CMSC 25025 Machine Learning and Large Scale Data Analysis

Homework #2

Xinyu Wei

1. PCA

$\min_{u, \{\lambda_i\}} \sum_{i=1}^{n} \|x_i - u - V_k\lambda_i\|^2$ V_k is d×k orthogonal

First fix u, let $\tilde{x}_i = x_i - u$. Now $\min_{\{\lambda_i\}} \sum_{i=1}^{n} \|\tilde{x}_i - V_k\lambda_i\|^2$

Claim: the minimizer is $\hat{\lambda}_i = V_k^T \tilde{x}_i$

Proof: $\|b - Ax\|^2$ is minimized when $A^TAx = A^Tb$ $b \in \mathbb{R}^k$. $A \in \mathbb{R}^{k\times d}$. $x \in \mathbb{R}^d$

Let $\mathbb{R}(A) = \{y \in \mathbb{R}^k : y = Ax, x \in \mathbb{R}^d\}$.

$\mathbb{R}(A)^\perp = \{x \in \mathbb{R}^d : Ax = 0\}$

Then $\mathbb{R}^k = \mathbb{R}(A) \oplus \mathbb{R}(A)^\perp$.

$b \in \mathbb{R}^k$ can be written as $b = b_1 + b_2$, where $b_1 \in \mathbb{R}(A)$, $b_2 \in \mathbb{R}(A)^\perp$

$b_1 - Ax \in \mathbb{R}(A)$.

$$\|b - Ax\|^2 = \|b_1 - Ax + b_2\|^2 = \|b_1 - Ax\|^2 + \|b_2\|^2 + \underbrace{(b_1 - Ax)^T b_2}_{=0} + \underbrace{b_2^T(b_1 - Ax)}_{=0}$$
$$= \|b_1 - Ax\|^2 + \|b_2\|^2$$

$\|b - Ax\|^2$ is minimized when $\|b_1 - Ax\|^2$ is minimized.

$b_1 = Ax$ has a solution

$x = A^Tb_1 = A^Tb_1 + A^Tb_2$ since $A^Tb_2 = 0$.

$= A^Tb$.

Thus, $A^TAx = A^Tb$.

$x = A^Tb$ is the minimizer of $\|b - Ax\|^2$

Going back to our problem, $\hat{\lambda}_i = V_k^T\tilde{x}_i$ is the minimizer of $\|\tilde{x}_i - V_k\lambda_i\|^2$,

thus the minimizer of $\sum_{i=1}^{n} \|\tilde{x}_i - V_k\lambda_i\|^2$.

Then we want to minimize $f(u) = \sum_{i=1}^{n} \|x_i - u - V_kV_k^T(x_i - u)\|^2 = \sum_{i=1}^{n} \|(I - V_kV_k^T)(x_i - u)\|^2$

$Df(u) = \sum_{i=1}^{n} 2(I - V_kV_k^T)(x_i - u) \cdot (I - V_kV_k^T) \cdot (-1) = 2(I - V_kV_k^T)^2 \cdot (nu - \sum_{i=1}^{n} x_i)$

$Df(u) = 0$ when $nu = \sum_{i=1}^{n} x_i$,

$\Rightarrow \hat{u} = \frac{1}{n}\sum_{i=1}^{n} x_i$

So $\hat{u} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\hat{\lambda}_i = V_k^T(x_i - \hat{u})$ are the minimizers. $\hat{u}$ is not unique because as long as it satisfies $(I - V_kV_k^T)^2 \cdot (u - \bar{x}) = 0$, the lost is minimized.

