ST509 Computational Statistics

Lecture 3: Regression

Seung Jun Shin

Department of Statistics
Korea University

E-mail: sjshin@korea.ac.kr

Introduction I

- Linear Regression assumes

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}, \qquad \mathrm{cov}(\mathbf{e}) = \sigma^2 \mathbf{I}$$

  where $\mathbf{X} \in \mathbb{R}^{n \times p}$ is called the *design matrix*.

- Our goal is to minimize

$$S(\mathbf{b}) = (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b})$$

  w.r.t. $\mathbf{b}$.

- Under a suitable conditions, the solution $\hat{\mathbf{b}}$ must the solution of the normal equation:

$$\mathbf{X}^T \mathbf{X}\mathbf{b} = \mathbf{X}^T \mathbf{y}$$

  which can be rewritten as

$$\mathbf{X}^T (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \mathbf{X}^T \hat{\mathbf{e}} = \mathbf{0}$$

  where $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\mathbf{b}}$ denotes the residual vector.

Introduction II

- The fitted value $\hat{\mathbf{y}}$ can be written as

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}} = \mathbf{P}_{\mathbf{X}}\mathbf{y}$$

  where $\mathbf{P}_{\mathbf{X}}$ is called the projection matrix since it projects $\mathbf{y}$ onto the column space of $\mathbf{X}$.

- The residual vector $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{P}_{\mathbf{X}})\mathbf{y}$.
- Thus $\hat{\mathbf{y}}$ and $\hat{\mathbf{e}}$ are orthogonal (i.e., $\hat{\mathbf{y}}^T\hat{\mathbf{e}} = \mathbf{0}$.)
- The solution $\hat{\mathbf{b}}$ and the error sum of squqres

$$\text{SSE} = S(\hat{\mathbf{b}}) = (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})^T(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) = \hat{\mathbf{e}}^T\hat{\mathbf{e}},$$

  are the most important quantities.

- At the next level, $(\mathbf{X}^T\mathbf{X})^{-1}$ and the (unscaled) covariance matrix of $\hat{\mathbf{b}}$

$$\text{cov}(\hat{\mathbf{b}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$$

  are also important.

Solving the Normal Equations I

- Assume $\mathbf{X}$ is of full-rank.
- One possible approach is to use CD.
  1. Compute $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}^T\mathbf{y}$.
  2. Factor $\mathbf{X}^T\mathbf{X} = \mathbf{L}\mathbf{L}^T$ via CD.
  3. Solve $\mathbf{L}\mathbf{w} = \mathbf{X}^T\mathbf{y}$ for $\mathbf{w}$.
  4. Compute SSE by $SSE = \mathbf{y}^T\mathbf{y} - \mathbf{w}^T\mathbf{w}$
  5. Solve $\mathbf{L}^T\mathbf{b} = \mathbf{w}$ for $\mathbf{b}$ to obtain $\hat{\mathbf{b}}$.
  6. Invert $\mathbf{L}$.
  7. Compute $\mathbf{L}^{-T}\mathbf{L}^{-1} = (\mathbf{X}^T\mathbf{X})^{-1}$.
- Drawbacks
  - Low accuracy (good for $\hat{\mathbf{b}}$ but not for $\hat{\mathbf{e}}$)
  - Impossible to use when $\mathbf{X}^T\mathbf{X}$ is (computationally) singular.

ex. For $\mathbf{X}^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1.01 & 1.01 \end{bmatrix}$, $\mathbf{X}^T\mathbf{X}$ is computationally singular with $d = 4$.

Gram-Schmidt Orthogonalization I

- A method for producing a sequence of orthonormal vectors from a set of linearly independent vectors.
- In the regression problems, the columns of $\mathbf{X} = (\mathbf{X}_{\bullet 1}, \cdots, \mathbf{X}_{\bullet p})$ are the linearly independent vectors.
- The objective is to find the following QR-factorization

$$\mathbf{X}_{n\times p} = \mathbf{Q}_{n\times p}\mathbf{R}_{p\times p}, \qquad \text{where } \mathbf{Q}^T\mathbf{Q} = \mathbf{I}.$$

  so that
  - $\mathbf{Q}$ has orthonormal columns $\mathbf{Q}_{\bullet 1}, \cdots, \mathbf{Q}_{\bullet p}$;
  - $\mathbf{R}$ is upper-triangular.

