# The Solution Of Generalized Lasso For Non Full Rank Case

Presented by Jaesung Hwang

Seoul National University

*postechiminuru@gmail.com*

February 18, 2019

# Outline

# Outline

# Introduction

Let $\mathbf{y} \in \mathbb{R}^n$ be a response vector and $\mathbf{X} \in \mathbb{R}^{n \times p}$ be a matrix of predictors.

- Generalized Lasso Problem(Tibshirani et al., 2011):
  Generalized Lasso problem is written as:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \frac{1}{2} ||\mathbf{y} - \mathbf{X}\beta||_2^2 + \lambda \|D\boldsymbol{\beta}\|_1$$

  where $D \in \mathbb{R}^{m \times p}$ is a specified penalty matrix.

# Introduction

- The conventional solution path algorithm $(\mathrm{rank}(X) < p)$
  - By adding a little $\epsilon$ ridge penalty, the algorithm for full rank $X$ matrix can be applied.
  - For a fixed $\epsilon > 0$, consider

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|D\beta\|_1 + \epsilon\|\beta\|_2^2$$

which is the same as

$$\min_{\beta} \frac{1}{2}\|y^* - (X^*)\beta\|_2^2 + \lambda\|D\beta\|_1$$

where $y^* = (y^T, 0)^T$ and $X^* = \begin{bmatrix} X \\ \epsilon \cdot I \end{bmatrix}$

- We propose an exact solution path of the Generalized Lasso problem.
- When the solution is nonuniqueness, we characterize all solution sets and find various kinds of solutions.

# Outline

# Dual problem

- Divide the $\beta$ as follows:

$$\beta = V\eta + W\tau$$

where $V$ is the matrix that is orthogonal basis elements for the row space of $X$ in its columns and $W$ is the matrix that is orthogonal basis elements for the null space of $X$ in its columns i.e. $XW = 0$.

- Primal problem

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|D\beta\|_1$$

Since $\beta = V\eta + W\tau$,

$$\min_{\eta,\tau} \frac{1}{2}\|y - XV\eta\|_2^2 + \lambda\|DV\eta + DW\tau\|_1$$

Using the auxiliary variable $z$, we rewrite this problem as:

$$\min_{\eta,z,\tau} \frac{1}{2}\|y - XV\eta\|_2^2 + \lambda\|z\|_1 \qquad \text{subject to } DV\eta + DW\tau = z$$

# Dual problem

- Lagrangian function

$$\mathcal{L}(\eta, z, \tau, u) = \frac{1}{2}\|y - XV\eta\|_2^2 + \lambda\|z\|_1 + u^T(DV\eta + DW\tau - z)$$

- Dual function

$$\mathcal{D}(u) = \min_{\eta, z, \tau} \mathcal{L}(\eta, z, \tau, u)$$

Since $\eta$, $z$ and $\tau$ are decoupled in the Lagrangian function, we can minimize the Lagrangian function with repect to $\eta$, $z$ and $\tau$ separately.

- Dual problem

$$\max_{u \in \mathbb{R}^m} \mathcal{D}(u)$$

# Dual problem

- We rewrite the dual problem:

$$\min_{u \in \mathbb{R}^m} \frac{1}{2} \|\tilde{y} - \tilde{D}^T u\|_2^2 \tag{1}$$

subject to $\|u\|_\infty \leq \lambda$, $(DW)^T u = 0$,
where $\tilde{y} = XX^+ y$, $\tilde{D} = DX^+$ and $X^+ = (X^T X)^+ X^T$.

