Gulf Journal of Mathematics
Vol 24, Issue 1 (2026) 1-26

ISSN: 2309-4966
https://doi.org/10.56947/ftfh7e69

GROUP LASSO FOR ZERO-INFLATED BERNOULLI REGRESSION MODEL

MOUHAMED NDOYE[1] and ABA DIOP[2*]

Abstract. Group Lasso is an extension of the Lasso regularization method that selects variables in predefined groups within regression models. In this study, we adapt Group Lasso to the Bernoulli regression model with zero inflation and introduce an efficient algorithm designed for high-dimensional problems, solving the related convex optimization task. The resulting Group Lasso estimator is proven to be statistically consistent and asymptotically Gaussian, even when the number of predictors greatly exceeds the sample size, provided the true underlying structure is sparse.

Keywords. Logistic regression, Zero inflation, Group variable selection, Penalized likelihood.
Subjclass: 62H12, 62J07, 62J12

1. Introduction and preliminaries

Binary regression models and their parameter estimations have been widely investigated in the literature, including asymmetric frameworks for rare events [15]. In addition the Group Lasso for variable selection in a zero-inflated logistic regression model is a powerful method, particularly when variables are naturally grouped. The group Lasso see [28, 2, 3, 1] is a regularization technique used in statistical modeling and machine learning that extends the traditional Lasso method. That is, group Lasso addresses variable selection problems by introducing an appropriate extension of the Lasso penalty discussed in [3]. Mathematically, the group Lasso can be expressed as an optimization problem where the objective is to maximize the penalized log-likelihood function subject to a constraint on the sum of L2 norms. It is especially useful when dealing with high-dimensional data where the number of predictors exceeds the number of observations. The group Lasso method enables the simultaneous selection of entire groups of variables that are most relevant to each component of the regression model, while accounting for the underlying data structure. Recent developments in statistical regression frameworks have focused on alternative robust estimation techniques

Date: Received: Jul 17, 2026; Revised: Aug 14, 2026; Accepted: Aug 21, 2026.
* Corresponding author

under various dependence assumptions [6]. This makes it an effective tool for feature selection in scenarios where predictors are naturally grouped. It is particularly valuable when the number of variables is large and the goal is to avoid overfitting.

The Group Lasso operates by penalizing groups of coefficients in the zero-inflated logistic regression model, in a manner analogous to the Lasso, but at the group level rather than the individual-variable level: when a group of variables is judged irrelevant, every coefficient within that group is shrunk to zero simultaneously, effectively removing the whole group from the model. This group-wise selection is particularly advantageous in contexts such as genomics, where variables may be naturally grouped.

In this context, the group structure of variables is present, for example, in biology, where a group may consist of variables sharing the same biological or chemical property. This is also the case for categorical variables, where each is represented in the design matrix by a group of dummy variables. In the case of categorical variables, the Lasso estimator selects individual dummy variables rather than the group of indicators that is, the variable in its entirety. In such situations, it is more judicious to consider selecting (or rejecting) variables by groups. This group structure can be accounted for using the Group Lasso estimator introduced by [28] for the linear model, and by [16] for the logistic model without zero-inflation. Variable selection (i.e., extracting relevant features for a given task) or subset selection (i.e., model selection, which is nonetheless known for its instability [10]) plays an increasingly important role in many fields, including genetics ([13], astronomy [29], and economics [8]. It is a crucial step for building a robust and interpretable model.

This article addresses the group Lasso penalty for zero-inflated logistic regression models [20, 11] first studied the group Lasso for logistic regression models and subsequently proposed a gradient descent algorithm to solve the corresponding constrained problem. Here, we present variable selection methods in a zero-inflated logistic regression model that allow for working directly on the penalized problem and whose convergence property does not depend on unknown constants, as in [11]. Our algorithms are efficient in the sense that they can handle high-dimensional problems. It should be noted that high-dimensional predictor spaces are found in many applications, particularly with higher-order interaction terms or basis expansions for additive logistic models where groups correspond to the basis functions of individual continuous covariates. We do not aim for an (approximate) path-following algorithm [21, 30, 12], yet our approaches are fast enough to compute a whole range of solutions for different penalty parameters on a (fixed) grid. Our approach is similar to that of [9], which presents a remarkably fast ("the fastest") implementation for large-scale logistic regression with the Lasso. We also present an asymptotic consistency theory for the group Lasso in high-dimensional problems where the predictor dimension is much larger than the sample size. This is very useful when the number of variables is high and the goal is to avoid overfitting.