Gram-Schmidt Orthogonalization II

- At ith iteration of GS, the ith column $\mathbf{X}_{\bullet i}$ is regression on $(\mathbf{Q}_{\bullet j}, \cdots, \mathbf{Q}_{\bullet i-1})$.
- Since explanatory variables are orthonormal, the regression coefficients are

$$R_{ji} = \mathbf{Q}_{\bullet j}^T \mathbf{X}_i$$

- Update $Q_{\bullet i}$ as

$$Q_{\bullet i} = \frac{\mathbf{X}_{\bullet i} - \sum_{j=1}^{i-1} R_{ji}\mathbf{Q}_{\bullet j}}{R_{ii}}$$

  where

$$R_{ii} = \left\| \mathbf{X}_{\bullet i} - \sum_{j=1}^{i-1} R_{ji}\mathbf{Q}_{\bullet j} \right\|.$$

Gram-Schmidt Orthogonalization III

- Notice $\mathbf{Q}_{\bullet i}$ is independent all $\mathbf{Q}_{\bullet j}, j = 1, \cdots, i-1$
- R_{11} and $\mathbf{Q}_{\bullet 1}$ and can be initialized as

$$R_{11} = \|\mathbf{X}_{\bullet 1}\|, \qquad \text{and} \qquad \mathbf{Q}_{\bullet 1} = \mathbf{X}_{\bullet 1}/R_{11}$$

  respectively.
- Notice that

$$\mathbf{X}_i = \sum_{j=1}^{i} R_{ji}\mathbf{Q}_{\bullet j}$$

  and this yields

$$\mathbf{X} = \mathbf{QR}.$$

Gram-Schmidt Orthogonalization IV

1. Initialize $R_{11} = \|\mathbf{X}_{\bullet 1}\|$ and $\mathbf{Q}_{\bullet 1} = \mathbf{X}_{\bullet 1}/R_{11}$
2. For $i = 2, \cdots, p$

   2.1 Update
   $$R_{ji} = \mathbf{Q}_{\bullet j}^T \mathbf{X}_i, \qquad j = 1, \cdots i-1.$$
   2.2 Update
   $$R_{ii} = \left\| \mathbf{X}_{\bullet i} - \sum_{j=1}^{i-1} R_{ji} \mathbf{Q}_{\bullet j} \right\|.$$
   2.3 Update
   $$\mathbf{Q}_{\bullet i} = \frac{\mathbf{X}_{\bullet i} - \sum_{j=1}^{i-1} R_{ji} \mathbf{Q}_{\bullet j}}{R_{ii}}.$$

Algorithm 1: Regular Gram-Schmidt (RGS) Orthogonalization

Gram-Schmidt Orthogonalization V

ex Apply RGS for

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \\ 1 & 6 & 36 \end{bmatrix}$$

1st

$$R_{11} = \|\mathbf{X}_{\bullet 1}\| = \sqrt{6} = 2.449$$

and

$$\mathbf{Q}_{\bullet 1} = \mathbf{X}_{\bullet 1}/R_{11} = \begin{bmatrix} .4082 \\ .4082 \\ .4082 \\ .4082 \\ .4082 \\ .4082 \end{bmatrix}.$$

Gram-Schmidt Orthogonalization VI

2nd

$$R_{12} = \mathbf{Q}_{\bullet 1}^T \mathbf{X}_{\bullet 2} = \begin{bmatrix} .4082 \\ .4082 \\ .4082 \\ .4082 \\ .4082 \\ .4082 \end{bmatrix}^T \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix} = 8.573,$$

$$\mathbf{X}_{\bullet 2} - R_{12}\mathbf{Q}_{\bullet 1} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix} - \begin{bmatrix} 3.499 \\ 3.499 \\ 3.499 \\ 3.499 \\ 3.499 \\ 3.499 \end{bmatrix} = \begin{bmatrix} -2.499 \\ -1.499 \\ -0.499 \\ 0.5010 \\ 1.510 \\ 2.501 \end{bmatrix}, \quad R_{22} = \|\mathbf{X}_{\bullet 2} - R_{12}\mathbf{Q}_{\bullet 1}\| = 4.183,$$

and

$$\mathbf{Q}_{\bullet 2} = (\mathbf{X}_{\bullet 2} - R_{12}\mathbf{Q}_{\bullet 1})/R_{22} = \begin{bmatrix} -.2499 \\ -.1499 \\ -.4990 \\ .5010 \\ 1.501 \\ 2.501 \end{bmatrix} /4.183 = \begin{bmatrix} -.5974 \\ -.3584 \\ -.1193 \\ .1198 \\ .3588 \\ .5979 \end{bmatrix}.$$