2. Lasso.

$f(\beta) = -t\beta + \frac{1}{2}\beta^2 + \lambda|\beta|$

WTS: the minimizer of $f(\beta)$ is $\beta = \text{sign}(t)\,[|t|-\lambda]_+$

Break it into the following 4 cases:

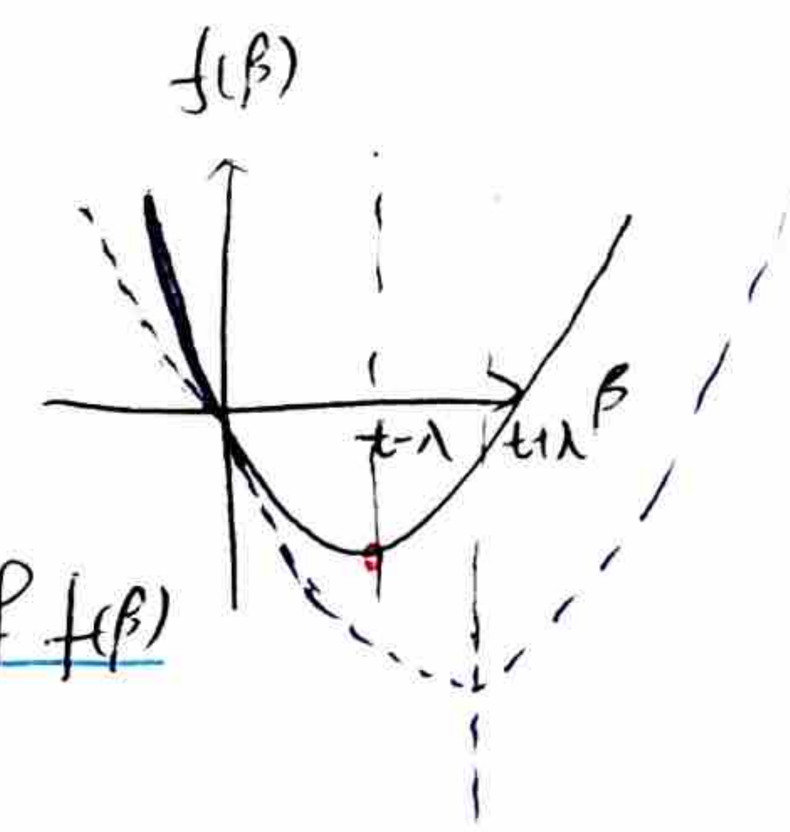
① $t>0$, $|t|-\lambda>0$

$$f(\beta) = \begin{cases} \frac{1}{2}\beta^2 + (\lambda - t)\beta & \beta>0 \\ \frac{1}{2}\beta^2 - (\lambda+t)\beta & \beta<0 \end{cases}$$

when $\beta>0$, $\beta_{*+} = -\frac{\lambda-t}{\frac{1}{2}\times 2} = t-\lambda>0$.

when $\beta<0$, $\beta_{*-} = -\frac{-(\lambda+t)}{\frac{1}{2}\times 2} = \lambda+t$

<1>. if $\lambda+t>0$



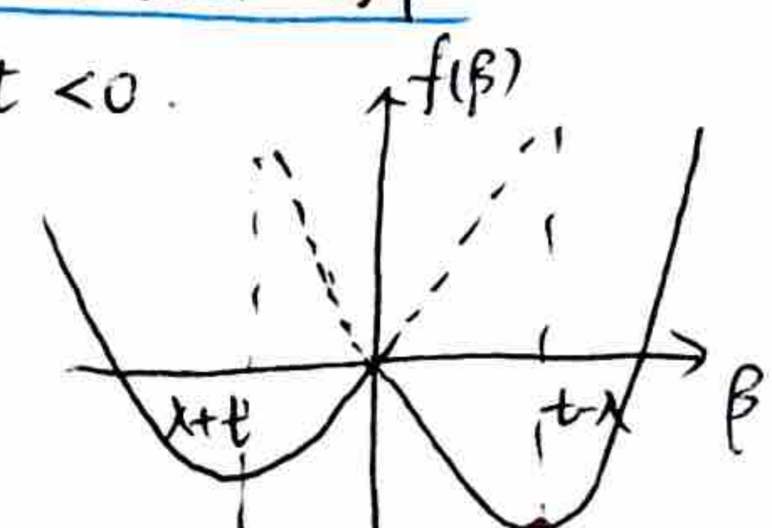
the minimizer of $f(\beta)$ is $t-\lambda$.