- A necessary and sufficient condition for $u$ to be a solution of the dual problem is that $u$ satisfy KKT conditions, since the dual problem is a convex problem.
- In order to find a solution path of the dual problem as $\lambda$ moves from $\infty$ to 0, the dual variable $u$ satisfying KKT conditions will be obtained.
  We define a boundary set $\mathcal{B}_\lambda$.

$$\mathcal{B}_\lambda = \{i : |u_i| = \lambda\}$$

# KKT condition

- For our problem (1), the KKT conditions are

$$(\tilde{D}\tilde{D}^T u)_i - (\tilde{D}\tilde{y})_i + \alpha\gamma_i + (DW\delta)_i = 0 \quad \text{for } i = 1, \cdots, m \tag{2}$$

where $u, \alpha, \gamma, \delta$ are subject to the constraints

$$\|u\|_\infty \leq \lambda \tag{3}$$

$$\alpha \geq 0 \tag{4}$$

$$\alpha(\|u\|_\infty - \lambda) = 0 \tag{5}$$

$$\|\gamma\|_1 \leq 1 \tag{6}$$

$$\gamma^T u = \|u\|_\infty \tag{7}$$

$$(DW)^T u = 0 \tag{8}$$

Constraints (6) and (7) say that $\gamma$ must be a subgradient of $\|u\|_\infty$ with respect to $u$.

# Algorithm Overview

- When $\lambda = \infty$, find a dual variable and lagrangian multipliers of the dual variable satisfying the KKT conditions.
- When $\lambda < \infty$, find a dual variable and lagrangian multipliers of the dual variable satisfying the KKT conditions.
- When $\lambda \leq \lambda_k$, find a dual variable and lagrangian multipliers of the dual variable satisfying the KKT conditions.
- Calculate event time $(\lambda_{k+1})$.

# How to find a dual variable and lagrangian multipliers ($\lambda = \infty$)

$\lambda = \lambda_0 = \infty$

- We can ignore the inequality constraint $\|u\|_\infty \leq \lambda$.
- The KKT conditions can be reduced to the following linear system.

$$\begin{bmatrix} \tilde{D}\tilde{D}^T & DW \\ (DW)^T & 0 \end{bmatrix} \begin{bmatrix} u(\lambda_0) \\ \delta(\lambda_0) \end{bmatrix} = \begin{bmatrix} \tilde{D}\tilde{y} \\ 0 \end{bmatrix} \tag{9}$$

- We solve the above linear system to obtain the dual variable $u(\lambda_0)$ and the lagrangian multiplier $\delta(\lambda_0)$ satisfying the KKT conditions.

# How to find a dual variable and lagrangian multipliers $(\lambda < \infty)$

$\lambda < \lambda_0 = \infty$

- $u(\lambda)$, $\delta(\lambda)$, $\alpha(\lambda)$ and $\gamma(\lambda)$ for satisfying KKT conditions are as follows:

$$u(\lambda) = \hat{u}(\lambda_0),\ \delta(\lambda) = \hat{\delta}(\lambda_0),\ \alpha(\lambda) = 0,$$

$$\gamma_i(\lambda) = \begin{cases} 1 \times \mathrm{sign}(\hat{u}_i(\lambda)) & \textit{If } i = \mathrm{argmax}_j\, |\hat{u}_j(\lambda)| \\ 0 & \textit{Otherwise} \end{cases}$$

- As the $\lambda$ decreases, the KKT condition $\|u(\lambda)\|_\infty \leq \lambda$ can be violated.
- Therefore, the $\lambda$, at which KKT condition is violated, is the maximum value of the absolute value of the dual variable $\hat{u}(\lambda)$. And insert the corresponding coordinate into boundary set $\mathcal{B}$.

$$\lambda_1 = \max_i(|\hat{u}_i(\lambda)|)$$

$$\mathcal{B}_{\lambda_1} = \mathcal{B}_{\lambda_0} \cup \{i|\, \mathrm{argmax}_i\, |\hat{u}_i(\lambda)|\}$$

where $\mathcal{B}_{\lambda_0} = \emptyset$.

