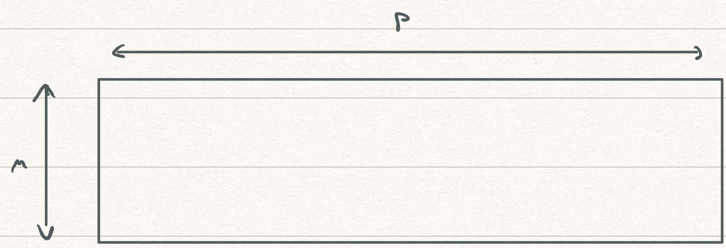
$$y = Ax + b$$

où $A \in \mathbb{R}^{n \times p}$, $b \in \mathbb{R}^n$, $y \in \mathbb{R}^n$, $x \in \mathbb{R}^p$

"grande dimension"

$A =$ 

$p \gg n$

<u>Connus</u> : y (observation) et A (design)

<u>Cible</u> : $x \in \mathbb{R}^p$ et/ou Ax

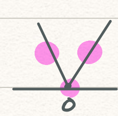
<u>perturbation</u> : $\varepsilon \in \mathbb{R}^n$

$$\hat{x}_\lambda \in \arg\min_{z \in \mathbb{R}^p} \left\{ \frac{1}{2} \|y - Az\|_2^2 + \lambda \|z\|_1 \right\}$$

(lasso)

où $\lambda > 0$

$\|z\|_1 = \sum_{i=1}^{p} |z_i|$



$\forall h, \forall v$ tq $|v| \leq 1$, $|t+h| \geq |t| + vh$ où $t=0$

$h \in \partial|.|(0)$

$\forall h \in \mathbb{R}^p, \forall v \in \mathbb{R}^p, \|v\|_\infty \leq 1$

$$\|t+h\|_1 \geq \|t\|_1 + \langle v, h\rangle \quad \text{où } \underline{t=0}$$

(Hölder $\langle v,h\rangle \leq \|h\|_1 \|v\|_\infty \leq \|h\|_1$)

$$\partial\|.\|_1(0) = \{v : \|v\|_\infty \leq 1\} = B_\infty$$

$t = (\pi, 0, -e)$

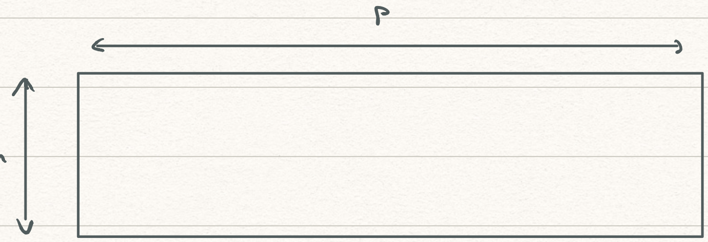
- $\partial\|.\|_1(t) = \left\{\left(\underset{v_1}{1}, \underset{v_2}{v}, \underset{v_3}{-1}\right) : |v| \leq 1\right\}$

$$\|t+h\|_1 = \sum_{i=1}^{p} |t_i + h_i| \geq \sum_{i=1}^{p} |t_i| + v_i h_i$$

- $sgn(t) = (v_1, \dots, v_p)$ où
  - $v_i = 1$ si $t_i > 0$
  - $v_i = -1$ si $t_i < 0$
  - $v_i \in [-1,1]$ si $t_i = 0$

$$\boxed{\partial\|.\|_1(t) = sgn(t)}$$

$$\bullet \quad \boxed{x_0 \in \arg\min_{Az = Ax} \|z\|_1 = \arg\min_{z} \left\{ \|z\|_1 + \mathbb{I}_{Az = Ax} \right\}}$$