In this article, we present the ZIBer model penalized by the group Lasso. Our main contribution is the development of variable selection methods for over-dispersed binary data (it should be noted that the ZIBer model can incorporate additional over-dispersion) and zero-inflated data [20] through the group Lasso penalized ZIBer model. The remainder of the article is organized as follows: Section 2 addresses the variable selection problem by penalizing the log-likelihood function and discusses the underlying statistical inference using the group Lasso. In Section 3, the asymptotic properties of the proposed estimators are studied, and their performance in finite dimensions is evaluated through simulations. Finally, Section 4 provides an application using real data set.

2. Methodology

We consider here the case where the explanatory variables have a group structure that is known a priori, a structure we aim to incorporate into the estimation procedure. The group Lasso method for the ZIBer model considers groups of variables instead of individual variables. Let n be an independent and random sample, p the number of explanatory variables, q and the number of zero-inflation variables.

2.1. **Model.**

The ZIBer model with group Lasso is a sophisticated statistical approach that combines two main components (the logistic part and the zero-inflation part) to model data with zero inflation.
The Logistic component of the model uses logistic regression to predict the probability that an observation belongs to a specific category (for example, the presence or absence of an event). It is particularly useful for binary or categorical data.

$$p_i = \mathbb{P}\left[y_i = 1 | x_i\right] = \frac{e^{\beta^T x_i}}{1 + e^{\beta^T x_i}}, \quad \beta \in \mathbb{R}^p \tag{2.1}$$

The inflation component models the excess of zeros in the data — that is, the case where the observed number of zeros exceeds what a standard distribution would predict — by separating zeros arising from a structural absence of the event from zeros arising from the ordinary sampling distribution:

$$\pi_i = \mathbb{P}\left[y_i = 0 | z_i\right] = \frac{e^{\gamma^T z_i}}{1 + e^{\gamma^T z_i}} \quad \gamma \in \mathbb{R}^q \tag{2.2}$$

By combining these two components with a group Lasso penalty, the ZIBer model automatically identifies the most relevant variables while properly accounting for zero inflation. The two linear predictors are:
Regression component:

$$logit(p_i) = \beta^T x_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} \tag{2.3}$$

Inflation component:

$$logit(\pi_i) = \gamma^T z_i = \gamma_1 z_{i1} + \gamma_2 z_{i2} + ... + \gamma_q z_{iq} \tag{2.4}$$

2.2. **Group Lasso.** Let X and Z be design matrices structured into G_n groups and G'_n groups, respectively. Each group g in X has dfx_g degrees of freedom, and each group g' in Z has $dfz_{g'}$ degrees of freedom, where $g \in \{1, \dots, G_n\}$ and $g' \in \{1, \dots, G'_n\}$. We define $df_g = dfx_g + dfz_g$ as the total degree of freedom of the $g-th$ predictor for the pair (X_i^g, Z_i^g). Let $d_{max} := \max_{g\in\{1,\dots,G_n+G'_n\}} df_g$, and $d_{min} := \min_{g\in\{1,\dots,G_n+G'_n\}} df_g$ the corresponding minimum. We also have $p = \sum_{g=1}^{G_n} dfx_g$ and $q = \sum_{g=1}^{G'_n} dfz_g$. The parameter $\theta = (\beta^T, \gamma^T)^T$ (where $\beta = \left(\beta_1^T, \dots, \beta_{G_n}^T\right)$ where $\beta_g = \left(\beta_{g,1}, \dots, \beta_{g,dfx_g}\right) \in \mathbb{R}^{dfx_g}$ and $\gamma = \left(\gamma_1^T, \dots, \gamma_{G'_n}^T\right)$ where $\gamma_g = \left(\gamma_{g,1}, \dots, \gamma_{g,dfz_g}\right) \in \mathbb{R}^{dfz_g}$ so that $\theta_g \in \mathbb{R}^{dfx_g+dfz_g}$) represents the coefficients of the zero-inflated logistic model. For $\beta \in \mathbb{R}^p$ et $\gamma \in \mathbb{R}^q$, β_g and γ_g denote the sub-vectors of β and γ whose indices correspond to those of the g-th group in X and Z, respectively. These degrees of freedom represent the complexity or flexibility associated with each group within the matrices. For instance, in a statistical modeling or machine learning context, degrees of freedom can be used to measure a model's capacity to fit the data while avoiding overfitting.