# How to find a dual variable and lagrangian multipliers $(\lambda \leq \lambda_k)$

$\lambda \leq \lambda_k$

- The solution is given by $\hat{u}_{\mathcal{B}_{\lambda_k}}(\lambda) = \lambda s$ for the boundary coordinates where $s$ is a sign vector.
- To satisfy KKT conditions, $u_{-\mathcal{B}_{\lambda_k}}(\lambda)$ and $\delta(\lambda)$ satisfy the following linear system with the inequality constraint i.e.

$$\begin{bmatrix} \tilde{D}_{-\mathcal{B}_{\lambda_k}} \tilde{D}^T_{-\mathcal{B}_{\lambda_k}} & (DW)_{-\mathcal{B}_{\lambda_k}} \\ (DW)^T_{-\mathcal{B}_{\lambda_k}} & 0 \end{bmatrix} \begin{bmatrix} u_{-\mathcal{B}_{\lambda_k}}(\lambda) \\ \delta(\lambda) \end{bmatrix} = \begin{bmatrix} \tilde{D}_{-\mathcal{B}_{\lambda_k}}(\tilde{y} - \lambda \tilde{D}^T_{\mathcal{B}_{\lambda_k}} s) \\ -\lambda (DW)^T_{\mathcal{B}_{\lambda_k}} s \end{bmatrix}$$
$$(\tilde{D}_i(\tilde{y} - \lambda \tilde{D}^T_{\mathcal{B}_{\lambda_k}} s - \tilde{D}^T_{-\mathcal{B}_{\lambda_k}} u_{-\mathcal{B}_{\lambda_k}}(\lambda)) - D_i W \delta(\lambda)) \times s_i \geq 0, \quad i \in \mathcal{B}_{\lambda_k} \tag{10}$$

# How to find a dual variable and lagrangian multipliers $(\lambda \leq \lambda_k)$

- The solution $u_{-\mathcal{B}_{\lambda_k}}(\lambda)$ and $\delta(\lambda)$ of the linear system (10) have the following form:

$$\begin{bmatrix} \hat{u}_{-\mathcal{B}_{\lambda_k}}(\lambda) \\ \hat{\delta}(\lambda) \end{bmatrix} = \begin{bmatrix} \hat{u}^{-}_{-\mathcal{B}_{\lambda_k}}(\lambda_k) \\ \hat{\delta}^{-}(\lambda_k) \end{bmatrix} + (\lambda_k - \lambda)H^{\dagger} \begin{bmatrix} \tilde{D}_{-\mathcal{B}_{\lambda_k}} \tilde{D}^T_{\mathcal{B}_{\lambda_k}} s \\ (DW)^T_{\mathcal{B}_{\lambda_k}} s \end{bmatrix} \quad (11)$$

where $\hat{u}^{-}_{-\mathcal{B}_{\lambda_k}}(\lambda_k)$ and $\hat{\delta}^{-}(\lambda_k)$ are the dual variable and the lagrangian multiplier before updating at $\lambda_k$ and $H = \begin{bmatrix} \tilde{D}_{-\mathcal{B}_{\lambda_k}} \tilde{D}^T_{-\mathcal{B}_{\lambda_k}} & (DW)_{-\mathcal{B}_{\lambda_k}} \\ (DW)^T_{-\mathcal{B}_{\lambda_k}} & 0 \end{bmatrix}$.

- To satisfy the KKT conditions, $\alpha(\lambda)$ and $\gamma(\lambda)$ are as followings:

$$\alpha(\lambda) = \|\tilde{D}_{\mathcal{B}_{\lambda_k}}(\tilde{y} - \tilde{D}^T \hat{u}(\lambda)) - D_{\mathcal{B}_{\lambda_k}} W \hat{\delta}(\lambda)\|_1$$

$$\gamma_{-\mathcal{B}_{\lambda_k}}(\lambda) = 0 \text{ and } \gamma_{\mathcal{B}_{\lambda_k}}(\lambda) = \frac{1}{\alpha}(\tilde{D}_{\mathcal{B}_{\lambda_k}}(\tilde{y} - \tilde{D}^T \hat{u}(\lambda)) - D_{\mathcal{B}_{\lambda_k}} W \hat{\delta}(\lambda))$$

# Event time ($\lambda \leq \lambda_k$)

Check the KKT conditions.

