{th} column by a_{kk}:

$$b_{ik} = a_{ik}/a_{kk};\ b_{kj} = a_{kj}/a_{kk}$$

2. Subtract $a_{ik}a_{kj}/a_{kk}$ from the other entries:

$$b_{ij} = a_{ij} - a_{ik}a_{kj}/a_{kk}$$

3. Invert k^{th} diagonal element:

$$b_{kk} = -1/a_{kk}$$

Proposition 1

Proposition

Suppose $V_{p\times m} = U_{p\times m}A_{m\times m}$*, and* $B = Sweep(A, k)$*. Then* $\hat{V} = \hat{U}B$*, where:*

- $\hat{U} = U$ *except that its* k^{th} *column is* v_k*;*
- $\hat{V} = V$ *except that its* k^{th} *coumn is* $-u_k$*.*

Propositon 1 proof:

By definition $v_{jl} = \sum_i u_{ji} a_{il}$ for all pairs j and l. After sweeping on a_{kk},

$$\begin{aligned}\hat{v}_{jk} = -u_{jk} &= -\frac{1}{a_{kk}} \left(v_{jk} - \sum_{i \neq k} u_{ji} a_{ik} \right) \\ &= \hat{u}_{jk} b_{kk} + \sum_{i \neq k} \hat{u}_{ji} b_{ik} = \sum_i \hat{u}_{ji} b_{ik},\end{aligned}$$

and for $l \neq k$,

$$\begin{aligned}\hat{v}_{jl} = v_{jl} &= \sum_{i \neq k} u_{ji} a_{il} + u_{jk} a_{kl} \\ &= \sum_{i \neq k} u_{ji} a_{il} + \left(v_{jk} - \sum_{i \neq k} u_{ji} a_{ik} \right) \frac{a_{kl}}{a_{kk}} \\ &= \sum_{i \neq k} \hat{u}_{ji} b_{il} + \hat{u}_{jk} b_{kl} = \sum_i \hat{u}_{ji} b_{il}.\end{aligned}$$

Thus, $\hat{V} = \hat{U} B$.

Propositions

Proposition (2)

If A is a symmetric invertible matrix and we sweep on each diagonal element of A the result is $B = Sweep(A, 1:n) = -A^{-1}$

Proposition (3)

If $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ and we sweep on each of the diagonal entries of A_{11}, we get $B = \begin{bmatrix} -A_{11}^{-1} & A_{11}^{-1}A_{12} \\ A_{21}A_{11}^{-1} & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}$

Proposition 4

We construct a matrix $\mathbf{Z} = [\mathbf{X}\mathbf{Y}]$, and let

$$A = \mathbf{Z}^\top \mathbf{Z} = \begin{bmatrix} \mathbf{X}^\top \mathbf{X} & \mathbf{X}^\top \mathbf{Y} \\ \mathbf{Y}^\top \mathbf{X} & \mathbf{Y}^\top \mathbf{Y} \end{bmatrix}$$

be the cross-product matrix. Then

$$\mathrm{SWP}[1:p]A = \begin{bmatrix} -(\mathbf{X}^\top \mathbf{X})^{-1} & (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} \\ \mathbf{Y}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} & \mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y} \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{\mathrm{Var}(\hat{\beta})}{\sigma^2} & \hat{\beta} \\ \hat{\beta}^\top & \mathrm{RSS} \end{bmatrix}$$

where $\mathrm{RSS} = \|\mathbf{Y} - \mathbf{X}\hat{\beta}\|_{\ell_2}^2$ is the residual sum of squares.

Question: What is the time complexity? $\mathcal{O}(p^3)$

Inverse sweep operator

We define the inverse sweep operator as follows: Inv-Sweep(k)

$$
\begin{cases}
a_{ik} & = -b_{ik}/b_{kk} \\
a_{kj} & = -b_{kj}/b_{kk} \\
a_{kk} & = -1/b_{kk} \\
a_{ij} & = b_{ij} - b_{ik}b_{kj}/b_{kk}
\end{cases}
$$

Sweep and Stepwise Regression

Suppose we "start with" the regression $y = X_1\beta_1 + \epsilon$ and we want to expand this to $y = X_1\beta_1 + X_2\beta_2 + \epsilon$, for a new regressor X_2. If we sweep

$$A = \begin{bmatrix} X_1^T \\ X_2^T \\ Y^T \end{bmatrix} \begin{bmatrix} X_1 & X_2 & Y = \end{bmatrix} \begin{bmatrix} X_1^T X_1 & X_1^T X_2 & X_1^T Y \\ X_2^T X_1 & X_2^T X_2 & X_2^T Y \\ Y^T X_1 & Y^T X_2 & Y^T Y \end{bmatrix}$$

on the diagonal entries corresponding to X_1 we get

$$\begin{bmatrix} -\left(X_1^T X_1\right)^{-1} & \hat{\beta}_{X_2|X_1} & \hat{\beta}_{Y|X_1} \\ \hat{\beta}_{X_2|X_1}^T & \left\|X_2 - \hat{X}_2\right\|^2 & \left(X_2 - \hat{X}_2\right)^T (Y - \hat{Y}) \\ \hat{\beta}_{Y|X_1}^T & \left(Y - \hat{Y}\right)^T \left(X_2 - \hat{X}_2\right) & \left\|Y - \hat{Y}\right\|^2 \end{bmatrix}$$

Sweep and Stepwise Regression

If we now further sweep on the diagonal entries corresponding to X_2 we have "added" X_2 to the regression:

$$\begin{bmatrix} -\left[\begin{pmatrix} X_1^T \\ X_2^T \end{pmatrix}\begin{pmatrix} X_1 & X_2 \end{pmatrix}\right]^{-1} & \hat{\beta}_{Y|X_1X_2} \\ \hat{\beta}_{Y|X_1X_2}^T & \|Y - \hat{Y}\|^2 \end{bmatrix}$$

where now $\hat{\beta}_{Y|X_1X_2}$ and $\hat{Y}$ are for regression on X_1X_2.