$= |t|-\lambda$

$= +[|t|-\lambda]_+$

$= \text{sign}(t)\,[|t|-\lambda]_+$

<2>. if $\lambda+t<0$.



$f(\beta_{*+}) = \frac{1}{2}\beta^2 + (\lambda-t)(t-\lambda)$

$= \frac{1}{2}\beta^2 - (\lambda-t)^2$

$f(\beta_{*-}) = \frac{1}{2}\beta - (\lambda+t)^2$

$f(\beta_{*+}) < f(\beta_{*-})$ because $(\lambda+t)^2 < (\lambda-t)^2$

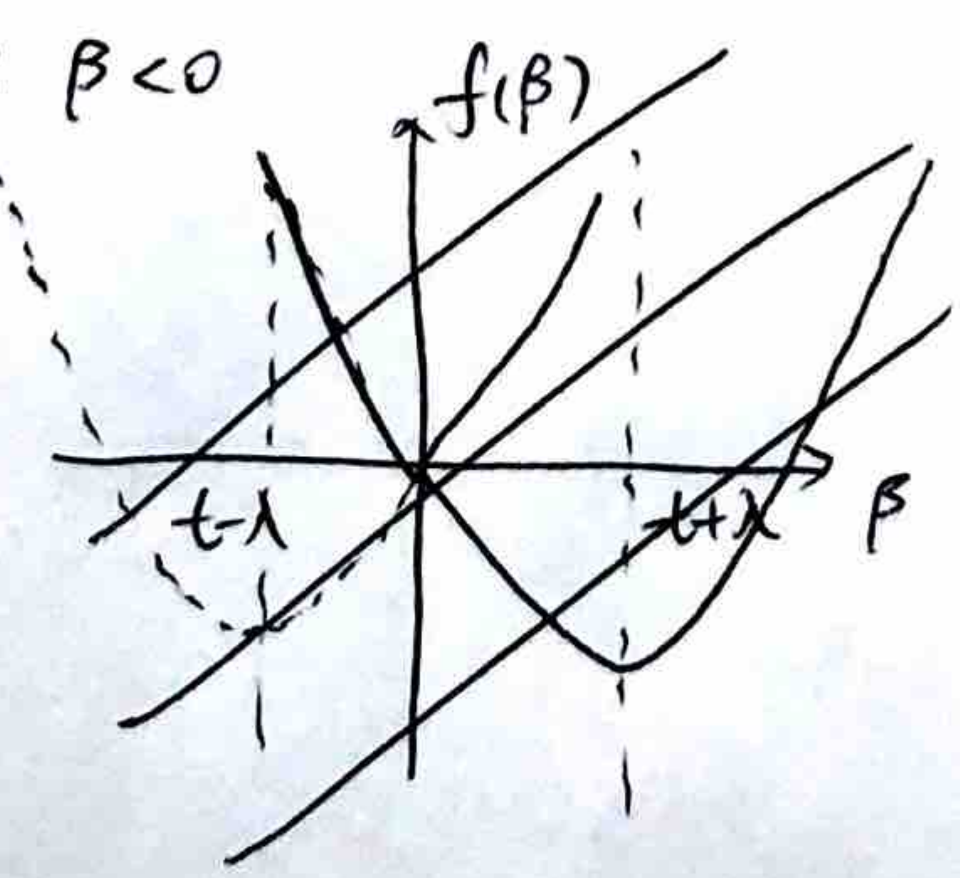
So the minimizer is $\beta_{*+} = t-\lambda = \text{sign}(t)[|t|-\lambda]_+$