- As we decrease $\lambda$, only two of the KKT conditions can be also violated:
  - The first is $\|u_{-\mathcal{B}_{\lambda_k}}(\lambda)\|_\infty \leq \lambda$
    - Insert the corresponding interior coordinates into the boundary set $\mathcal{B}_{\lambda_k}$.
  - The second is $\gamma^T u = \|u\|_\infty = \lambda$
    Since $\gamma^T u = \gamma_{\mathcal{B}_{\lambda_k}}^T u_{\mathcal{B}_{\lambda_k}}$ and $\|\gamma\|_1 = 1$, the second condition is $\mathrm{sign}(\gamma_{\mathcal{B}_{\lambda_k}}(\lambda)) = \mathrm{sign}(u_{\mathcal{B}_{\lambda_k}}(\lambda))$.
    There are two possibilities because $\gamma(\lambda)$ is related to $u(\lambda)$ and $\delta(\lambda)$.
    - $\delta(\lambda)$ changes.
    - One of the boundary set, which violate the condition, left out the boundary set $\mathcal{B}_{\lambda_k}$.

# Event time ($\lambda \le \lambda_k$)

- $\|u_{-\mathcal{B}_{\lambda_k}}(\lambda)\|_\infty \le \lambda$

$$h_{k+1} = \max_{i \in -\mathcal{B}_{\lambda_k}} \frac{u_i(\lambda_k) - \lambda_k \times l_i}{-l_i \pm 1} \tag{12}$$

- $\mathrm{sign}(\gamma_{\mathcal{B}_{\lambda_k}}(\lambda)) = \mathrm{sign}(u_{\mathcal{B}_{\lambda_k}}(\lambda))$

$$lv_{k+1} = \max_{i \in \mathcal{B}_{\lambda_k}} t_i^{(\text{leave})} \tag{13}$$

where $t_i^{(leave)} = \begin{cases} \frac{\xi_i s_i}{\eta_i s_i} & \text{if } \xi_i s_i < 0 \text{ and } \eta_i s_i < 0 \\ 0 & \text{otherwise.} \end{cases}$ and $\alpha \gamma_{\mathcal{B}_{\lambda_k}} = \boldsymbol{\xi} - \lambda \boldsymbol{\eta}$

- Thus we can know the next event time $\lambda_{k+1}$ that the KKT is violated by calculating a next hit time and a next leave time.

$$\lambda_{k+1} = \max\{h_{k+1}, lv_{k+1}\}$$

- Update the boundary set or update the $\delta$ according to whether next event time is a hit time or a leave time, and then calculate variables to satisfy the KKT.

# At leave time, when does the $\delta(\lambda)$ change?($\lambda \leq \lambda_k$)

- At leave time, there are two possibilities.
  - $\delta(\lambda_{k+1})$ changes.
  - One of the boundary set left out the boundary set $\mathcal{B}_{\lambda_k}$.
- The coordinate $i$, which is violated, is left out the boundary set $\mathcal{B}_{\lambda_k}$ temporarily.
- Let $l_i$ be the slope of $u_i$ and $s_i$ be the sign of $u_i$. So if $l_i \times s_i > 1$, then leave the coordinate $i$ out of the boundary set $\mathcal{B}_{\lambda_k}$, otherwise, $\delta(\lambda)$ changes.
- If $\delta(\lambda_{k+1})$ changes, then boudary set $(\mathcal{B}_{\lambda_{k+1}})$ is the same as $\mathcal{B}_{\lambda_k}$, so the slopes of $\hat{u}$ and $\delta$ does not change.
  To find $\delta(\lambda_{k+1})$, we solve the following linear system:

$$\begin{gathered}(DW)_{-\mathcal{B}_{\lambda_{k+1}}}\delta(\lambda_{k+1}) = \tilde{D}_{-\mathcal{B}_{\lambda_{k+1}}}(\tilde{y} - \tilde{D}^T\hat{u}(\lambda_s)) - (\lambda_s - \lambda_k)(DW)_{-\mathcal{B}_{\lambda_{k+1}}}\delta_l \\ -D_iW\delta(\lambda_{k+1}) \times s_i \geq \tilde{D}_i(\tilde{D}^T\hat{u}(\lambda_s) - \tilde{y}) \times s_i + (\lambda_s - \lambda_k)D_iW\delta_l \times s_i, \\ \text{for } \ i \in \mathcal{B}_{\lambda_{k+1}}\end{gathered} \tag{14}$$

where $\lambda_s = \lambda_{k+1} - \epsilon$

# Exact Dual Solution Path Algorithm

- Start with $k = 0$, $\lambda_0 = \infty$, $\mathcal{B}_0 = \emptyset$, and $\boldsymbol{s} = \emptyset$.
- Solve the linear block system (9) and $\lambda_1 = \max_i |\hat{u}_i|$, the corresponding coordinate is put in $\mathcal{B}_0$, and and the sign of the corresponding value is put in $\boldsymbol{s}$.
- While $\lambda_k > 0$:
  1. If the prior event is a hit event, calculate (11) to obtain the slope of $\hat{u}_{-\mathcal{B}_{\lambda_k}}$ and $\hat{\delta}$, $l$ and $\delta_l$.
  2. Compute the next hit time $h_{k+1}$ using (12).
  3. Compute the next leave time $lv_{k+1}$ using (13).
  4. $\lambda_{k+1} = \max(h_{k+1}, lv_{k+1})$
  5. If $h_{k+1} > lv_{k+1}$, then add the hitting coordinate and sign to $\mathcal{B}_{\lambda_k}$ and **s** and move to next iteration.
     Otherwise, move to next step.
  6. The coordinate $i$, which is violated, is left out the boundary set $\mathcal{B}_{\lambda_k}$ temporarily ($\tilde{\mathcal{B}}_{\lambda_k}$).
  7. Calculate (11) to obtain the slope of $\hat{u}_{-\tilde{\mathcal{B}}_{\lambda_k}}$ and $\hat{\delta}$, $\tilde{l}$ and $\tilde{\delta}_l$.
  8. If $\tilde{l}_i \times s_i > 1$, then leave the coordinate $i$ out of the boundary set $\mathcal{B}_{\lambda_k}$ and assign $\tilde{l}$, $\tilde{\delta}_l$ to $l$, $\delta_l$, otherwise, $\delta(\lambda)$ changes by solving (14).

# Recover a primal solution path from the dual solution path

- A primal solution path can be recovered from the dual solution path through the primal - dual relationships.
  - The relationship between $\eta$ and $u$ is:

    $$V\eta = (X^T X)^+ (X^T y - D^T u)$$

    This relationship can be obtained by setting the gradient of the $\mathcal{L}(\eta, z, \tau, u)$ with respect to $\eta$ equal to zero.
  - The relationship between $\tau$ and $u$ is:

    $$D_{-\mathcal{B}} W\tau = \tilde{D}_{-\mathcal{B}}(\tilde{D}^T u - \tilde{y})$$
    $$D_{\mathcal{B}} W\tau \times \mathrm{sign}(u_{\mathcal{B}}) \geq \tilde{D}_{\mathcal{B}}(\tilde{D}^T u - \tilde{y}) \times \mathrm{sign}(u_{\mathcal{B}})$$

    This relationship can be obtained by setting the gradient of the $\mathcal{L}(\eta, z, \tau, u)$ with respect to $z$ equal to zero.
  - $\beta = V\eta + W\tau$