$A =$ 

$p >> m$

$$\mathcal{L}(\underset{\in \mathbb{R}^p}{z}, \underset{\in \mathbb{R}^m}{c}) = \|z\|_1 + \langle Az - Ax, c \rangle$$

$$\sup_c \mathcal{L}(z,c) = \|z\|_1 + \mathbb{I}_{Az = Ax}$$ <u>primal</u>

<u>c fixé</u> $\quad \inf_z \mathcal{L}(z,c)$

$$= \inf_z \left\{ \langle Az, c \rangle + \|z\|_1 - \langle Ax, c \rangle \right\}$$

$$= -\langle Ax, c \rangle + \inf_z \left\{ \langle z, A^T c \rangle + \|z\|_1 \right\}$$

$$= -\langle Ax, c \rangle - \boxed{\sup_z \left\{ \langle A^T c, z \rangle - \|z\|_1 \right\}}$$

- Si $\|A^T c\|_\infty \leq 1$ alors

$$|\langle A^T c, z\rangle| \leq \|A^T c\|_\infty \, \|z\|_1$$

$$\Rightarrow \quad \langle A^T c, z\rangle - \|z\|_1 \leq 0 = \langle A^T c, 0\rangle - \|0\|_1$$

- Si $\|A^T c\|_\infty > 1$ alors

$v = A^T c = (v_1, \dots, v_p)$ $\quad \exists k_0$ tq $|v_{k_0}| > 1$

on suppose $v_{k_0} > 1$,

$$z = (0, \dots, 0, \underset{\substack{\uparrow \\ k_0}}{t}, 0 \dots 0) \qquad \|z\|_1 = t$$

$$t > 0 \qquad \langle z, A^T c\rangle = v_{k_0} t$$

$$\longrightarrow \quad \langle z, A^T c\rangle - \|z\|_1 = \underbrace{(v_{k_0} - 1)}_{>0} t \xrightarrow{t \to +\infty} +\infty$$

Fenchel dual

$$\boxed{\|.\|_1^*(A^T c) = \mathbb{C}_{\|A^T c\|_\infty \leq 1}}$$

$$\inf_{z} \mathcal{L}(z,c) = -\langle Ax, c\rangle - \mathbb{C}_{\|A^T c\|_\infty \leq 1}$$

<u>Dual</u>: $$\sup_{c} \left\{ -\langle Ax, c\rangle - \mathbb{C}_{\|A^T c\|_\infty \leq 1} \right\}$$

$$= - \inf_{c} \left\{ \langle Ax, c\rangle + \mathbb{C}_{\|A^T c\|_\infty \leq 1} \right.$$

$$= \underbrace{- \inf_{\|A^T c\|_\infty \leq 1} \left\{ \langle Ax, c\rangle \right\}}_{\text{Dual}}$$

<u>Condition Slater</u>: $\|c\|_2 < \varepsilon \Rightarrow \|A^T c\|_\infty \leq 1$

↳ No duality gap

Primal = Dual

$$\begin{cases} x_0 \in \arg\min\limits_{Az = Ax} \|z\|_1 & \text{Primal} \\ c_0 \in \arg\left\{ -\min\limits_{\|A^T c\|_\infty \leq 1} \langle Ax, c\rangle \right\} & \text{Dual} \end{cases}$$

No duality gap $\Rightarrow$ $\boxed{\begin{cases} \|x_0\|_1 = -\langle Ax, c_0\rangle \\ Ax_0 = Ax \\ \|A^T c_0\|_\infty \leq 1 \end{cases}}$