Let $y_i \in \{0, 1\}$ be a binary response variable from independent and identically distributed (i.i.d.) observations. Let X_i and Z_i be $n \times dfx_i$ and $n \times dfz_i$ matrices, respectively, representing the columns of the design matrix corresponding to the i-th predictor. In other words, assuming the block matrices X_g (resp Z_g) are of full rank, we can perform block orthonormalization—for example, via QR decomposition to obtain $X_g^T X_g = I_{dfx_g}$ (resp $Z_{g'}^T Z_{g'} = I_{dfz_{g'}}$) for all $g = \{1, \dots, G_n\}$ (resp $g = \{1, \dots, G_n\}$). For $i = \{1, \dots, n\}$ $X_i = \left(X_i^1, \dots, X_i^{G_n}\right)$ where $X_i^g = \left(X_{i,1}^g, \dots, X_{i,dfx_g}^g\right)$ and $Z_i = \left(Z_i^1, \dots, Z_i^{G'_n}\right)$ où $Z_i^g = \left(Z_{i,1}^g, \dots, Z_{i,dfx_g}^g\right)$. This decomposition is often natural in biology and microarray data when the covariates represent gene expressions, for example, [22, 27]. We allow the number of groups to increase with the sample size n in order to accommodate cases where $G_n + G'_n$ exceeds n. By using such a design matrix, the group Lasso estimator does not depend on the dummy variable coding scheme. We chose a rescaled version, $X_g^T X_g = nI_{dfx_g}$ and $Z_g^T Z_g = nI_{dfz_g}$, to ensure that the parameter estimates remain on the same scale as the sample size n varies. After parameter estimation, the estimates must be transformed back to correspond to the original encoding. We study the estimation and prediction properties of the group Lasso in high-dimensional contexts where the number of groups exceeds the sample size: $G_n + G'_n \gg n$. Let $H^\star = \{g : \theta^{\star g} \neq 0\}$ denote the set of group indices for which the corresponding sub-vectors of $\theta^\star$ are non-zero, and let $s^\star := |H^\star|$. This set characterizes the sparsity of the model, and $H^{\star c}$ denotes its complement the set of group indices not included in $H^\star$. Both $H^\star$ and $s^\star$ depend on n, though we omit this dependence for notational simplicity. In general, it is impossible to estimate all unknown parameters from the data unless we assume that the true parameter is group-sparse;we therefore assume that $\theta^\star$ is partitioned according to the grouping of (X, Z), with only a few groups being relevant. That is $G_n + G'_n \gg s^\star$.

2.3. **Penalized log-likelihood function.** The complete-data log-likelihood function is given by:

$$\ell_n(\theta) = \sum_{i=1}^{n} \left[J_i log\left(\pi_i + (1-\pi_i)(1-p_i)\right) + (1-J_i) log\left((1-\pi_i)p_i\right)\right] \tag{2.5}$$

$$\begin{aligned} \ell_n(\theta) &= \sum_{i=1}^{n} \left[J_i log\left(e^{\gamma^T z_i} + (1+e^{\beta^T x_i})^{-1}\right) - log\left(1+e^{\gamma^T z_i}\right)\right] \\ &+ \left[(1-J_i)\left(y_i \beta^T x_i - log(1+e^{\beta^T x_i})\right)\right] := \sum_{i=1}^{n} \ell_i(\theta) \end{aligned} \tag{2.6}$$

where $J_i := 1_{\{y_i=0\}}$ and $1 - J_i := 1_{\{y_i=1\}}$.

2.4. **Group Lasso penalty function.** Let $\mathcal{P}(\theta, \lambda)$ be the group Lasso penalty function. It is defined as follows:

$$\mathcal{P}(\theta, \lambda) = \lambda \sum_{g=1}^{G} w_g \left\| (\beta_g, \gamma_g)^T \right\|_2, \tag{2.7}$$

where g indexes the groups of variables and $\lambda \geq 0$ is the regularization (tuning) parameter. The regularization parameter $\lambda \geq 0$ controls the level and strength of the penalty, thereby influencing the selection process, and w_g represents the weight adjusted for the size of group g. In order to independently select variables in each component, we apply a separate penalty:

$$\mathcal{P}(\theta, \lambda) = \lambda_1 \sum_{g=1}^{G_n} s(df x_g) \left\|\beta_g\right\|_2 + \lambda_2 \sum_{g=1}^{G'_n} s(df z_g) \left\|\gamma_g\right\|_2 \tag{2.8}$$

where, $\|\|_2$ refers to the Euclidean norm. We generally choose $s(t) = \sqrt{t} = w_g$, so as to penalize larger groups more heavily. We also assume there exists a constant $B > 0$ such that $\mathcal{P}(\theta, \lambda) \leq B$.