# Recover primal solution

**Theorem**

*Primal variable $\tau(\lambda)$ is same as $-\delta(\lambda)$ which is a lagrange multiplier of dual variable $u(\lambda)$.*

- Through this algorithm, the primal solution, $\beta$, has the following relation with $u$ and $\delta$.

$$\beta = V\eta(\lambda) + W\tau(\lambda) = (X^T X)^{+}(X^T y - D^T u(\lambda)) - W\delta(\lambda)$$

# Outline

# Non-uniqueness of the solution

- When the solution of the generalized Lasso problem is given $\hat{\beta}_\lambda$, is it a unique solution?
- Clearly, if $\mathrm{null}(X) \cap \mathrm{null}(D) \neq \{0\}$, then the solution of the generalized Lasso problem is not unique for any $\lambda > 0$.
- But, if $\mathrm{null}(X) \cap \mathrm{null}(D) = \{0\}$, the uniqueness of the solutions depends on $\lambda$.
- For given $\hat{\beta}_\lambda$, we characterize the solutions set.

# Non-uniqueness of the solution

**Lemma**

*If $\hat{\beta}_1$, $\hat{\beta}_2$ are the solutions of the generalized Lasso problem at $\lambda$, then*

$$\frac{1}{2}\|y - X\hat{\beta}_1\|_2^2 = \frac{1}{2}\|y - X\hat{\beta}_2\|_2^2$$

- This means that $\hat{\beta}_2 = \hat{\beta}_1 + \gamma$ where $\gamma \in \mathrm{null}(X)$.

# Non-uniqueness of the solution

- Let $\hat{\beta}_\lambda$ be the solution of Generalized Lasso problem at $\lambda$.
- We can also define an active set : $\mathcal{A}_\lambda = \{i : d_i^T \hat{\beta}_\lambda \neq 0\}$.

**Theorem (Sufficient and necessary condition for non-uniqueness)**

$\exists \gamma \in \mathrm{null}(X)$ *such that*

$$\sum_{k \in \mathcal{A}_\lambda} \mathrm{sign}(d_k^T \hat{\beta}_\lambda) d_k^T \gamma + \sum_{k \in -\mathcal{A}_\lambda} |d_k^T \gamma| = 0 \tag{15}$$

*,if and only if the solution of Generalized Lasso problem is not unique at* $\lambda$.

# The Characterization of Solutions Set

## Theorem

*Suppose $\hat{\beta}_\lambda$ is the solution of Generalized Lasso problem at $\lambda$. At $\lambda$, all of the solutions ($\tilde{\beta}_\lambda$) follow the formula:*

$$\tilde{\beta}_\lambda = \hat{\beta}_\lambda + \gamma, \qquad \gamma \in \Gamma$$

*where* $\Gamma = \{\gamma \in \mathrm{null}(X) : \sum_{k \in \mathcal{A}_\lambda} \mathrm{sign}(d_k^T \hat{\beta}_\lambda) d_k^T \gamma + \sum_{k \in -\mathcal{A}_\lambda} |d_k^T \gamma| = 0,\ \mathrm{sign}(d_k^T(\hat{\beta}_\lambda + \gamma)) = \mathrm{sign}(d_k^T \hat{\beta}_\lambda) \text{ for } k \in \mathcal{A}_\lambda\}$.

- For given $\lambda$, the signs of $d_k^T \beta$ in the active set of all solutions are unchanged.
- We just characterize the set $\Gamma$ to find all solutions of Generalized Lasso problem for a given $\hat{\beta}_\lambda$.