$$\boxed{\eta_0 = -A^T c_0}$$

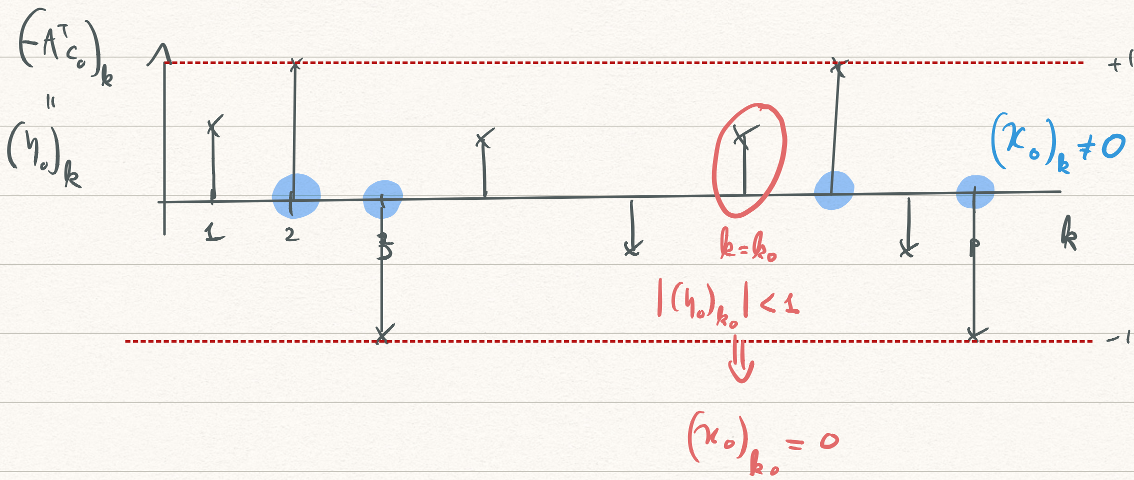
- $$\boxed{\begin{aligned} \langle x_0, \eta_0 \rangle &= -\langle x_0, A^T c_0\rangle \\ &= -\langle Ax_0, c_0\rangle \\ &= -\langle Ax, c_0\rangle \\ &= \|x_0\|_1 \\ &= \|x_0\|_1 + \mathbb{I}_{\|\eta_0\|_\infty \leq 1} \end{aligned}}$$

- $$\boxed{\langle x_0, \eta_0 \rangle = \|x_0\|_1 + \mathbb{I}_{\|\eta_0\|_\infty \leq 1}}$$

$$\sum_{k=1}^{p} \underline{|(x_0)_k|} = \sum_{k=1}^{p} \underline{(x_0)_k}\,\underline{(\eta_0)_k}$$

$$\Leftrightarrow \quad \eta_0 = \operatorname{sgn}(x_0)$$

$$\Leftrightarrow \quad \boxed{-A^T c_0 = \operatorname{sgn}(x_0)}$$



F convex $\qquad F(x) = \| \ \|_1$

$F^* \qquad F^*(\eta) = \mathbb{C}_{\|\eta\|_\infty \leq 1}$

$\cdot \ F^*(\eta) = \boxed{\sup_{z} \left\{ \langle \eta, z \rangle - F(z) \right\}}$

$$\Big\downarrow \partial$$

$$\eta - \partial F(z) = 0$$

$$\Longleftrightarrow \boxed{\eta \in \partial F(z^*)}$$

$\cdot \ F^{**}(x) = F(x) = \boxed{\sup_{\rho} \left\{ \langle \rho, x \rangle - F^*(\rho) \right\}}$

$$x \in \partial F^*(\rho^*)$$

$$\langle x_0, y_0 \rangle = F(x_0) + F^*(y_0)$$

$$\Rightarrow \quad F^*(y_0) = \boxed{\langle x_0, y_0 \rangle - F(x_0)}$$

$$y_0 \in \partial F(x_0)$$

$$\Rightarrow \quad F^{**}(x_0) = F(x_0) = \boxed{\langle x_0, y_0 \rangle - F^*(y_0)}$$

$$\underline{x_0 \in \partial F^*(y_0)}$$

$\hookrightarrow \quad |y_{0,k}| < 1 \Rightarrow (x_0)_k = 0$

x fixé et je cherche $\underline{x_0 = x}$

Si A est telle que $\exists\, c$ tq

certificat dual $\boxed{\eta = -A^T c}$ vérifie : (+ $\boxed{\underline{A_S \text{ inversible}}}$)