2.5. **Minimizing the negative log-likelihood for the group Lasso.** Our objective is to minimize the penalized negative log-likelihood function under the group Lasso, defined as:

$$\mathcal{S}_\lambda(\theta) = -\ell_n(\theta) + \mathcal{P}(\theta, \lambda) \tag{2.9}$$

$\mathcal{S}_\lambda(\theta)$ is a convex loss function; furthermore, we penalize the coefficients $\theta = (\beta, \gamma)$ using the group Lasso via:

$$\min_{\theta} \left[\mathcal{S}_\lambda(\theta)\right] \tag{2.10}$$

Thus, the group Lasso formulation can be represented as a minimization of $\mathcal{S}_\lambda(\theta)$. Consequently, the penalized negative log-likelihood estimator minimizes the negative log-likelihood function penalized by the group Lasso (for $\lambda = (\lambda_1, \lambda_2) > 0$). In other words, the group Lasso estimator, which achieves group-level sparsity, is obtained as the solution to the following convex optimization problem:

$$\hat{\theta}_\lambda = arg \min_{\theta} \left[\mathcal{S}_\lambda(\theta)\right] \tag{2.11}$$

Note that if all groups are of size 1, we recover the Lasso estimator. The group Lasso allows for the simultaneous selection and estimation of variables, much like the Lasso. Furthermore, the group Lasso estimator defined in (2.11) relies on the assumption that the groups form a partition. The group Lasso penalty combines an L_2 norm within each group with an L_1-type norm across groups, which induces selection at the group level rather than the individual coefficient level positioning the estimator between the Lasso and Ridge penalties. Concretely, the penalty term is the sum of the Euclidean norms of the coefficient groups, so that the estimator behaves like a Lasso at the group level [17, 14, 28], while remaining invariant to orthogonal transformations applied within each group, a property inherited from the Ridge penalty [28].

Note that we do not penalize the intercept. However, as shown by Lemma (2.1), the minimum of (2.9) is achieved. Let s be a function used to scale the penalty relative to the dimension of the parameter vector θ_g. Unless otherwise specified, we use $s(t) = \sqrt{t}$ (where t can be df_g) to ensure that the penalty term is of the same order as the number of parameters df_g. The same rescaling is used in [28].

Lemma 2.1. *Assume that* $0 < \sum_{i=1}^{n} y_i < n$. *For* $\lambda \geq 0$ *and* $s(t) > 0$ *for all* $t \in \mathbb{N}$, *the minimum of the optimization problem (2.9) is achieved.*

The condition on the observed data in Lemma 2.1. If X and Z are of full rank, the minimizer of $S_\lambda(.)$ is unique; otherwise, the set of minimizers is a convex set, all of whose elements achieve the same minimum value. The L_2 structure of the group penalty implies that, generically, either $\hat{\theta}_g = 0$ or $\hat{\theta}_{g,j} \neq 0$ for every $j \in \{1, \ldots, df_g\}$ (indices omitted for brevity). A geometric interpretation of this particular sparsity property is provided by [28].

2.6. **Selection of the penalty parameter λ.** The tuning parameter λ_n depends on the sample size n, the number of groups $G_n + G'_n$, and the degrees of freedom dfx_g and $dfz_{g'}$ within each group. Assuming the degrees of freedom per group are fixed, the regularization parameter λ_n can be chosen on the order of $log(G + G')$. In this case, it can be shown that the group Lasso is globally consistent, under certain additional regularity and sparsity conditions. When $\lambda = 0$, we recover the classical unpenalized maximum likelihood for the full model. For very large values of λ, all parameters are estimated as exactly zero; furthermore, an increase in λ_n leads to a decrease in θ_g toward zero, meaning that certain blocks simultaneously become null and groups of predictors are eliminated from the model, leading to model underfitting. Therefore, λ must be carefully chosen to achieve a proper balance between goodness-of-fit and sparsity. All questions related to correct estimation and model selection are, in fact, conditioned by the proper choice of λ, as this value determines the size of the model selected by the Lasso. The parameter λ is tuned via cross-validation ([23]) or by minimizing AIC and BIC criteria. These criteria provide guidance on how to adjust the degree of regularization in light of the prediction problem, but do not provide desirable guidance regarding the estimation or selection problem. We focus here on prediction properties. In practice, given a grid of tuning parameters $\lambda_1, \ldots, \lambda_k$, we

run the group Lasso algorithm for each λ_i. Each λ_i is linked to a subset S_{λ_i} of selected variables. For each λ_i, we perform a zero-inflated logistic regression model using the selected subset S_{λ_i} and calculate prediction errors using leave-one-out cross-validation. The regularization parameter λ is used to adjust the trade-off between loss minimization and the search for a group-level sparse solution. The group Lasso offers the same advantages as the Lasso. Finally, the optimal subset of selected variables is chosen as the subset that minimizes the prediction error. These methods were implemented using the R software packages glmnet and grplasso.

