Example 6.4 *(Orthonormal design matrix)*
Consider an orthonormal design matrix $\mathbf{X}$, i.e. $\mathbf{X}^\top\mathbf{X} = \mathbf{I}_{pp} = (\mathbf{X}^\top\mathbf{X})^{-1}$. The lasso estimator then is:

$$\hat{\beta}_j(\lambda_1) \quad = \quad \text{sign}(\hat{\beta}_j)(|\hat{\beta}_j| - \tfrac{1}{2}\lambda_1)_+,$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} = \mathbf{X}^\top\mathbf{Y}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$ and $\hat{\beta}_j$ its j-th element and $f(x) = (x)_+ = max\{x, 0\}$. This expression for the lasso regression estimator can be obtained as follows. Rewrite the lasso regression loss criterion:

$$\begin{aligned}
\min_{\boldsymbol{\beta}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1 &= \min_{\boldsymbol{\beta}} \mathbf{Y}^\top\mathbf{Y} - \mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{Y} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} + \lambda_1\sum\nolimits_{j=1}^{p}|\beta_j| \\
&\propto \min_{\boldsymbol{\beta}} -\hat{\boldsymbol{\beta}}^\top\boldsymbol{\beta} - \boldsymbol{\beta}^\top\hat{\boldsymbol{\beta}} + \boldsymbol{\beta}^\top\boldsymbol{\beta} + \lambda_1\sum\nolimits_{j=1}^{p}|\beta_j| \\
&= \min_{\beta_1,\ldots,\beta_p}\sum\nolimits_{j=1}^{p}\big(-2\hat{\beta}_j^{\text{OLS}}\beta_j + \beta_j^2 + \lambda_1|\beta_j|\big) \\
&= \sum\nolimits_{j=1}^{p}(\min_{\beta_j} -2\hat{\beta}_j\beta_j + \beta_j^2 + \lambda_1|\beta_j|).
\end{aligned}$$

The minimization problem can thus be solved per regression coefficient. This gives:

$$\min_{\beta_j} -2\hat{\beta}_j\beta_j + \beta_j^2 + \lambda_1|\beta_j| \quad = \quad \begin{cases} \min_{\beta_j} -2\hat{\beta}_j\beta_j + \beta_j^2 + \lambda_1\beta_j & \text{if} \quad \beta_j > 0, \\ \min_{\beta_j} -2\hat{\beta}_j\beta_j + \beta_j^2 - \lambda_1\beta_j & \text{if} \quad \beta_j < 0. \end{cases}$$

The minimization within the sum over the covariates is with respect to each element of the regression parameter separately. Optimization with respect to the j-th one gives:

$$\hat{\beta}_j(\lambda_1) \quad = \quad \begin{cases} \hat{\beta}_j - \frac{1}{2}\lambda_1 & \text{if} \quad \hat{\beta}_j(\lambda_1) > 0 \\ \hat{\beta}_j + \frac{1}{2}\lambda_1 & \text{if} \quad \hat{\beta}_j(\lambda_1) < 0 \\ 0 & \text{otherwise} \end{cases}$$

Example 6.5 *(Orthogonal design matrix)*
The analytic solution of the lasso regression estimator for experiments with an orthonormal design matrix applies to those with an orthogonal design matrix. This is illustrated by a numerical example. Use the lasso estimator with $\lambda_1 = 10$ to fit the linear regression model to the response data and the design matrix:

$$\begin{aligned} \mathbf{Y}^\top &= \begin{pmatrix} -4.9 & -0.8 & -8.9 & 4.9 & 1.1 & -2.0 \end{pmatrix}, \\ \mathbf{X}^\top &= \begin{pmatrix} 1 & -1 & 3 & -3 & 1 & 1 \\ -3 & -3 & -1 & 0 & 3 & 0 \end{pmatrix}. \end{aligned}$$

Note that the design matrix is orthogonal, i.e. its columns are orthogonal (but not normalized to one). The orthogonality of $\mathbf{X}$ yields a diagonal $\mathbf{X}^\top\mathbf{X}$, and so is its inverse $(\mathbf{X}^\top\mathbf{X})^{-1}$. Here $\text{diag}(\mathbf{X}^\top\mathbf{X}) = (22, 28)$. Rescale $\mathbf{X}$ to an orthonormal design matrix, denoted $\tilde{\mathbf{X}}$, and rewrite the lasso regression loss function to:

$$\begin{aligned} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1 &= \left\| \mathbf{Y} - \mathbf{X}\begin{pmatrix} \sqrt{22} & 0 \\ 0 & \sqrt{28} \end{pmatrix}^{-1}\begin{pmatrix} \sqrt{22} & 0 \\ 0 & \sqrt{28} \end{pmatrix}\boldsymbol{\beta} \right\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1 \\ &= \|\mathbf{Y} - \tilde{\mathbf{X}}\gamma\|_2^2 + (\lambda_1/\sqrt{22})|\gamma_1| + (\lambda_1/\sqrt{28})|\gamma_2|, \end{aligned}$$

where $\gamma = (\sqrt{22}\beta_1, \sqrt{28}\beta_2)^\top$. By the same argument this loss can be minimized with respect to each element of γ separately. In particular, the soft-threshold function provides an analytic expression for the estimates of γ:

$$\begin{aligned} \hat{\gamma}_1(\lambda_1/\sqrt{22}) &= \text{sign}(\hat{\gamma}_1)[|\hat{\gamma}_1| - \tfrac{1}{2}(\lambda_1/\sqrt{22})]_+ &= -[9.892513 - \tfrac{1}{2}(10/\sqrt{22})]_+ &= -8.826509, \\ \hat{\gamma}_2(\lambda_1/\sqrt{28}) &= \text{sign}(\hat{\gamma}_2)[|\hat{\gamma}_2| - \tfrac{1}{2}(\lambda_1/\sqrt{28})]_+ &= [5.537180 - \tfrac{1}{2}(10/\sqrt{28})]_+ &= 4.592269, \end{aligned}$$

where $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are the ordinary least square estimates of γ_1 and γ_2 obtained from regressing $\mathbf{Y}$ on the corresponding column of $\tilde{\mathbf{X}}$. Rescale back and obtain the lasso regression estimate: $\hat{\boldsymbol{\beta}}(10) = (-1.881818, 0.8678572)^\top$. □

Question 6.1

Find the lasso regression solution for the data below for a general value of λ and for the straight line model $Y = \beta_0 + \beta_1 X + \varepsilon$ (only apply the lasso penalty to the slope parameter, not to the intercept). Show that when λ_1 is chosen as 14, the lasso solution fit is $\hat{Y} = 40 + 1.75X$. Data: $\mathbf{X}^\top = (X_1, X_2, \ldots, X_8)^\top = (-2, -1, -1, -1, 0, 1, 2, 2)^\top$, and $\mathbf{Y}^\top = (Y_1, Y_2, \ldots, Y_8)^\top = (35, 40, 36, 38, 40, 43, 45, 43)^\top$.