$$\left\{ \eta_k = \frac{x_k}{|x_k|} \quad \text{quand } x_k \neq 0 \right.$$

η_k est tel que $|\eta_k| < 1$

quand $x_k = 0$

Alors : $\boxed{x_0 = x}$

Dem : $\|x_0\|_1 \leq \|x\|_1 = \langle \eta , x \rangle$

$= \langle -A^T c , x \rangle$

$= \langle -c , Ax \rangle$

$= \langle -c , Ax_0 \rangle$

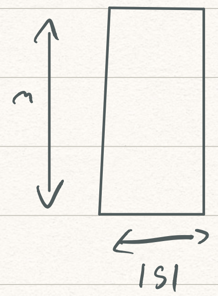
$= \langle -A^T c , x_0 \rangle$

$\langle x_0 , \eta \rangle \geqslant \|x_0\|_1$

Holder $\langle x_0 , \eta \rangle = \|x_0\|_1$

$\Rightarrow (x_0)_k = 0 \iff x_k = 0$

(F) $\quad Ax_0 = Ax \iff A_S x_0 = A_S x$

$A_S =$ 

$S = \{k : x_k \neq 0\}$

support de x

$\iff x_0 = x.$

$$\boxed{\hat{x} \in \arg\min_{\substack{z,u \\ u = Az}} \left\{ \frac{1}{2} \| y - \underbrace{Az}_{u} \|_2^2 + \lambda \|z\|_1 \right\}}$$ (Primal)

$$\mathcal{L}(\underbrace{z, u}_{\text{primal}}, \underbrace{c}_{\text{dual}}) = \frac{1}{2}\|y - u\|_2^2 + \langle u - Az, c \rangle + \lambda \|z\|_1$$

$$\sup_{c} \mathcal{L} = \frac{1}{2}\|y - u\|_2^2 + \lambda\|z\|_1 + \mathbb{C}_{u=Az}$$ (Primal)

$$\inf_{z,u} \mathcal{L} = \inf_{z,u} \left\{ \frac{1}{2} \|y-u\|_2^2 + \langle u, c \rangle + \lambda \left[\|z\|_1 - \left\langle z, \frac{A^T c}{\lambda} \right\rangle \right] \right\}$$

$$= \inf_{u} \left\{ \frac{1}{2} \|y-u\|_2^2 + \langle u, c \rangle \right\} + \lambda \inf_{z} \left\{ \|z\|_1 - \left\langle z, \frac{A^T c}{\lambda} \right\rangle \right\}$$

$$= - \sup_{u} \left\{ \underbrace{\langle u, -c \rangle - \frac{1}{2} \|y-u\|_2^2}_{G(u)} \right\} - \lambda \underbrace{\sup_{z} \left\{ \left\langle z, \frac{A^T c}{\lambda} \right\rangle - \|z\|_1 \right\}}_{F^*\left(\frac{A^T c}{\lambda}\right) = \mathbb{1}_{\left\| \frac{A^T c}{\lambda} \right\|_\infty \leq 1}}$$

$$\nabla G(u) = -c - (u - y) = 0$$

$$\Leftrightarrow \underline{u = y - c}$$

$$G(y-c) = \langle y-c, -c \rangle - \frac{1}{2} \|y - y - c\|_2^2$$

$$= -\langle y, c \rangle + \|c\|_2^2 - \frac{1}{2} \|c\|_2^2$$

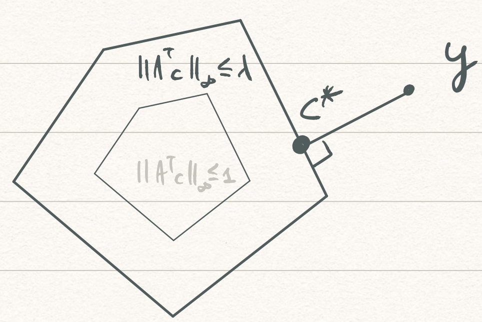
$$G(y - c) = \frac{1}{2}\|y - c\|_2^2 - \frac{1}{2}\|y\|_2^2$$

$$\operatorname*{Inf}_{u,z} \mathcal{L}(z, u, c) = \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y - c\|_2^2 - \iota_{\|A^T c\|_\infty \leq \lambda}$$

