① The SVD can be written as: $A = U\Sigma V^T$ and the pseudoinverse as: $A^+ = V\Sigma^+ U^T$
where: if $m>n$, then:

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{m\times n} \qquad \Sigma^+ = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n} \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{m\times n} \quad ; r = \operatorname{rank}(A) \; ; \sigma_i \neq 0$$

if any $\sigma_k = 0$ then its inverse is 0.

The alternative definition of pseudoinverse is:

$$A^\dagger = \lim_{\lambda\to+0} (A^TA+\lambda I)^{-1}A^T = \lim_{\lambda\to+0} \left((U\Sigma V^T)^T(U\Sigma V^T)+\lambda I\right)^{-1}(U\Sigma V^T)^T =$$

Points for problems: 1, 1, 1, 1, 1
Total: 5

$$= \lim_{\lambda\to+0} (V\Sigma^TU^TU\Sigma V^T+\lambda I)^{-1} V\Sigma^TU^T = \begin{bmatrix} UU^T = U^TU = I \\ VV^T = V^TV = I \end{bmatrix} =$$

$$= \lim_{\lambda\to+0} (V\Sigma^T\Sigma V^T+\lambda I)^{-1} V\Sigma^TU^T = \lim_{\lambda\to+0} (V\Sigma^T\Sigma V^T + V\lambda V^T)^{-1} V\Sigma^TU^T$$

$$= \lim_{\lambda\to+0} \left[V(\Sigma^T\Sigma+\lambda I)V^T\right]^{-1} V\Sigma^TU^T = \begin{bmatrix} (AB)^{-1} = B^{-1}A^{-1} \\ V^T = V^{-1}; V = V^{-T} \end{bmatrix} = \lim_{\lambda\to+0} V(\Sigma^T\Sigma+\lambda I)^{-1}V^TV\Sigma^TU^T$$

$$= V\left[\lim_{\lambda\to+\lambda} (\Sigma^T\Sigma+\lambda I)^{-1}\Sigma^T\right]U^T$$

The matrix $\Sigma^T\Sigma$ will always be diagonal since if $A\in\mathbb{R}^{m\times n} \Rightarrow \Sigma\in\mathbb{R}^{m\times n} \Rightarrow \Sigma^T\Sigma\in\mathbb{R}^{n\times n}$
this implies that:

$$\Sigma^T\Sigma+\lambda I = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}_{n\times n} + \lambda \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{pmatrix}_{n\times n} = \begin{pmatrix} \sigma_1^2+\lambda & \cdots & & 0 \\ 0 & \sigma_2^2+\lambda & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & & \cdots & \sigma_n^2+\lambda \end{pmatrix}_{n\times n}$$

$$\Rightarrow (\Sigma^T\Sigma+\lambda I)^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2+\lambda} & \cdots & & 0 \\ 0 & \frac{1}{\sigma_2^2+\lambda} & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & & \frac{1}{\sigma_n^2+\lambda} \end{pmatrix}_{n\times n} \Rightarrow \lim_{\lambda\to+0} \begin{pmatrix} \frac{1}{\sigma_1^2+\lambda} & \cdots & 0 \\ \vdots & & \\ 0 & \cdots & \frac{1}{\sigma_n^2+\lambda} \end{pmatrix}_{n\times n} = \begin{pmatrix} \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n^2} \end{pmatrix}_{n\times n} \equiv L$$

$$\Rightarrow L\Sigma^T \quad \begin{pmatrix} \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & & \\ 0 & \cdots & \frac{1}{\sigma_n^2} \end{pmatrix}_{n\times n} \begin{pmatrix} \sigma_1 & \cdots & & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & & & \vdots \\ & & & \sigma_m \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{n\times m} = \begin{pmatrix} \frac{1}{\sigma_1} & \cdots & 0 & \cdots & 0 \\ 0 & & 0 & & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n} & \cdots & 0 \end{pmatrix}_{n\times m} = \Sigma^+$$

finally:

$$A^\dagger = V\Sigma^+U^T = A^+$$

Note: When the matrix A is not full rank, then the terms of $\Sigma^T\Sigma+\lambda I$ for σ_{r+1}; $r = \operatorname{rank}(A)$ will be $1/\lambda$, this ensures the term $(\Sigma^T\Sigma+\lambda I)$ is always invertible, but this calculation of the pseudoinverse might be not convenient for very small singular values. ∎
The same proof can be done when $n>m$.