3. Algorithm for the group Lasso

3.1. **Block Coordinate Descent (BCD).** Parameter estimation is more computationally demanding than in the case of generalized linear regression models. The algorithm proposed by [28] relies on sequentially solving a system of nonlinear equations, which are both necessary and sufficient to minimize the penalized residual sum of squares by groups. This algorithm thus constitutes a specific case of the block coordinate descent algorithm. To address the more complex case of zero-inflated logistic regression, we can resort to a block coordinate descent algorithm. The numerical convergence of this approach is demonstrated by building upon the results of [24], as explained below:

Step	Block Coordinate Descent
Step 1	Consider $\theta = (\beta, \gamma) \in \mathbb{R}^{p+1} \times \mathbb{R}^{q+1}$ as a vector of initial parameters
Step 2	$\theta_0 := argmin_{\theta_0} S_\lambda(\theta)$
Step 3	for $g \in \{1, \ldots, G\}$ if $\|X_g^T(y - p_{-g})\|_2 + \|Z_g^T(y - \pi_{-g})\|_2 \leq \lambda\sqrt{df_g}$ $\theta_g = (0, 0)$ else $\theta_g := argmin_{\theta_g} S_\lambda(\theta)$ end end
Step 4	Repeat steps 2 and 3 until a convergence criterion is met.

Table 1. Logistic Group Lasso algorithm using block coordinate descent minimization.

We iterate through the parameter groups and minimize the objective function S_λ while keeping all parameter groups fixed except for the current one. This leads us to the algorithm presented in Table 1, where θ_{-g} denotes the parameter vector θ with θ_g set to 0, while all other components remain unchanged. In step (3), we first check whether the minimum is located at the non-differentiable point $\theta_g = 0$. If it is not, a standard numerical minimizer such as a Newton-type algorithm can be used to find the optimal solution with respect to θ_g. In this case, the values from the previous iteration can be used as starting values to save computation

time. If the group was not included in the model during the last iteration, we first take a small step in the direction opposite to the gradient of the negative log-likelihood function to ensure that we start at a differentiable point.

Proposition 3.1. *Steps (2) and (3) of the block coordinate descent algorithm perform group-wise minimization of S_λ and are well-defined in the sense that the corresponding minima are achieved. Furthermore, if we denote by $\hat{\theta}(t)$ the parameter vector after t block updates, then every limit point of the sequence $\{\hat{\theta}(t)\}_{t\geq 0}$ is a minimum point of S_λ.*

As presented in Proposition 3.1, the key lies in the separable structure of the differentiable part of S_λ. These properties apply to the group Lasso for generalized linear regression models with or without zero-inflation.

As it can be shown that the iterations remain within a compact set, the existence of a limit point is guaranteed. The main drawback of such an algorithm is that the block minimization for the active groups must be performed numerically. However, for small to medium-scale problems in terms of the dimensions p (for the logistic model) and q (for the inflation model), and group sizes df_g, this proves to be sufficiently fast. For large-scale applications, it would be beneficial to have a closed-form solution for block updates, as in [28]. This point will be addressed in the following subsection.

3.2. **Block Coordinate Gradient Descent (BCGD).** The main idea of the Block Coordinate Gradient Descent method of [25] consists in combining a quadratic approximation of the group Lasso penalized log-likelihood with an additional line search. By using a second-order Taylor expansion at $\hat{\theta}^{(t)}$ and replacing the Hessian of the group Lasso penalized log-likelihood function $\ell(.)$ with an appropriate matrix $H^{(t)}$, we define:

$$\begin{aligned} M_\lambda^{(t)}(d) &= -\left[\ell(\hat{\theta}^{(t)}) + d^T\nabla\ell(\hat{\theta}^{(t)}) + \frac{1}{2}d^T H^{(t)} d\right] \\ &+ \lambda_1\sum_{g=1}^{G}\sqrt{df_g}\|\hat{\beta}_g^t + d'_g\|_2 + \lambda_2\sum_{g=1}^{G'}\sqrt{df_g}\|\hat{\gamma}_g^t + d"_g\|_2 \\ &\approx S_\lambda(\hat{\theta}^{(t)} + d) \end{aligned} \tag{3.1}$$