<u>Dual</u>

$$\sup_{c : \|A^T c\|_\infty \leq \lambda} \left\{ \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y - c\|_2^2 \right\}$$

$$= - \boxed{\inf_{c : \|A^T c\|_\infty \leq \lambda} \left\{ \frac{1}{2}\|y - c\|_2^2 \right\}} + \frac{1}{2}\|y\|_2^2$$



c^* proj orthogonal de y sur $\underbrace{\{c : \|A^T c\|_\infty \leq \lambda\}}_{\text{polytope convexe}}$

No duality Gap (Slater) $\quad \{c : \|A^T c\|_\infty \le \lambda\} \neq \emptyset$

$$\begin{cases} \hat{x} \text{ sol Primal} & (\leftrightarrow z) \\ \hat{y} = A\hat{x} & (\leftrightarrow u) \\ c^* \text{ sol duale} & \end{cases}$$

Primal: $\quad \frac{1}{2}\|y - \hat{y}\|_2^2 + \lambda \|\hat{x}\|_1$

$\|$

Dual: $\quad -\frac{1}{2}\|y - c^*\|_2^2 + \frac{1}{2}\|y\|_2^2$

$\|$

$$\mathcal{L}(\hat{x}, \hat{y}, c^*) = \frac{1}{2}\|y - \hat{y}\|_2^2 + \langle \hat{y} - A\hat{x}, c^* \rangle + \lambda \|\hat{x}\|_1$$

$$= \frac{1}{2}\|y - \hat{y}\|_2^2 + \langle \hat{y}, c^* \rangle$$

$$+ \lambda \left[\|\hat{x}\|_1 - \langle \frac{A\hat{x}}{\lambda}, c^* \rangle \right]$$

$$\hat{y} = y - c^*$$

$$\langle \frac{A^T c^*}{\lambda}, \hat{x} \rangle = \|\hat{x}\|_1$$

$$\eta = \frac{A^T c^*}{\lambda} = \frac{1}{\lambda}\left[A^T (\hat{y} - y) \right] = \frac{1}{\lambda} A^T (A\hat{x} - y)$$

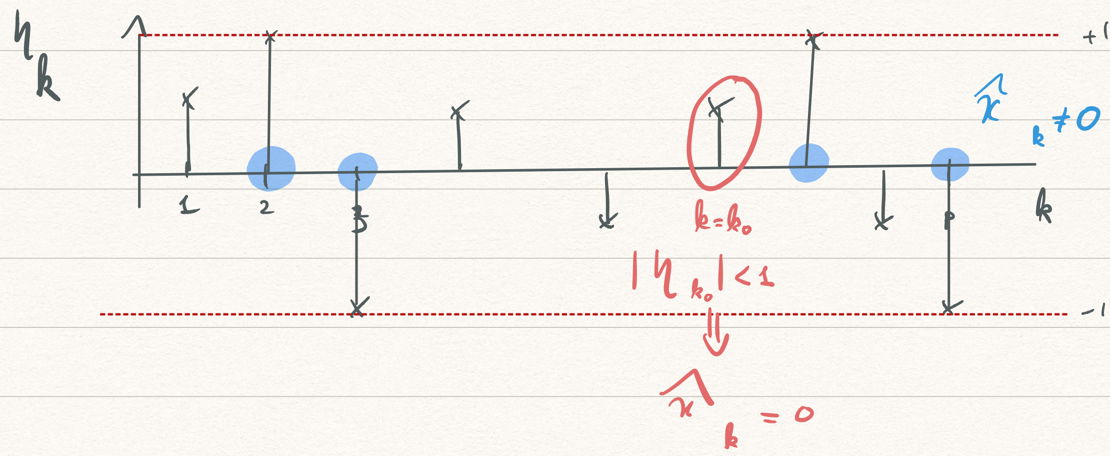
$\longrightarrow$

$$\|\eta\|_\infty \leq 1$$

$$\langle \eta , \hat{x} \rangle = \|\hat{x}\|_1 \qquad \eta = \mathrm{sgn}(\hat{x})$$

$$\hat{y} = y - c^*$$

$$\eta = \frac{A^T c^*}{\lambda} = \frac{1}{\lambda} A^T (A\hat{x} - y)$$



Théorème $A \in \mathbb{R}^{n \times p}$

$F : \mathbb{R}^n \longrightarrow \mathbb{R}$ $\quad (F(u) = \frac{1}{2} \|y-u\|_2^2)$