(2) $L(\beta) = \frac{1}{n}\|y - X\beta\|^2 + \lambda\|\beta\|^2 \Rightarrow \frac{\partial L}{\partial \beta} = -\frac{2}{n} X^T(y - X\beta) + 2\lambda\beta$

Solving for β:

$$-\frac{2}{n}X^Ty + \frac{2}{n}X^TX\beta + 2\lambda\beta = 0 \Rightarrow \cancel{\frac{2}{n}}(X^TX + \lambda n I)\beta = \cancel{\frac{2}{n}}X^Ty$$

$$\Rightarrow \hat{\beta}_\lambda^{ridge} = (X^TX + n\lambda I)^{-1}X^Ty$$

taking the SVD of $X = U\Sigma V^T$:

$$\hat{\beta}_\lambda^{ridge} = ((U\Sigma V^T)^T U\Sigma V^T + n\lambda I)^{-1}(U\Sigma V^T)^T y = (V\Sigma^T U^T U \Sigma V^T + n\lambda I)^{-1} V\Sigma^T U^T y =$$

$$= (V\Sigma^T\Sigma V^T + n\lambda I)^{-1} V\Sigma^T U^T y = V(\Sigma^T\Sigma + n\lambda I)^{-1} V^T V \Sigma^T U^T y =$$

$$= V(\Sigma^T\Sigma + n\lambda I)^{-1}\Sigma^T U^T y$$

taking the limit when $\lambda \to +0$ and assuming $n\lambda$ also tends to $+0$.

$$\lim_{\lambda \to +0} \hat{\beta}_\lambda^{ridge} = V\underbrace{\left(\lim_{\lambda \to +0} (\Sigma^T\Sigma + n\lambda I)^{-1}\Sigma^T\right)}_{\Sigma^+} U^T y = V\Sigma^+ U^T y = X^+ y \qquad (1)$$

On the other hand:

$$\hat{\beta}^{ls} = (X^TX)^+X^Ty = X^+(X^T)^+X^Ty = X^+X^{-T}(X^T(X^T)^+X^T)y = X^+X^{-T}X^Ty = X^+y \qquad (2)$$

finally by (1) and (2)

$$\hat{\beta}^{ls} = (X^TX)^+X^Ty = X^+y = \lim_{\lambda \to +0} \hat{\beta}_\lambda^{ridge}$$

∎

③ Prove that:

$$\|\tfrac{1}{n}X^TX - I_d\|_\infty \leq \mathcal{E} \Rightarrow \quad (1-\mathcal{E}s)\|\beta\|^2 \leq \tfrac{1}{n}\|X\beta\|^2 \leq (1+\mathcal{E}s)\|\beta\|^2; \quad \forall \beta \in \mathbb{R}^d \text{ s.t. } \|\beta\|_0 \leq s$$

$$s \in \{1, 2, \dots, [1/\mathcal{E}]\}$$

The expresion $\frac{1}{n}\|X\beta\|^2$ can be written as:

$$\frac{1}{n}\|X\beta\|^2 = \frac{1}{n}(X\beta)^T X\beta = \frac{1}{n}\beta^T X^T X\beta + \beta^T\beta - \beta^T\beta = \beta^T\left(\frac{X^TX}{n} - I\right)\beta + \underbrace{\beta^T\beta}_{\|\beta\|^2}$$