where $d = (d', d")^T \in \mathbb{R}^{p+1}\times\mathbb{R}^{q+1}$.
We now consider the minimization of $M_\lambda^{(t)}(.)$ with respect to the $g-th$ group of penalized parameters. This means we restrict ourselves to vectors d where $d_k = (0,0)^T$ for $k = g$. Furthermore, we assume that the corresponding submatrix $H_{gg}^{(t)}$ of dimension $df_g \times df_g$ is of the form: $H_{gg}^{(t)} = h_g^{(t)}.I_{df_g}$ for some scalar $h_g^{(t)} \in \mathbb{R}\times\mathbb{R}$. If $\|\nabla\ell(\hat{\theta}^{(t)})_g - h_g^{(t)}\hat{\theta}_g^{(t)}\|_2 \leq \lambda\sqrt{df_g}$, the minimizer of (3.1) are $d_g^{'(t)} = -\hat{\beta}_g^{(t)}$ and $d_g^{"(t)} = -\hat{\gamma}_g^{(t)}$ among other things $d_g^{(t)} = -\hat{\theta}_g^{(t)}$ else

$$d_g^{(t)} = -\frac{1}{h_g^{(t)}}\left[\nabla\ell(\hat{\theta}^{(t)})_g - \lambda\sqrt{df_g}\frac{\nabla\ell(\hat{\theta}^{(t)})_g - h_g^{(t)}\hat{\theta}_g^{(t)}}{\|\nabla\ell(\hat{\theta}^{(t)})_g - h_g^{(t)}\hat{\theta}_g^{(t)}\|_2}\right]. \tag{3.2}$$

If $d^{(t)} \neq 0$, an inexact line search using the Armijo rule must be performed: Let $\alpha^{(t)}$ be the largest value in $\{\alpha_0 \delta^l\}_{l \geq 0}$ such that:

$$S_\lambda(\hat{\theta}^{(t)} + \alpha^{(t)} d^{(t)}) - S_\lambda(\hat{\theta}^{(t)}) \leq \alpha^{(t)} \sigma \Delta^{(t)}$$

If $d^{(t)} \neq 0$, an inexact line search using the Armijo rule must be performed: Let $\alpha^{(t)}$ be the largest value in $\{\alpha_0 \delta^l\}_{l \geq 0}$ such that $S_\lambda(\hat{\theta}^{(t)} + \alpha^{(t)} d^{(t)}) - S_\lambda(\hat{\theta}(t)) \leq \alpha^{(t)} \sigma \Delta^{(t)}$, where $0 < \delta < 1$, $0 < \sigma < 1, \alpha_0 > 0$, and $\Delta^{(t)}$ is the improvement in the objective function S_λ when using a linear approximation for the log-likelihood, i.e.:

$$\Delta^{(t)} = -(d^{(t)})^T \nabla \ell(\hat{\theta}^{(t)}) + \lambda \sqrt{df_g} \|\hat{\theta}_g^{(t)} + d_g^{(t)}\|_2 - \lambda \sqrt{df_g} \|\hat{\theta}_g^{(t)}\|_2. \tag{3.3}$$

Finally, we define $\theta^{(t+1)} = \theta^{(t)} + \alpha^{(t)} d^{(t)}$. The algorithm is described in Table 2. When minimizing $M_\lambda^{(t)}$ with respect to a penalized group, it is first necessary to check whether the minimum is located at a non-differentiable point, as indicated above. For the (unpenalized) intercept, this is not necessary, and the solution can be calculated directly.

$$d_0^{(t)} = -\frac{1}{h_0^{(t)}} \nabla(\hat{\theta}^{(t)})_0 \tag{3.4}$$

For a general matrix $H^{(t)}$, the minimization with respect to the g-th group of parameters depends on $H^{(t)}$ only through the corresponding submatrix $H_{gg}^{(t)}$. To ensure a reasonable quadratic approximation in (3.1), $H_{gg}^{(t)}$ is ideally chosen to be close to the corresponding submatrix of the Hessian of the log-likelihood function. By restricting ourselves to matrices of the form $H_{gg}^{(t)} = h_g^{(t)}.I_{df_g}$, one possible choice is :

$$h_g^{(t)} = -\max\left\{diag\left[-\nabla^2 \ell(\hat{\theta}^{(t)})_{gg}\right], c_*\right\} \tag{3.5}$$

where $c_* > 0$ is a lower bound to guarantee convergence (see proposition 3.2). The matrix $H^{(t)}$ does not necessarily need to be recomputed at each iteration. Under certain moderate conditions.