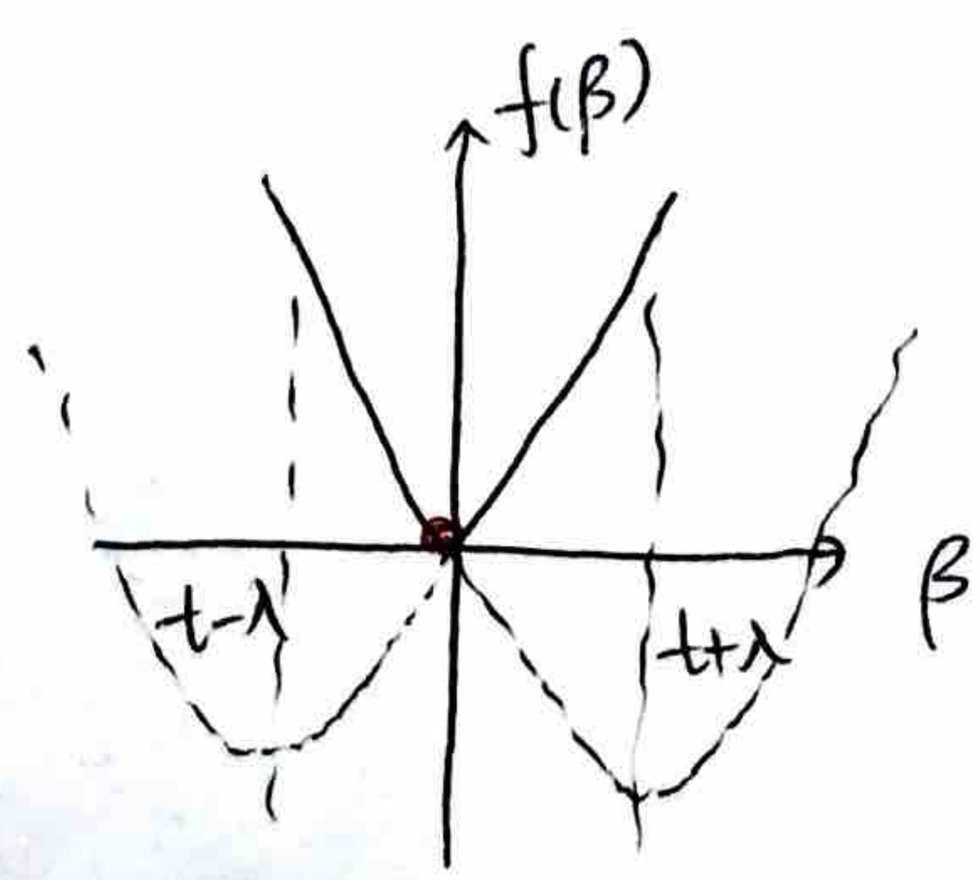
② $t>0$, $|t|-\lambda<0$

$$f(\beta) = \begin{cases} \frac{1}{2}\beta^2 + (\lambda-t)\beta & \beta>0 \\ \frac{1}{2}\beta^2 - (\lambda+t)\beta & \beta<0 \end{cases}$$