$G : \mathbb{R}^p \longrightarrow \mathbb{R}$ $\quad (G(z) = \lambda \|z\|_1)$

Dualité forte (No dualité Gap)

$$\min_{z \in \mathbb{R}^p} \left\{ F(Az) + G(z) \right\} = \max_{c \in \mathbb{R}^n} \left\{ -F^*(c) - G^*(-A^T c) \right\}$$

Solution (z^*, c^*) est un point selle de Lagrangien

$$\min_{z} \max_{c} \left\{ \langle Az, c \rangle + G(z) - F^*(c) \right\}$$

$\oplus \quad A^T c^* \in \partial G(z^*)$

$$y = Ax + b$$

$$\hat{y} = A\hat{x}$$

$$\hat{x} = \hat{x}_\lambda \in \arg\min_{z} \left\{ \frac{1}{2} \|y - Az\|_2^2 + \lambda \|z\|_1 \right\}$$

<u>Théorème</u> : <u>Si</u> $\lambda \geq 3 \|A^T b\|_\infty$ <u>alors</u>

$$\|\underbrace{A\hat{x}}_{\hat{y}} - Ax\|_2^2 \leq \inf_{\substack{z : \\ z \neq 0}} \left\{ \|Az - Ax\|_2^2 + \frac{\lambda^2}{\kappa^2} \|z\|_0 \right\}$$

où

$$s = \#\{k : x_k \neq 0\} = \#S$$

$$\|z\|_0 = \#\{k : z_k \neq 0\}$$

$$\kappa = \min_{v :\ 5\|v_S\|_1 > \|v_{S^c}\|_1} \left\{ \frac{\sqrt{s}\, \|Av\|_2}{\|v_S\|_1} \right\}$$

$$(v_S)_k = \begin{cases} v_k & \text{si } k \in S \\ 0 & \text{sinon} \end{cases}$$

$$S = \{k : x_k \neq 0\}$$

4.2.3 Oracle Risk Bound

We have proved in Chapter 2 that the risk of the model selection estimator (2.9) can be nicely bounded; see Theorem 2.2 and Exercise 2.9.2, part A. We derive in this section a risk bound for the Lasso estimator $\widehat{f}_\lambda = \mathbf{X}\widehat{\beta}_\lambda$, which is similar, at least for some classes of design matrix $\mathbf{X}$.

The best risk bounds available in the literature involve the so-called compatibility constant

$$\kappa(\beta) = \min_{v \in \mathscr{C}(\beta)} \left\{ \frac{\sqrt{|m|}\, \|\mathbf{X}v\|}{|v_m|_1} \right\},$$
$$\text{where } m = \mathrm{supp}(\beta) \text{ and } \mathscr{C}(\beta) = \{v \in \mathbb{R}^p :\ 5|v_m|_1 > |v_{m^c}|_1\}. \quad (4.8)$$

This compatibility constant is a measure of the lack of orthogonality of the columns of $\mathbf{X}_m$; see Exercise 4.5.3. We emphasize that it can be very small for some matrices $\mathbf{X}$. We refer again to Exercise 4.5.3 for a simple lower bound on $\kappa(\beta)$.

$X \leftarrow A$
$\beta^* \leftarrow x$
$\hat{\beta} \leftarrow \hat{x}$
$\varepsilon \leftarrow b$

A deterministic bound

We first state a deterministic bound and then derive a risk bound from it.

Theorem 4.1 A deterministic bound

For $\lambda \geq 3|\mathbf{X}^T\varepsilon|_\infty$ we have

$$\|\mathbf{X}(\widehat{\beta}_\lambda - \beta^*)\|^2 \leq \inf_{\beta \in \mathbb{R}^p \setminus \{0\}} \left\{ \|\mathbf{X}(\beta - \beta^*)\|^2 + \frac{\lambda^2}{\kappa(\beta)^2} |\beta|_0 \right\}, \quad (4.9)$$

with $\kappa(\beta)$ defined by (4.8).