Step	Block Coordinate Gradient Descent
Step 1	Let $\theta = (\beta, \gamma) \in \mathbb{R}^{p+1} \times \mathbb{R}^{q+1}$ be a vector of initial parameters.
Step 2	for $g \in \{1, \ldots, G + G'\}$ $H_{gg} := h_g(\theta).I_{df_g}$ $d := argmin_{d \mid d_k = 0, k \neq g} M_\lambda(d)$ if $d \neq 0$ $\alpha :=$ Line search $\theta := \theta + \alpha.d$ end end
Step 3	Repeat steps 2 and 3 until a convergence criterion is met.

TABLE 2. Group Lasso Logistic Algorithm using Block Coordinate Gradient Descent.

Under certain conditions on $H^{(t)}$, the convergence of the algorithm is ensured, as shown by [28] and the proof of proposition 3.2.
Standard choices for the tuning parameters are, for instance, $\alpha_0 = 1$, $\delta = 0,5$ and $\sigma = 0,1$. Other definitions of $\Delta^{(t)}$, for example to include the quadratic part of the improvement, are also possible. We refer the reader to [25] for further details and proofs that $\Delta^{(t)} < 0$ for $d^{(t)} = 0$ and that the line search can always be performed.

Proposition 3.2. *If $H_{gg}^{(t)}$ is chosen according to equation (3.5), then every limit point of the sequence $\{\hat{\theta}(t)\}_{t\geq 0}$ is a minimum point of S_λ.*

Remark 3.3. When cycling through the coordinate blocks, we could restrict ourselves to the current active set and visit the remaining blocks, for instance every 10 iterations, to update the active set. In high-dimensional cases, for example, this modification reduces computation times by approximately 40% compared to what is shown in Figure 2. Furthermore, it is also possible to update the coordinate blocks non-cyclically or all at once, which would allow for a parallelizable approach while maintaining the convergence result.

Remark 3.4. The Block Coordinate Gradient Descent algorithm can also be applied to the Group Lasso in other models. For instance, any generalized linear model where the response y has a distribution from the exponential family belongs to this class. This is available in our R package grplasso.

A similar algorithm is presented in [12], where a global upper bound on the Hessian is used to solve the Lasso problem for multinomial logistic regression. This approach can also be used with the Group Lasso penalty, yielding a closed-form solution for a block update. However, the upper bound is not tight enough for moderate and small values of λ, generally leading to excessively slow convergence. For zero-inflated generalized linear models, the LARS algorithm [7, 18] is highly efficient for computing the Lasso solution path $\{\hat{\theta}_\lambda\}_{\lambda\geq 0}$. For zero-inflated logistic regression, approximate path-following algorithms have been proposed [21, 30, 7]. But with the Group Lasso penalty, some of them are not applicable [21] or do not necessarily converge to a minimum point of S_λ [30], and none of them seem to be computationally faster than iteratively working over a fixed grid of penalty parameters λ. This latter point was also observed by [9] for large-scale Lasso logistic regression applications.
To compute the solutions $\hat{\theta}_g$ over a grid of the penalty parameter $0 < \lambda_K < \cdots < \lambda_1 < \lambda_{max}$, we can, for instance, start at a given λ_{max}:

$$\lambda_{max} = \max_{g=1,\ldots,G} \frac{1}{\sqrt{df_g}} \|X_g^T(y-\bar{y})\|_2 + \max_{g=1,\ldots,G'} \frac{1}{\sqrt{df_g}} \|Z_g^T(y-\bar{y})\|_2$$

Where only the intercept β_0 is in the logistic model and γ_0 is in the inflation model. We then use $\hat{\theta}_{\lambda_k}$ as the starting value for $\hat{\theta}_{\lambda_{k+1}}$ and proceed iteratively until $\hat{\theta}_{\lambda_K}$, with λ_K being equal or close to zero. Instead of updating the Hessian approximation $H^{(t)}$ at each iteration, we can use a constant matrix based on the

previous parameter estimates $\hat{\theta}_{\lambda_k}$ to save computation time, i.e.:

$$H_{gg}^{(t)} = h_g(\hat{\theta}_{\lambda_k}) I_{df_g}$$

For the estimation of $\hat{\theta}_{\lambda_{k+1}}$, cross-validation can then be used to choose the parameter λ. Most often, we aim for a minimum negative log-likelihood score on the test sample.