$\beta_{*+} = t-\lambda<0$

$\beta_{*-} = t+\lambda>0$





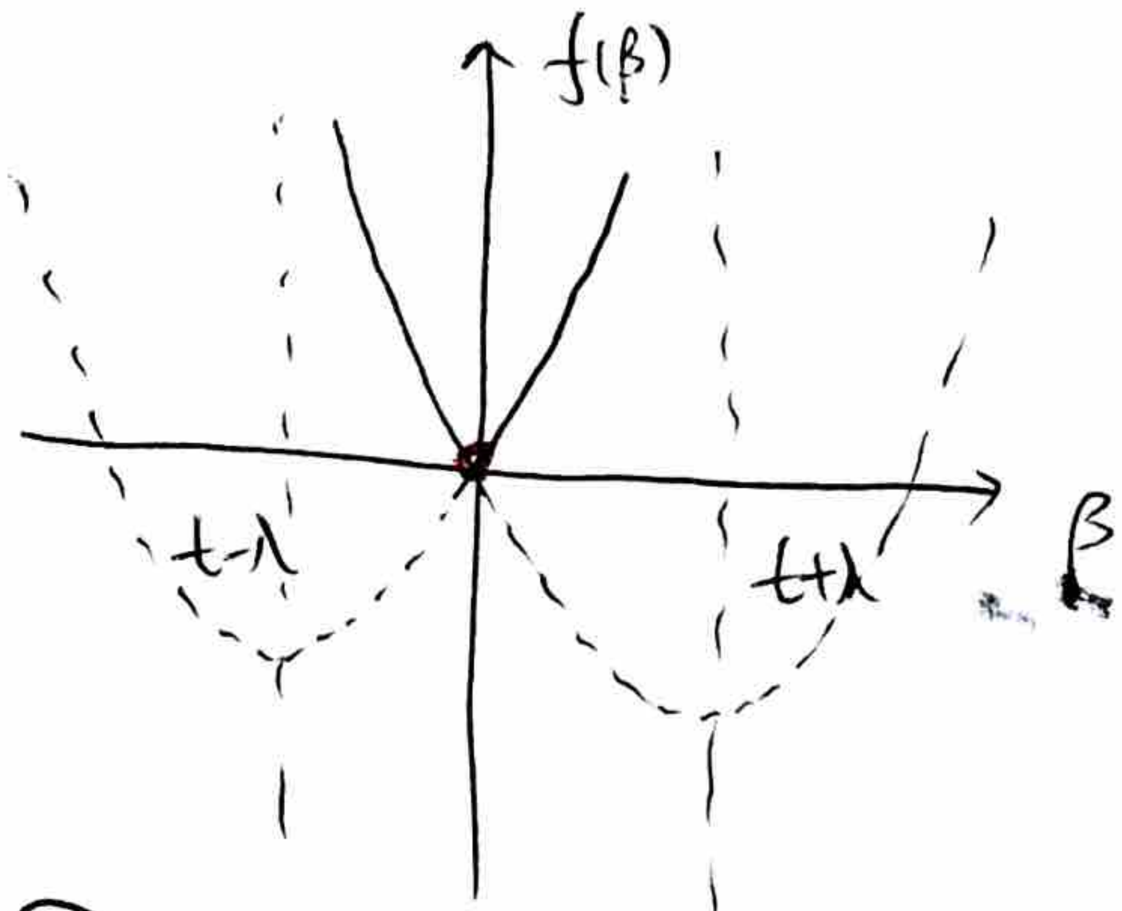
the minimizer is $\beta=0 = +0 = +[|t|-\lambda]_+ = \text{sign}(t)\,[|t|-\lambda]_+$

③. $t<0, \quad |t|-\lambda<0$

$$f(\beta)=\begin{cases}\frac{1}{2}\beta^2+(\lambda-t)\beta & \beta>0\\ \frac{1}{2}\beta^2-(\lambda+t)\beta & \beta<0\end{cases}$$

$\beta_{s+}=t-\lambda<0$ because $\left.\begin{matrix}|t|-\lambda<0\\ t<0\end{matrix}\right\} \Rightarrow -\lambda-t<0 \Rightarrow -(\lambda+t)<0 \Rightarrow \left.\begin{matrix}\lambda+t>0\\ t<0\end{matrix}\right\} \Rightarrow \lambda>0 \Rightarrow t-\lambda<0$

$\beta_{s-}=t+\lambda>0$



the minimizer is $\beta=0=-0=-[|t|-\lambda]_+$
$=\text{sign}(t)[|t|-\lambda]_+$

④. $t<0, \quad |t|-\lambda>0$

$$f(\beta)=\begin{cases}\frac{1}{2}\beta^2+(\lambda-t)\beta & \beta>0\\ \frac{1}{2}\beta^2-(\lambda+t)\beta & \beta<0\end{cases}$$

$\beta_{s-}=t+\lambda<0$ because $\left.\begin{matrix}|t|-\lambda>0\\ t<0\end{matrix}\right\} \Rightarrow -t-\lambda>0 \Rightarrow t+\lambda<0$