Gram-Schmidt Orthogonalization VII

3rd

$$R_{13} = \mathbf{Q}_{\bullet 1}^T \mathbf{X}_{\bullet 3} = \begin{bmatrix} .4082 \\ .4082 \\ .4082 \\ .4082 \\ .4082 \\ .4082 \end{bmatrix}^T \begin{bmatrix} 1 \\ 4 \\ 9 \\ 16 \\ 25 \\ 36 \end{bmatrix} = 37.15$$

and

$$R_{23} = \mathbf{Q}_{\bullet 2}^T \mathbf{X}_{\bullet 3} = \begin{bmatrix} -.5974 \\ -.3584 \\ -.1193 \\ .1198 \\ .3588 \\ .5979 \end{bmatrix}^T \begin{bmatrix} 1 \\ 4 \\ 9 \\ 16 \\ 25 \\ 36 \end{bmatrix} = 29.31.$$

$$\mathbf{X}_{\bullet 3} - R_{13}\mathbf{Q}_{\bullet 1} - R_{23}\mathbf{Q}_{\bullet 2} = \begin{bmatrix} 1 - 15.16 + 17.51 \\ 4 - 15.16 + 10.51 \\ 9 - 15.16 + 3.497 \\ 16 - 15.16 - 3.511 \\ 25 - 15.16 - 10.52 \\ 36 - 15.16 - 17.52 \end{bmatrix} = \begin{bmatrix} 3.350 \\ -.6600 \\ -2.663 \\ -2.671 \\ -.6800 \\ 3.320 \end{bmatrix},$$

Gram-Schmidt Orthogonalization VIII

$$R_{33} = 6.113,$$

and

$$\mathbf{Q}_{\bullet 3} = \begin{bmatrix} 3.350 \\ -.6600 \\ -2.663 \\ -2.671 \\ -.6800 \\ 3.320 \end{bmatrix} /6.113 = \begin{bmatrix} 0.5480 \\ -0.1080 \\ -0.4356 \\ -0.4369 \\ -0.1112 \\ 0.5431. \end{bmatrix}$$

Gram-Schmidt Orthogonalization IX

- Connection to CD: **R** is merely the transpose of **L** (with possibly different sign).
- Normal equation becomes

$$\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y} \qquad \Leftrightarrow \qquad \mathbf{L}\mathbf{L}^T\mathbf{b} = \mathbf{L}\mathbf{Q}^T\mathbf{y}$$

  where $\mathbf{L} = \mathbf{R}^T$.
- This yields

$$\mathbf{R}\mathbf{b} = \mathbf{Q}^T\mathbf{y}$$

  which is a simple triangular system to be solved.
- But this is still poorly conditioned problem, and a better orthonormalization method is still required.
- There is a modified GS, but it is still not good enough.

Householder Transformations for Least Squares I

- Householder transformation (HT) is a simple but powerful tool.
- For an arbitrary vector $\mathbf{u}$, the matrix

$$\mathbf{U} = \mathbf{I} - d\mathbf{u}\mathbf{u}^T$$

  is symmetric and orthogonal when $d = 2/\mathbf{u}^T\mathbf{u}$.
- For any vector $\mathbf{x}$, we can find $\mathbf{u}$ and s such that

$$\mathbf{U}\mathbf{x} = s\mathbf{e}_1$$

- We call $\mathbf{U}\mathbf{x}$ is a HT of $\mathbf{x}$.
- One set solutions of $\mathbf{u}$ and s is

$$\mathbf{u} = \mathbf{x} + s\mathbf{e}_1, \quad \text{and} \quad s^2 = \mathbf{x}^T\mathbf{x}.$$

ex. Apply HT to $\mathbf{x}^T = (-1, 2, -2, 4)^T$.

Householder Transformations for Least Squares II

- LS problem is unaffected by rotation. That is, for an orthogonal matrix $\mathbf{U}$,
$$S(\mathbf{b}) = \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 = \|\mathbf{U}\mathbf{y} - \mathbf{U}\mathbf{X}\mathbf{b}\|^2.$$
- If $\mathbf{UX}$ has a simple form, then the problem become much easier.
- This motivates the use of HT for the regression problem.
- Apply HT to find an orthogonal matrix $\mathbf{U}$ such that
$$\mathbf{UX} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \begin{array}{c} p \\ n-p \end{array}$$
where $\mathbf{R}$ is upper triangular.