Proof. The proof mainly relies on the optimality condition (4.2) for (4.4) and some simple (but clever) algebra. $\mathcal{L} = \|y - Az\|_2^2 + \lambda\|z\|_1$

Optimality condition: We have $0 \in \partial\mathscr{L}(\widehat{\beta}_\lambda)$. Since any $\widehat{w} \in \partial\mathscr{L}(\widehat{\beta}_\lambda)$ can be written as $\widehat{w} = -2\mathbf{X}^T(Y - \mathbf{X}\widehat{\beta}_\lambda) + \lambda\widehat{z}$ with $\widehat{z} \in \partial|\widehat{\beta}_\lambda|_1$, using $Y = \mathbf{X}\beta^* + \varepsilon$ we obtain that there exists $\widehat{z} \in \partial|\widehat{\beta}_\lambda|_1$ such that $2\mathbf{X}^T(\mathbf{X}\widehat{\beta}_\lambda - \mathbf{X}\beta^*) - 2\mathbf{X}^T\varepsilon + \lambda\widehat{z} = 0$. In particular, for all $\beta \in \mathbb{R}^p$

$$2\langle \mathbf{X}(\widehat{\beta}_\lambda - \beta^*), \mathbf{X}(\widehat{\beta}_\lambda - \beta)\rangle - 2\langle \mathbf{X}^T\varepsilon, \widehat{\beta}_\lambda - \beta\rangle + \lambda\langle \widehat{z}, \widehat{\beta}_\lambda - \beta\rangle = 0. \quad (4.10)$$

Convexity: Since $|.|_1$ is convex, the subgradient monotonicity ensures that $\langle \widehat{z}, \widehat{\beta}_\lambda - \beta\rangle \geq \langle z, \widehat{\beta}_\lambda - \beta\rangle$ for all $z \in \partial|\beta|_1$. Therefore, Equation (4.10) gives

for all $\beta \in \mathbb{R}^p$ and for all $z \in \partial|\beta|_1$ we have,

$$2\langle \mathbf{X}(\widehat{\beta}_\lambda - \beta^*), \mathbf{X}(\widehat{\beta}_\lambda - \beta)\rangle \leq 2\langle \mathbf{X}^T\varepsilon, \widehat{\beta}_\lambda - \beta\rangle - \lambda\langle z, \widehat{\beta}_\lambda - \beta\rangle. \quad (4.11)$$

The next lemma provides an upper bound on the right-hand side of (4.11).

$m \subset S$

Lemma 4.2

We set $m = \mathrm{supp}(\beta)$. There exists $z \in \partial|\beta|_1$, such that for $\lambda \geq 3|\mathbf{X}^T\varepsilon|_\infty$ we have

1. *the inequality* $2\langle \mathbf{X}^T\varepsilon, \widehat{\beta}_\lambda - \beta\rangle - \lambda\langle z, \widehat{\beta}_\lambda - \beta\rangle \leq 2\lambda|(\widehat{\beta}_\lambda - \beta)_m|_1$,
2. *and* $5|(\widehat{\beta}_\lambda - \beta)_m|_1 > |(\widehat{\beta}_\lambda - \beta)_{m^c}|_1$ *when* $\langle \mathbf{X}(\widehat{\beta}_\lambda - \beta^*), \mathbf{X}(\widehat{\beta}_\lambda - \beta)\rangle > 0$.

Proof of the lemma

1. Since $\partial|z|_1 = \left\{z \in \mathbb{R}^p : z_j = \mathrm{sign}(\beta_j) \text{ for } j \in m \text{ and } z_j \in [-1,1] \text{ for } j \in m^c\right\}$, we can choose $z \in \partial|\beta|_1$, such that $z_j = \mathrm{sign}([\widehat{\beta}_\lambda - \beta]_j) = \mathrm{sign}([\widehat{\beta}_\lambda]_j)$ for all $j \in m^c$. Using the duality bound $\langle x, y\rangle \leq |x|_\infty |y|_1$, we have for this choice of z