$\beta_{s+}=t-\lambda$

<1>. if $t-\lambda<0, \quad \lambda>0$

the minimizer is

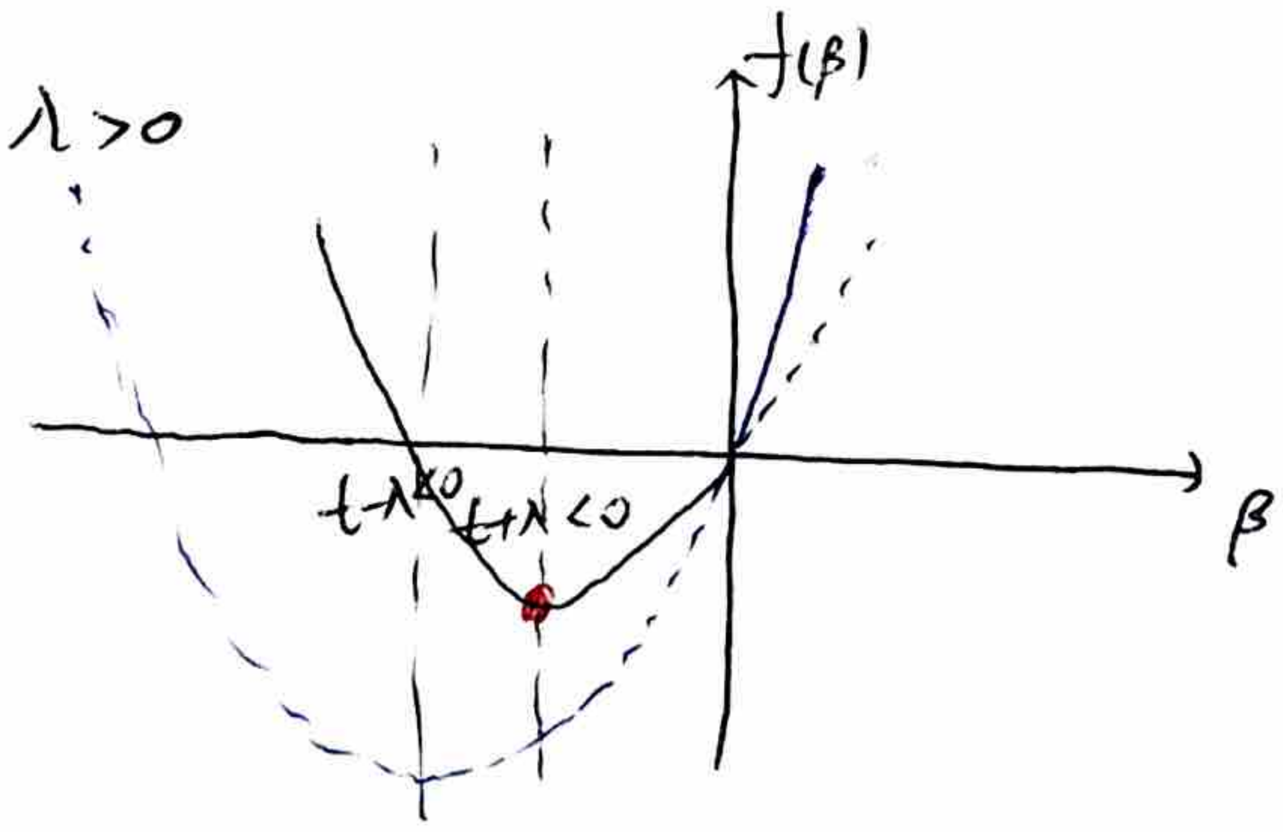
$\beta_{s-}=t+\lambda$

$=-|t|+\lambda$

$=-(|t|-\lambda)$

$=-[|t|-\lambda]_+$

$=\text{sign}(t)[|t|-\lambda]_+$



<2>. if $t-\lambda>0, \quad \lambda<0$.