Householder Transformations for Least Squares III

ex Apply the HT to $\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$. That is, find an orthogonal matrix $\mathbf{U}$ such that $\mathbf{UX}$ is upper triangluar.

1. Apply HT to the first column of $\mathbf{X}$: $\mathbf{c}^{(1)} = (1, 1, 1, 1)$:
   1.1 $s_1 = 2$ and $\mathbf{u}^{(1)} = \mathbf{c}^{(1)} + s_1\mathbf{e}_1 = (3, 1, 1, 1)^T$, which yields $\mathbf{U}^{(1)} = \mathbf{I}_4 - d_1\mathbf{u}^{(1)}\mathbf{u}^{(1)T} = \mathbf{U}_1$.
   1.2 Thus

$$\mathbf{U}_1\mathbf{X} = \begin{bmatrix} -1/2 & -1/2 & -1/2 & 1/2 \\ -1/2 & 5/6 & -1/6 & -1/6 \\ -1/2 & -1/6 & 5/6 & -1/6 \\ -1/2 & -1/6 & -1/6 & 5/6 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} -2 & -5 \\ 0 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix}$$

Householder Transformations for Least Squares IV

2. Apply HT to the second column (except the 1st elem.) of $\mathbf{U}_1\mathbf{X}$: $\mathbf{c}^{(2)} = (0, 1, 2)$:
   2.1 $s_2 = \sqrt{5}$ and $\mathbf{u}^{(2)} = \mathbf{c}^{(2)} + s_2\mathbf{e}_1 = (\sqrt{5}, 1, 2)^T$, which yields $\mathbf{U}^{(2)} = \mathbf{I}_3 - d_2\mathbf{u}^{(2)}\mathbf{u}^{(2)T}$ and $\mathbf{U}_2 = \mathrm{Diag}(\mathbf{I}, \mathbf{U}^{(2)})$
   2.2 Thus

$$\mathbf{U}_2\mathbf{U}_1\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1/\sqrt{5} & -2/\sqrt{5} \\ 0 & -1/\sqrt{5} & 4/5 & -2/\sqrt{5} \\ 0 & -2/\sqrt{5} & -2/\sqrt{5} & 1/5 \end{bmatrix} \begin{bmatrix} -2 & -5 \\ 0 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -2 & -5 \\ 0 & -\sqrt{5} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

3 Finally, we have

$$\mathbf{U}\mathbf{X} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}$$

where $\mathbf{U} = \mathbf{U}_2\mathbf{U}_1$ and $\mathbf{R} = \begin{bmatrix} -2 & -5 \\ 0 & -\sqrt{5} \end{bmatrix}$.

Householder Transformations for Least Squares V

1. Initialize $\mathbf{X}^{(1)} = \mathbf{X}$ and $\mathbf{U} = \mathbf{I}_p$.
2. For $i = 1, \cdots, p-1$;

   2.1 $\mathbf{c}^{(i)}$ = vector of the last $n-i+1$ elements of the ith column of $\mathbf{X}^{(i)}$.

   2.2 Apply the householder transformation to $\mathbf{c}^{(i)}$:

   - $s_i^2 = \mathbf{c}^{(i)T}\mathbf{c}^{(i)}$
   - $\mathbf{u}^{(i)} = \mathbf{c}^{(i)} + s_i \mathbf{e}_1$
   - $d_i = 2/(\mathbf{u}^{(i)T}\mathbf{u}^{(i)}) = (s_i^2 + sc_1^{(i)})^{-1}$
   - $\mathbf{U}^{(i)} = \mathbf{I}_p - d_i \mathbf{u}^{(i)}\mathbf{u}^{(i)T}$

   to update

   $$\mathbf{U}_i = \begin{bmatrix} \mathbf{I}_{i-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}^{(i)} \end{bmatrix}$$

   2.3 Update $\mathbf{U} = \mathbf{U}_i\mathbf{U}$ and $\mathbf{X}^{(i+1)} = \mathbf{U}_i\mathbf{X}^{(i)}$.

3. Return

$$\mathbf{U}, \mathbf{U}\mathbf{y}, \text{ and } \mathbf{X}^{(p)} = \mathbf{U}\mathbf{X}.$$