- We can characterize the solutions set as follows:

$$\left\{\hat{\beta}_\lambda + \gamma \ : \ \gamma \in \bigcup_{\boldsymbol{s} \in \mathcal{S}} \Gamma_{\boldsymbol{s}}\right\}$$

where $\Gamma_s = \{WH_s\psi : F_sH_s\psi \geq 0, \ \ MH_s\psi \geq N\}$ and
$\mathcal{S}$ is a set of all combinations of sign vectors $\in (-1, 1)^{|-\mathcal{A}_\lambda|}$.
$F_s = s \times \left[d_k^T W\right]_{k \in -\mathcal{A}_\lambda}$ where $s$ is a sign vector
$H_s$ is the basis of the null space of $\mathbf{1}^T \left[\mathrm{sign}(d_k^T \hat{\beta}) d_k^T W\right]_{k \in \mathcal{A}_\lambda} + \mathbf{1}^T F_s$
$M = \mathrm{sign}\left(\left[d_k^T\right]_{k \in \mathcal{A}_\lambda} \hat{\beta}_\lambda\right) \times \left(\left[d_k^T W\right]_{k \in \mathcal{A}_\lambda}\right)$
$N = -\mathrm{sign}\left(\left[d_k^T\right]_{k \in \mathcal{A}_\lambda} \hat{\beta}_\lambda\right) \times \left(\left[d_k^T\right]_{k \in \mathcal{A}_\lambda} \hat{\beta}_\lambda\right)$.

# The Characterization of Solutions Set

- All $\Gamma_s$ have three cases:
  - All $\Gamma_s$ are zero vector set.
    - The given solution is unique solution.
  - All $\Gamma_s$ are the same set, not zero vector set.
    - The given solution is the solution with the largest active set.
  - Some $\Gamma_s$ are the same set which is not a zero vector set, and all other $\Gamma_s$ are zero vector sets.
    - The given solution is not the solution with the largest active set.
    - The signs of the coordinates contained in the largest active set are fixed. (by Theorem)
- Therefore $\Gamma$ is same as $\Gamma_s$ for a specific $s$.

# Types of the solutions

- We can express other solutions as follows:

$$\tilde{\beta}_\lambda = \hat{\beta}_\lambda + C\psi \quad \text{subject to } A\psi \geq B$$

where $A = \begin{bmatrix} F_s H_s \\ MH_s \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ N \end{bmatrix}$ and $C = \begin{bmatrix} WH_s \end{bmatrix}$.

- We can find the various kinds of solutions
  - The largest active set solution
  - The smallest active set solution
  - $l_2$ minimal solution
  - $l_\infty$ maximal solution

# How to find the sign vector

- In order to find $\Gamma_s$ which is not a zero vector set, we have to know the signs of the coordinates ($D_i\tilde{\beta}$ where $\tilde{\beta}$ has the solution with the largest active set) contained in the largest active set.
- Recall the derivation of the dual function. Since $\eta$, $z$ and $\tau$ are decoupled in the Lagrangian function, we can minimize the Lagrangian function with repect to $\eta$, $z$ and $\tau$ separately.

# How to find the sign vector

- The minimization over $z$ for given $u$ is as follows:

$$\min_z \lambda \|z\|_1 - u^T z$$

Since the problem is convex problem, the minimizer $\hat{z}$ satisfy the KKT condition:

$$\lambda\gamma - u = 0$$

where $\gamma$ is a subgradient of $\|\cdot\|_1$ at $\hat{z}$ i.e.

$$\gamma_i = \begin{cases} \mathrm{sign}(\hat{z}_i) & \text{if } \hat{z}_i (= D_i\hat{\beta}) \neq 0 \\ \left[-1, 1\right] & \text{if } \hat{z}_i (= D_i\hat{\beta}) = 0 \end{cases}$$

which is equivalent to

$$u_i = \begin{cases} \lambda \mathrm{sign}(\hat{z}_i) & \text{if } \hat{z}_i (= D_i\hat{\beta}) \neq 0 \\ \left[-\lambda, \lambda\right] & \text{if } \hat{z}_i (= D_i\hat{\beta}) = 0 \end{cases}$$

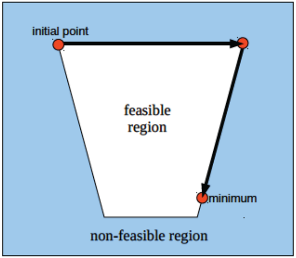
- This means that the boundary set $\mathcal{B}_\lambda$ contains the active set $\mathcal{A}_\lambda$.