On the other hand, the expresion

$$\left|\beta^T\left(\frac{X^TX}{n} - I\right)\beta\right| \leq \left[|\mu^T M v| \leq \|M\|_\infty \|\mu\|_1 \|v\|_1\right] \leq \left\|\frac{X^TX}{n} - I\right\|_\infty \|\beta^T\|_1 \|\beta\|_1 \leq$$

$$\leq \mathcal{E}\|\beta\|_1^2 \leq \left[\|\beta\|_1 \leq \sqrt{s}\,\|\beta\|\right] \leq \mathcal{E}s\|\beta\|^2$$

since $supp(\beta) = \{1, 2, \dots, [1/\mathcal{E}]\}$

$$\Rightarrow \quad -\mathcal{E}s\|\beta\|^2 \leq \underbrace{\beta^T\left(\frac{X^TX}{n} - I\right)\beta}_{= \frac{1}{n}\|X\beta\|^2 - \|\beta\|^2} \leq \mathcal{E}s\|\beta\|^2$$

$$\Rightarrow \quad (1-\mathcal{E}s)\|\beta\|^2 \leq \frac{1}{n}\|X\beta\|^2 \leq (1+\mathcal{E}s)\|\beta\|^2 \qquad \blacksquare$$

④ In the case where $\frac{X^TX}{n} = I_d$ the problem can be reduced to:

$$\frac{1}{n}\|y - X\beta\|^2 = \frac{1}{n}(y - X\beta)^T(y - X\beta) = \frac{1}{n}(y^T - \beta^T X^T)(y - X\beta) = (y^Ty - y^TX\beta - \beta^TX^Ty + \beta^T\underbrace{X^TX}_{nI_d}\beta)\frac{1}{n}$$

$$= \frac{1}{n}y^Ty - \frac{1}{n}y^TX\beta - \frac{1}{n}\beta^TX^Ty + \beta^T\beta + \frac{1}{n^2}y^TXX^Ty - \frac{1}{n^2}y^TXX^Ty$$

$$= \left(\frac{1}{n^2}y^TXX^Ty - \frac{1}{n}y^TX\beta - \frac{1}{n}\beta^TX^Ty + \beta^T\beta\right) + \frac{1}{n}y^Ty - \frac{1}{n^2}y^TXX^Ty$$

$$= \left(\frac{1}{n}y^TX\left(\frac{X^Ty}{n} - \beta\right) - \beta^T\left(\frac{X^Ty}{n} - \beta\right)\right) + \frac{1}{n}y^T\left(I - \frac{XX^T}{n}\right)y$$

$$= \left(\frac{1}{n}y^TX - \beta^T\right)\left(\frac{X^Ty}{n} - \beta\right) + \frac{1}{n}y^T\left(I - \frac{XX^T}{n}\right)y \Rightarrow \frac{1}{n}\|y - X\beta\|^2 = \left(\frac{X^Ty}{n} - \beta\right)^T\left(\frac{X^Ty}{n} - \beta\right) + \frac{1}{n}y^T\left(I - \frac{XX^T}{n}\right)y$$

Minimizing this expression separately:

$$L = \left\|\frac{X^Ty}{n} - \beta\right\|^2 + \frac{1}{n}y^T\left(I - \frac{XX^T}{n}\right)y \Rightarrow \frac{\partial L}{\partial \beta} = 2\left(\beta - \frac{X^Ty}{n}\right) = 0 \Rightarrow \hat{\beta} = \frac{X^Ty}{n} \qquad (1)$$

As the value of L does not depend on β for the second term, the minimization problem can be rewritten as:

$$\underset{\beta}{\operatorname{argmin}} \left\|\frac{X^Ty}{n} - \beta\right\|^2 + \lambda^2\|\beta\|_0 = \underset{\beta}{\operatorname{arg\,min}} \left\|\frac{X^Ty}{n} - \beta\right\|^2 + \lambda^2\sum_{i=1}^{m} \mathbb{1}\{\beta_i \neq 0\}$$