Algorithm 2: Householder Transformation For Least Squares

Householder Transformations for Least Squares VI

- Partion $\mathbf{Uy}$ into its first p and last $n-p$ elements:

$$\mathbf{Uy} = \begin{bmatrix} \mathbf{z}_{(1)} \\ \mathbf{z}_{(2)} \end{bmatrix}$$

- LS problem is now

$$\|\mathbf{y} - \mathbf{Xb}\|^2 = \|\mathbf{U}(\mathbf{y} - \mathbf{X})\|^2 = \left\| \begin{bmatrix} \mathbf{z}_{(1)} \\ \mathbf{z}_{(2)} \end{bmatrix} - \begin{bmatrix} \mathbf{Rb} \\ \mathbf{0} \end{bmatrix} \right\|^2 = \underbrace{\|\mathbf{z}_{(1)} - \mathbf{Rb}\|^2}_{\text{SSR}} + \underbrace{\|\mathbf{z}_{(2)}\|^2}_{\text{SSE}}$$

- We can get $\mathbf{b}$ by solving

$$\mathbf{Rb} = \mathbf{z}_{(1)}.$$

- Notice that

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} \mathbf{R}^T & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} = \mathbf{R}^T\mathbf{R}$$

and thus

$$(\mathbf{X}^T\mathbf{X})^{-1} = \mathbf{R}^{-1}\mathbf{R}^{-T}$$

Summary: HT for LS

1. Create and multiply $\mathbf{U} = \mathbf{U}_p \cdots \mathbf{U}_1$ on $\mathbf{X}$ and $\mathbf{y}$. (Algorithm 2)
2. Solve $\mathbf{R}\mathbf{b} = \mathbf{z}_{(1)}$ to obtain $\hat{\mathbf{b}}$.
3. SSR $= \|\mathbf{z}_{(1)}\|^2$ and SSE $= \|\mathbf{z}_{(1)}\|^2$.
4. Invert $\mathbf{R}$ and multiply to get $(\mathbf{X}^T\mathbf{X})^{-1} = \mathbf{R}^{-1}\mathbf{R}^{-T}$.
5. Compute predicted values $\hat{\mathbf{y}}$ and residuals $\hat{\mathbf{e}}$

$$\hat{\mathbf{y}} = \mathbf{Q}\begin{bmatrix} \mathbf{z}_{(1)} \\ \mathbf{0}_{n-p} \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{e}} = \mathbf{Q}\begin{bmatrix} \mathbf{0}_p \\ \mathbf{z}_{(2)} \end{bmatrix}.$$

where $\mathbf{Q} = \mathbf{U}^T = \mathbf{U}^{-1}$. That is, we have

$$\mathbf{X} = \mathbf{Q}\begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \qquad \text{(QR decomposition of } \mathbf{X}\text{)}$$

Givens Transformations for Least Squares I

- Consider the 2×2 orthogonal matrix

$$\mathbf{U} = \begin{bmatrix} \sin t & \cos t \\ -\cos t & \sin t \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

  where $a^2 + b^2 = 1$.

- For $\mathbf{x} = (x_1, x_2)^T$, let

$$a = \frac{x_1}{s} \quad \text{and} \quad b = \frac{x_2}{s}$$

  with $s = \sqrt{x_1^2 + x_2^2}$.

- We then have

$$\mathbf{Ux} = \frac{1}{s} \begin{bmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} s \\ 0 \end{bmatrix}.$$

- This is the Givens transformation of $\mathbf{x}$.

Givens Transformations for Least Squares II

- For $\mathbf{x}$, we have for any pairs of $i = 1, \cdots, p$ and $j = 1, \cdots, n$ with $i < j$,

$$\mathbf{U}_{ij}\mathbf{x} = \begin{bmatrix} \mathbf{I}_{i-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & x_i/s_{ij} & \mathbf{0} & x_j/s_{ij} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{j-i-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -x_j/s_{ij} & \mathbf{0} & x_i/s_{ij} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{n-j} \end{bmatrix} \begin{bmatrix} \cdots \\ x_i \\ \cdots \\ x_j \\ \cdots \end{bmatrix} = \begin{bmatrix} \cdots \\ s_{ij} \\ \cdots \\ 0 \\ \cdots \end{bmatrix}$$

  where $s_{ij} = \sqrt{x_i^2 + x_j^2}$.
- Finally we have

$$\mathbf{U}_{pn} \cdots \mathbf{U}_{12}\mathbf{X} = \mathbf{U}\mathbf{X} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}$$

  with $\mathbf{R}$ being upper triangular.