$$\begin{aligned}
&2\langle \mathbf{X}^T\varepsilon, \widehat{\beta}_\lambda - \beta\rangle - \lambda\langle z, \widehat{\beta}_\lambda - \beta\rangle \\
&\quad = 2\langle \mathbf{X}^T\varepsilon, \widehat{\beta}_\lambda - \beta\rangle - \lambda\langle z_m, (\widehat{\beta}_\lambda - \beta)_m\rangle - \lambda\langle z_{m^c}, (\widehat{\beta}_\lambda - \beta)_{m^c}\rangle \\
&\quad \leq 2|\mathbf{X}^T\varepsilon|_\infty|\widehat{\beta}_\lambda - \beta|_1 + \lambda|(\widehat{\beta}_\lambda - \beta)_m|_1 - \lambda|(\widehat{\beta}_\lambda - \beta)_{m^c}|_1 \\
&\quad \leq \frac{5\lambda}{3}|(\widehat{\beta}_\lambda - \beta)_m|_1 - \frac{\lambda}{3}|(\widehat{\beta}_\lambda - \beta)_{m^c}|_1 \qquad (4.12) \\
&\quad \leq 2\lambda|(\widehat{\beta}_\lambda - \beta)_m|_1,
\end{aligned}$$

   where we used $3|\mathbf{X}^T\varepsilon|_\infty \leq \lambda$ and $|\widehat{\beta}_\lambda - \beta|_1 = |(\widehat{\beta}_\lambda - \beta)_{m^c}|_1 + |(\widehat{\beta}_\lambda - \beta)_m|_1$ for the Bound (4.12).
2. When $\langle \mathbf{X}(\widehat{\beta}_\lambda - \beta^*), \mathbf{X}(\widehat{\beta}_\lambda - \beta)\rangle > 0$, combining (4.11) with (4.12) give the inequality $5|(\widehat{\beta}_\lambda - \beta)_m|_1 > |(\widehat{\beta}_\lambda - \beta)_{m^c}|_1$. □

We now conclude the proof of Theorem 4.1. Al-Kashi formula gives

$$2\langle \mathbf{X}(\widehat{\beta}_\lambda - \beta^*), \mathbf{X}(\widehat{\beta}_\lambda - \beta)\rangle = \|\mathbf{X}(\widehat{\beta}_\lambda - \beta^*)\|^2 + \|\mathbf{X}(\widehat{\beta}_\lambda - \beta)\|^2 - \|\mathbf{X}(\beta - \beta^*)\|^2.$$

When this quantity is nonpositive, we have directly (4.9). When this quantity is positive, we can combine it with (4.11) and apply successively the first part of the above lemma, the second part of the lemma with (4.8), and finally $2ab \leq a^2 + b^2$ to get that for all $\beta \in \mathbb{R}^p$

$$\begin{aligned}
\|\mathbf{X}(\widehat{\beta}_\lambda - \beta^*)\|^2 + \|\mathbf{X}(\widehat{\beta}_\lambda - \beta)\|^2 &\leq \|\mathbf{X}(\beta - \beta^*)\|^2 + 2\lambda|(\widehat{\beta}_\lambda - \beta)_m|_1 \\
&\leq \|\mathbf{X}(\beta - \beta^*)\|^2 + \frac{2\lambda\sqrt{|\beta|_0}}{\kappa(\beta)}\|\mathbf{X}(\widehat{\beta}_\lambda - \beta)\| \\
&\leq \|\mathbf{X}(\beta - \beta^*)\|^2 + \frac{\lambda^2|\beta|_0}{\kappa(\beta)^2} + \|\mathbf{X}(\widehat{\beta}_\lambda - \beta)\|^2.
\end{aligned}$$

The proof of Theorem 4.1 is complete. □

If the tuning parameter λ of the Lasso estimator is such that $\lambda \geq 3|\mathbf{X}^T\varepsilon|_\infty$ with high probability, then (4.9) holds true with high probability for this choice of λ. We state in the next corollary such a risk bound in the Gaussian setting (2.3).