Log-Pseudolikelihood function: $\ell(\beta) = \sum_{i=1}^{n} \{ z_i [X_i^T\beta + \eta A_i^T \underline{z}] - \log[1+\exp(X_i^T\beta + \eta A_i^T \underline{z})] \}$

where $(\underline{z})_{n\times1}$ is a vector of binary responses,

$(X_i)_{(p+1)\times1}$ is a vector of covariates for the ith observation, $i=1,\dots,n$

$(A_i)_{n\times1}$ is a vector of 0s and 1s where $A_{ij} = \begin{cases} 1, & \text{if obs. } i \text{ and } j \text{ are neighbors} \\ 0, & \text{o.w.} \end{cases}$

$(\beta)_{(p+1)\times1}$ is a vector of regression parameters $(\beta = (\beta_0, \beta_1, \dots, \beta_p))$

$(\eta)_{1\times1}$ is a scalar parameter for measuring spatial dependence

Let $\beta_{1:p}$ denote the parameter vector $(\beta_1, \dots, \beta_p)$.

Then let $d_\beta(\beta_0, \beta_{1:p}, \eta) = \frac{\partial \ell(\beta)}{\partial \beta_{1:p}}$ and $H_\beta(\beta_0, \beta_{1:p}, \eta) = \frac{\partial \ell^2(\beta)}{\partial \beta_{1:p} \partial \beta_{1:p}^T}$.

In the 2010 JRSSB paper,

$$y^* = (B^{-1})^T \{ d_\beta(\hat{\beta}_0, \hat{\beta}_{1:p}^{(m-1)}, \hat{\eta}) + H_\beta(\hat{\beta}_0, \hat{\beta}_{1:p}^{(m-1)}, \hat{\eta}) \hat{\beta}_{1:p}^{(m-1)}$$

$$X^* = B \operatorname{diag}(\{\lambda_j^{-1}\}_{j=1}^p)$$

$$\beta^* = \operatorname{diag}(\{\lambda_j\}_{j=1}^p) \beta_{1:p}$$

$$I(\hat{\beta}^{(m-1)}) = H_\beta(\hat{\beta}_0, \hat{\beta}_{1:p}^{(m-1)}, \hat{\eta}) = B^T B$$

In my code, "mat.Hbeta" is H_β, "mat.A" is B, "mat.X.beta" is X^*, "beta.d1" is d_β, and "vec.y.beta" is y^*.

* In other words, we need to compute the gradient vector and Hessian matrix for the likelihood (as in the Zhu paper) and substitute the quantities above into the LARS function.

* Note that $\hat{\beta}_0$ and $\hat{\eta}$ are estimated as the maximum pseudolikelihood estimators at the beginning and fixed.

* This is only for the uncentered model. The centered and spatial-temporal models work the same way except the likelihoods are different.

Centered model: terms with $X_i^T\beta + \eta A_i^T \underset{\sim}{Z}$ are replaced by $X_i^T\beta + \eta A_i^T(\underset{\sim}{Z} - \underset{\sim}{\mu})$ where

$(\underset{\sim}{\mu})_{n\times 1}$ is a vector of the "independence" means, where

$$\mu_i = \frac{\exp(X_i^T\beta)}{1+\exp(X_i^T\beta)}$$

Spatial-temporal model: terms with $X_i^T\beta + \eta A_i^T \underset{\sim}{Z}$ are replaced by $X_{it}^T\beta + \eta A_i^T \underset{\sim}{Z}_t + \sum_{s=1}^{S} \tau_s \underset{\sim}{Z}_{t-s}$,

where X_{it} is a vector of covariates for observation plot i at time t

$\underset{\sim}{Z}_t$ is the $n\times 1$ vector of binary responses at time t

τ_s is the autoregression coefficient for time lag s, $s=1,\dots,S$