The terms of the vector $\frac{X^Ty}{n} - \beta$ are as follows $\frac{1}{n}\left(\sum_{j=1}^{n} x_{jk}y_j\right) - \beta_k$ for the k-term, setting $\beta_k = 0$ will lead to: its contribution to the norm to be:

$$\left(\frac{1}{n}\left(\sum_{j=1}^{d} x_{jk}y_j\right)\right)^2 = \frac{1}{n^2}\left(\sum_{j=1}^{d} x_{jk}y_j\right)^2 = \frac{1}{n^2}(X_k^Ty)^2$$

Setting this β_k to 0 will be effective only if $\frac{1}{n^2}(X_k^Ty)^2 \leq \lambda^2 \Rightarrow |X_k^Ty| \leq n\lambda \qquad (2)$

then, by (1) and (2) it's possible to find the closed form solution:

$$\hat{\beta}_k^{bic} = \begin{cases} 0 & ; \ |X_k^Ty| \leq n\lambda \\ \frac{1}{n}X_k^Ty & ; \ \text{otherwise} \end{cases}$$

∎

(5) It's necessary to show that:

$$\min_{\substack{J\subseteq\{1,..,d\}\\ |J|\leq s}} RE(0,J)=\lambda>0 \Rightarrow (1-\delta)\|\beta\|^2 \leq \frac{\|X\beta\|^2}{n} \leq (1+\delta)\|\beta\|^2$$

1st approach

It's possible to write:

$$\frac{\|X\beta\|^2}{n} = \frac{1}{n}\|X\beta_J+X\beta_{J^c}\|^2 = \frac{1}{n}\|X\beta_J\|^2+\frac{1}{n}\|X\beta_{J^c}\|^2+\frac{2}{n}(X\beta_J)^T(X\beta_{J^c}) \leq$$

$$\leq \frac{1}{n}\|X\beta_J\|^2+\frac{1}{n}\|X\beta_{J^c}\|^2+\frac{2}{n}|(X\beta_J)^T(X\beta_{J^c})| \qquad (1)$$

from here:

$$\frac{\|X\beta\|^2}{n} \geq \frac{1}{n}\|X\beta_J\|^2-\frac{2}{n}|(X\beta_J)^T(X\beta_{J^c})| = \frac{1}{n}\|X\beta_J\|^2-\frac{2}{n}|\beta_J^TX^TX\beta_{J^c}| \geq$$

$$\geq \left[\frac{1}{n}\frac{\|X\beta_J\|^2}{\|\beta_J\|^2}\geq K>0\right] \geq K\|\beta_J\|^2-\frac{2}{n}|\beta_J^TX^TX\beta_{J^c}| \qquad (2)$$

the term:

$$|\beta_J^TX^TX\beta_{J^c}| \leq \|X^TX\|_\infty\|\beta_J^T\|_1\|\beta_{J^c}\|_1 \leq \left[\|\beta_J^T\|_1=\|\beta_J\|_1\leq c\|\beta_{J^c}\|_1\right] \leq$$

$$\leq \|X^TX\|_\infty c\|\beta_{J^c}\|_1^2 \leq \left[\|\beta_{J^c}\|_1\leq\sqrt{|J^c|}\|\beta_{J^c}\|\right] \leq$$

$$\leq \underbrace{\|X^TX\|_\infty}_{a} c|J^c|\|\beta_{J^c}\|^2 = ac|J^c|\|\beta_{J^c}\|^2$$

Taking the limit when $c\to0$ implies $X^TX\approx I$ which is equivalent to the ε-incoherence condition to hold for a very small ε, in other words

$$\lim_{c\to0} ac|J^c|\|\beta_{J^c}\|^2=0 \Rightarrow |\beta_J^TX^TX\beta_{J^c}|=0 \Rightarrow X^TX=I \text{ since this leads to } \beta_J^T\beta_{J^c}=0$$