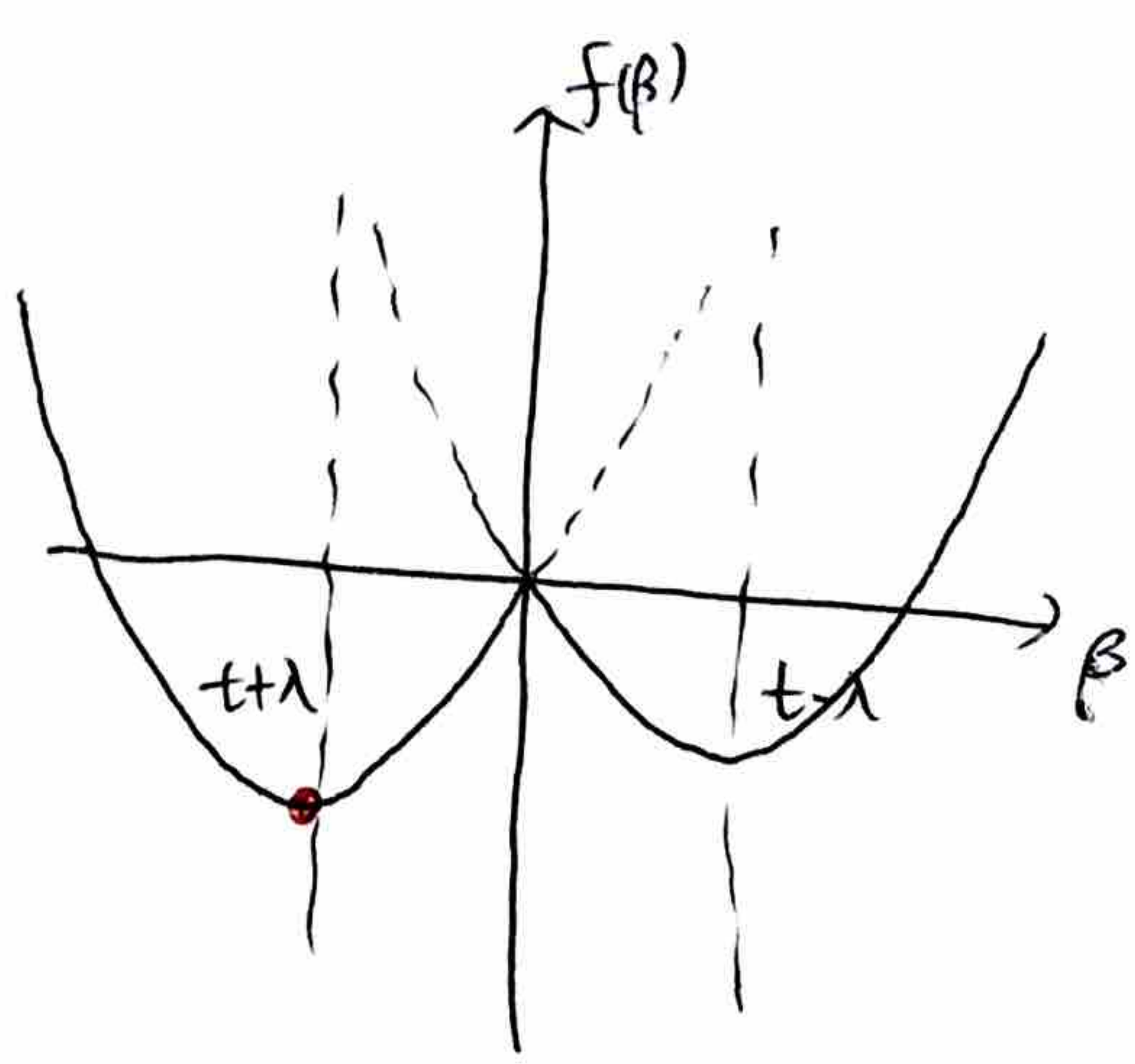
$f(\beta_{s+})=\frac{1}{2}\beta^2-(t-\lambda)^2$

$f(\beta_{s-})=\frac{1}{2}\beta^2-(\lambda+t)^2$

Since $t<0, \lambda<0$, $|\lambda+t|>|t-\lambda|$

$(\lambda+t)^2>(t-\lambda)^2$

$f(\beta_{s-})<f(\beta_{s+})$



So the minimizer is $\beta_{s-}=t+\lambda=\text{sign}(t)[|t|-\lambda]_+$

Based on ①,②,③,④, the minimizer of $f(\beta)$ is $\beta=\text{sign}(t)[|t|-\lambda]_+$