Givens Transformations for Least Squares III

ex Apply the GT to $\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$. That is, find an orthogonal matrix $\mathbf{U}$ such that $\mathbf{UX}$ is upper triangluar.

(1,2) we have $s_{12} = \sqrt{2}$ and hence

$$\mathbf{U}_{12}\mathbf{X} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 3/\sqrt{2} \\ 0 & -1/\sqrt{2} \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$$

(1,3) we have $s_{13} = \sqrt{3}$ and hence

$$\mathbf{U}_{13}\mathbf{U}_{12}\mathbf{X} = \begin{bmatrix} \sqrt{2/3} & 0 & 1/\sqrt{3} & 0 \\ 0 & 1 & 0 & 0 \\ -1/\sqrt{3} & 0 & \sqrt{2/3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 3/\sqrt{2} \\ 0 & -1/\sqrt{2} \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & 2\sqrt{3} \\ 0 & -1/\sqrt{2} \\ 0 & \sqrt{3/2} \\ 1 & 4 \end{bmatrix}$$

Givens Transformations for Least Squares IV

(2,3) we have $s_{23} = \sqrt{2}$ and hence

$$\mathbf{U}_{23}\mathbf{U}_{13}\mathbf{U}_{12}\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1/2 & \sqrt{3}/2 & 0 \\ 0 & -\sqrt{3}/2 & -1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 2\sqrt{3} \\ 0 & -1/\sqrt{2} \\ 0 & \sqrt{3/2} \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & 2\sqrt{3} \\ 0 & \sqrt{2} \\ 0 & 0 \\ 1 & 4 \end{bmatrix}$$

(1,4) we have $s_{14} = 2$ and hence

$$\mathbf{U}_{14}\cdots\mathbf{U}_{12}\mathbf{X} = \begin{bmatrix} \sqrt{3}/2 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1/2 & 0 & 0 & \sqrt{3}/2 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 2\sqrt{3} \\ 0 & \sqrt{2} \\ 0 & 0 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 0 & \sqrt{2} \\ 0 & 0 \\ 0 & \sqrt{3} \end{bmatrix}$$

(2,4) we have $s_{24} = \sqrt{5}$ and hence

$$\mathbf{U}_{24}\cdots\mathbf{U}_{12}\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{2/5} & 0 & \sqrt{3/5} \\ 0 & 0 & 1 & 0 \\ 0 & -\sqrt{3/5} & 0 & \sqrt{2/5} \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 0 & \sqrt{2} \\ 0 & 0 \\ 0 & \sqrt{3} \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 0 & \sqrt{5} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Givens Transformations for Least Squares V

Row For $j = 2, \cdots, n$

Column For $i = 1, \min(j-1, p)$

1. Compute $\mathbf{U}_{ij}$: let $s_{ij} = \sqrt{x_{ii}^2 + x_{ji}^2}$,

$$\mathbf{U}_{ij} = \begin{bmatrix} \mathbf{I}_{i-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & x_{ii}/s_{ij} & \mathbf{0} & x_{ji}/s_{ij} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{j-i-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -x_{ji}/s_{ij} & \mathbf{0} & x_{ii}/s_{ij} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{n-j} \end{bmatrix}$$

2. Update $\mathbf{X} = \mathbf{U}_{ij}\mathbf{X}$, and $\mathbf{U} = \mathbf{U}_{ij}\mathbf{U}$.
3. Then we have

$$\mathbf{U}\mathbf{X} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad \mathbf{U}\mathbf{y} = \mathbf{z}$$

with $\mathbf{R}$ being upper triangular. Alternatively, letting $\mathbf{Q} = \mathbf{U}^T = \mathbf{U}^{-1}$ we can rewrite

$$\mathbf{X} = \mathbf{Q} \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}.$$

Algorithm 3: Givens Transformation for Least Squares.

- After getting $\mathbf{R}$ and $\mathbf{U}$, remaining steps are identical to those of HT.

Reference

- Monahan, J. F. (2011). Numerical Methods of Statistics, Cambridge University Press. Chapter 4.