# $l_2$ minimal solution

- We can find the $l_2$ minimal solution by solving the problem:

$$\min_{\psi} \|\hat{\beta}_\lambda + C\psi\|_2^2 \quad \text{subject to } A\psi \geq B$$

- Since this problem is quadratic program and $C = WH_s$ is full column rank, if $\psi = -(C^T C)^{-1} C^T \hat{\beta}_\lambda$ does not satisfy the inequality constraint, then the solution of the problem is on the boundary of the feasible set.
- By using the Binding-Direction Primal Active-set algorithm, we find the solution of the problem.



# $l_\infty$ maximal solution

- We can find the $l_\infty$ maximal solution by solving the problem:

$$\max_{\psi} \|\hat{\beta}_\lambda + C\psi\|_\infty \quad \text{subject to } A\psi \geq B$$

- Since $\|\cdot\|_\infty$ is the convex function, the maximizer is on the boundary of the feasible set.

# Simple Example

- We characterize other solutions for a given $\hat{\beta}_\lambda$ of the following two simple examples.
- The two examples have the same $y$ and $X$, but different penalty matrix $D$.

$$y = \begin{bmatrix} 1000 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}$$

1. $D = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0.1 & 0 & 0 \\ 0 & 0.1 & 0 \end{bmatrix}$
2. $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$

- The basis matrix $W$ of the null space of $X$ is as following:

$$W = \begin{bmatrix} 0 & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \\ 1 & 0 \end{bmatrix}$$

# Simple Example

When $\lambda$ is 2000 for **the first example** and $\hat{\beta}_\lambda$ we found is $\begin{bmatrix} 400 & 400 & 400 \end{bmatrix}^T$.
An active set $\mathcal{A}$ is $\{3,4\}$. And a sign set $\mathcal{S}$ is $\{[1,1]^T, [-1,1]^T, [1,-1]^T, [-1,-1]^T\}$.

- $s = [1,1]^T$
  $\Gamma_s = \left\{ [0 \quad 0 \quad 0]^T \right\}$
- $s = [-1,1]^T$
  $\Gamma_s = \left\{ [0 \quad 0 \quad 0]^T \right\}$
- $s = [1,-1]^T$
  $\Gamma_s = \left\{ [0 \quad 0 \quad 0]^T \right\}$
- $s = [-1,-1]^T$
  $\Gamma_s = \left\{ [0 \quad 0 \quad 0]^T \right\}$

Therefore, in this example, $\hat{\beta}_\lambda = \begin{bmatrix} 400 & 400 & 400 \end{bmatrix}^T$ is a unique solution for a given $\lambda = 2000$.

# Simple Example

We characterize the other solutions when $\lambda$ is 700 for **the second example** and $\hat{\beta}_\lambda$ we found is $\begin{bmatrix} 200 & 100 & 100 \end{bmatrix}^T$.
An active set $\mathcal{A}$ is $\{1, 2\}$. And a sign set $\mathcal{S}$ is $\{[1], [-1]\}$.

- $s = [1]$
  $\Gamma_s = \left\{ [-\psi \quad \psi \quad \psi]^T : -100 \leq \psi \leq 200 \right\}$
- $s = [-1]$
  $\Gamma_s = \left\{ [-\psi \quad \psi \quad \psi]^T : -100 \leq \psi \leq 200 \right\}$

Therefore, in this example, for a given $\lambda = 700$, we can characterize other solutions as follows:

$$\tilde{\beta}_\lambda = \begin{bmatrix} 200 - \psi \\ 100 + \psi \\ 100 + \psi \end{bmatrix} \quad , -100 \leq \psi \leq 200$$

# Outline

# Future work

- Real data analysis for nonuniqueness solution

# The End