Replacing in (1):

$$\frac{\|X\beta\|^2}{n\|\beta_J\|^2} \geq K>0$$

and because of (1) when $c\to0$

$$\frac{\|X\beta\|^2}{n\|\beta_J\|^2} < \frac{1}{n}\frac{\|X\beta_J\|^2}{\|\beta_J\|^2}+\frac{1}{n}\frac{\|X\beta_{J^c}\|^2}{\|\beta_J\|^2} \leq \left[\frac{1}{\|\beta_J\|^2}\leq\frac{|J|}{\|\beta_{J^c}\|_1}\leq\frac{|J^c|c}{\|\beta_J\|_1}\right] \leq K+\frac{1}{n}|J^c|c\frac{\|X\beta_{J^c}\|^2}{\|\beta_J\|_1} =$$

$$= K \text{ when } c\to0 \Rightarrow \frac{\|X\beta\|^2}{n\|\beta_J\|^2} < K \text{ (it's a contradiction)}$$

Then $\frac{\|X\beta\|^2}{n\|\beta_J\|^2}$ does not meet the RIP ■

2nd attempt

It's necessary to prove that

$$\text{If} \min_{\substack{J\subseteq\{1,\dots,d\}\\|J|\leq s}} RE(0,J)=\lambda>0 \quad \Rightarrow \quad (1-\delta)\|\beta\|^2\leq \frac{\|X\beta\|^2}{n}\leq(1+\delta)\|\beta\|^2 \quad (RIC)$$

for $\delta\in(0,1)$ and $\|\beta\|_0\leq s$

In this case I am assuming that the minimization RE condition also holds for the eigenvalues of $\frac{X^TX}{n}$ $\Rightarrow \lambda_{min}\left(\frac{X^TX}{n}\right)=\frac{\|X\beta_J\|^2}{n\|\beta_J\|^2}=K>0$ for the cone $\|\beta_J\|_1\leq c\|\beta_{J^c}\|_1$

then:

$$\frac{\|X\beta_J\|^2}{n\|\beta_J\|^2}\leq\frac{\|\beta_J\|^2+\left|\beta_J^T\left(\frac{X^TX}{n}-I\right)\beta_J\right|}{\|\beta_J\|^2} \quad (1)$$

Considering the inequality:

$$\left|\beta_J^T\left(\frac{X^TX}{n}-J\right)\beta_J\right|\leq\underbrace{\left\|\frac{X^TX}{n}-I\right\|_\infty}_{a}\|\beta_J^T\|_1\|\beta_J\|_1=a\|\beta_J\|_1^2\leq a\,c\|\beta_{J^c}\|_1^2\leq a\,c\,|J^c|\|\beta_{J^c}\|^2$$

Then taking the limit at the expression on the right it gives 0 when $c\to0$, this implies that the matrix $\frac{X^TX}{n}=I$; meaning that the matrix $\frac{X^TX}{n}$ is orthonormal.

So if $c\to0$, it does not matter if the matrix X is ε-incoherent, it's possible to bound the eigenvalues of: X^TX by:

$$K<\frac{\|X\beta_J\|^2}{n\|\beta_J\|^2}\leq1 \quad \text{from (1) when } c\to0;\ K=\frac{\|X\beta_J\|^2}{n\|\beta_J\|^2}$$

which mean that it does not follow the RIC property ∎

In the case when $c>0$ and the restricted eigenvalue condition holds, it's possible to ensure ε-incoherence by random sampling of X from the normal distribution and getting:

$$\left\|\frac{1}{n}X^TX-I\right\|\leq c\sqrt{\frac{s}{n}} \quad ;\ X\overset{i.i.d}{\sim}N(0,1)$$

then this ensures RIC. for $\delta=cs\sqrt{\frac{s}{n}}$