If we want to address $Y = X_2\beta_2 + \epsilon$, we could apply the inverse sweep operator on X1 to remove X1 from the regression

Other Application of sweep operator

Perform sweep operator on multivariate normal distribution $x \sim N(\mu, \Omega)$

$$\begin{bmatrix} \Omega & x-\mu \\ x^t-\mu^t & 0 \end{bmatrix}$$

If we sweep on Ω, we get the quadratic form $-(x-\mu)^t\Omega^{-1}(x-\mu)$ in the lower right block of the swept matrix.

If we partition X as $(Y^t, Z^t)^t$ and sweep on the upper left block of

$$\begin{pmatrix} \Omega_Y & \Omega_{YZ} & \mu_Y - y \\ \Omega_{ZY} & \Omega_Z & \mu_Z \\ \mu_Y^t - y^t & \mu_Z^t & 0 \end{pmatrix},$$

then the conditional mean $\mathrm{E}(Z \mid Y = y) = \mu_Z + \Omega_{ZY}\Omega_Y^{-1}(y - \mu_Y)$ and conditional variance $\mathrm{Var}(Z \mid Y = y) = \Omega_Z - \Omega_{ZY}\Omega_Y^{-1}\Omega_{YZ}$ are immediately available.

Why Sweep Operator?

1. Efficient for computing the central statistics used in multiple regression
   - Start from a square correlation or covariance matrix.
   - Compute multiple correlation, residual variance, regression slopes, and standard errors of slopes, plus some other
2. An efficient way to compute a whole series of regressions, as in stepwise regression.
3. (Drawback) If it is applied many times to the same matrix, rounding error can accumulate.

Matrix Decomposition

QR decomposition decomposes a matrix $\mathbf{X}$ into a product $\mathbf{X} = QR$ where Q is an orthogonal matrix and R is an upper triangular matrix. It does not require the computation of the cross-product matrix such as $\mathbf{X}^\top \mathbf{X}$.

Let $Q = (q_1, q_2, \ldots, q_n)$ be an orthogonal matrix, $Q^\top Q = QQ^\top = I$, then Q forms an orthogonal basis:

(1) For each vector q_i, $\|q_i\| = 1$.

(2) For any two different vectors q_i and q_j, $\langle q_i, q_j \rangle = 0$, i.e., $q_i \perp q_j$.

For any vector v, we have

(1) Analysis: $u_i = \langle v, q_i \rangle = q_i^\top v$ is the coordinate of v on the axis q_i, for $i = 1, ..., n$, i.e., $u = Q^\top v$.

(2) Synthesis: $v = \sum_{i=1}^{n} q_i u_i = Qu$.

From (1) and (2), Q and $Q^\top$ are inverse of each other.

Gram-Schmidt Process

Sketch it!

$$
\begin{aligned}
u_1 &= a_1, & q_1 &= \frac{u_1}{\|u_1\|_2}, \quad \tilde{r}_{11} = \frac{1}{\|u_1\|_2} \\
u_2 &= a_2 - (q_1^T a_2) q_1, & q_2 &= \frac{u_2}{\|u_2\|_2}, \quad \tilde{r}_{12} = -\frac{q_1^T a_2}{\|u_2\|_2}, \quad \tilde{r}_{22} = \frac{1}{\|u_2\|_2} \\
&\vdots \\
u_n &= a_n - \sum_{j=1}^{n-1} (q_j^T a_n) q_j, & q_n &= \frac{u_n}{\|u_n\|_2}, \quad \tilde{r}_{jn} = -\frac{q_j^T a_n}{\|u_n\|_2}, \quad \tilde{r}_{nn} = \frac{1}{\|u_n\|_2}
\end{aligned}
$$

$$
A \begin{bmatrix} \tilde{r}_{11} & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}
\begin{bmatrix} 1 & \tilde{r}_{12} & & \\ & \tilde{r}_{22} & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}
\cdots
\begin{bmatrix} 1 & & & \tilde{r}_{1n} \\ & 1 & & \tilde{r}_{2n} \\ & & \ddots & \vdots \\ & & & \tilde{r}_{n1} \end{bmatrix} = Q
$$

Projection Matrix

In Gram-Schmidt, we can treat each vector q_j as a product of projection of A_j,where the projection matrix is $P_j = I - Q_{j-1}Q_{j-1}^T$ and it satisfies the following:

1. $P^2 = P$
2. $I - P$ is also a projection matrix

$$Q_j = \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_j \\ | & | & & | \end{bmatrix}$$

$$q_1 = \frac{P_1 a_1}{||P_1 a_1||} \qquad P_1 = I$$

$$q_2 = \frac{P_2 a_2}{||P_2 a_2||}$$

$$\vdots$$