4. Algebraic Properties

Let (x_i, z_i, y_i) be independent and identically distributed copies of the random vector (x, z, y). The loss function used is the penalized log-likelihood (a negative log-likelihood function):

$$\begin{aligned} \ell(\theta) &= J \times log\left(e^{\gamma^T z} + (1 + e^{\beta^T x})^{-1}\right) - log\left(1 + e^{\gamma^T z}\right) \\ &+ (1 - J)\left(y\beta^T x - log(1 + e^{\beta^T x})\right) \end{aligned} \tag{4.1}$$

with $J := 1_{\{y=0\}}$
Assuming that the degrees of freedom per group are fixed, the smoothing parameter λ can be taken to be of order $log(G)$. As noted in Section 2.6, assuming fixed degrees of freedom per group, the smoothing parameter λ may be taken of order log(G); under this choice, the group Lasso is globally consistent given suitable regularity and sparsity conditions, as detailed below. It should be noted that a judicious choice of the (tuning) regularization parameter λ will depend on the sample size n, the number of groups G, and the degrees of freedom within each group df_g. It can then be demonstrated that the Group Lasso is globally consistent under certain additional regularity and sparsity conditions.

This section details this asymptotic result.
The theoretical risk is defined as follows:

$$\mathcal{R}(\theta) = \mathbb{E}\left[\ell(\theta)\right] \tag{4.2}$$

and its empirical counterpart:

$$\mathcal{R}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta_i) \tag{4.3}$$

The Logistic Group Lasso estimator $\hat{\theta}_\lambda$ is the minimizer of:

$$\mathcal{S}_\lambda(\theta) = n\mathcal{R}_n(\theta) + \mathcal{P}(\theta, \lambda) \tag{4.4}$$

$$\frac{\mathcal{S}_\lambda(\theta)}{n} = \mathcal{R}_n(\theta) + \frac{\mathcal{P}(\theta, \lambda)}{n} \tag{4.5}$$

Now, let θ^0 be a minimizer satisfying:

$$\theta^0 \in argmin_\theta \mathcal{R}(\theta) \tag{4.6}$$

If the model is correctly specified, then we have:

$$\mathbb{E}(y \mid x) = p_{\beta^0}(x) \tag{4.7}$$

$$\mathbb{E}(y \mid z) = \pi_{\gamma^0}(z) \tag{4.8}$$

There are several ways to measure the quality of the estimation procedure. We will use the global measure.

$$d^2(\eta_{\hat{\beta}_\lambda}, \eta_{\beta^0}) = \mathbb{E}\left[|\eta_{\hat{\beta}_\lambda}(x) - \eta_{\beta^0}(x)|^2\right] \tag{4.9}$$

$$d^2(\eta_{\hat{\gamma}_\lambda}, \eta_{\gamma^0}) = \mathbb{E}\left[|\eta_{\hat{\gamma}_\lambda}(z) - \eta_{\gamma^0}(z)|^2\right] \tag{4.10}$$

Where η is chosen as the $logit()$ function.
The following assumptions are made:
Assumption 1
We assume that for some constant $0 < \epsilon \leq \frac{1}{2}$, for all x and z we have :

$$\epsilon \leq p_{\beta^0}(x) \leq 1 - \epsilon \tag{4.11}$$

$$\epsilon \leq \pi_{\gamma^0}(z) \leq 1 - \epsilon \tag{4.12}$$

Assumption 2
We shall require the matrix to : $\Sigma_x = \mathbb{E}[xx^T]$ and $\Sigma_z = \mathbb{E}[zz^T]$ are non-singular. Let υ^2 denote the smallest eigenvalue of Σ.
Assumption 3
Let x_g and z_g be the g-th predictors in x and z, respectively. We normalize x_g and z_g such that they have an identical inner product matrix, specifically $\mathbb{E}[x_g x_g^T] = I_{df\,x_g}$ and $\mathbb{E}[z_g z_g^T] = I_{df\,z_g}$. With this normalization, we further assume that for a constant $\mathcal{L}_n$, we have:

$$\max_x \max_g x_g^T x_g \leq nL_n^2 \quad \text{et} \quad \max_z \max_g z_g^T z_g \leq nL_n^2 \tag{4.13}$$

Since we use the previous normalization, the smallest possible order for L_n^2 is $L_n^2 = O\left(\frac{1}{n}\right)$. For categorical predictors, $L_n^2 = O\left(\frac{1}{n}\right)$ corresponds to the balanced case where, in each category, the probability of finding an individual within that category is bounded away from zero and one.

Consistency can then be demonstrated in the following sense. Let θ_g^0 denote the elements of the vector θ^0 corresponding to the g-th group. Let N_0 be the number of non-zero group effects, i.e., the number of vectors θ_g^0 satisfying $\|\theta_g^0\| \neq 0$. There exist universal constants $C1$, $C2$, $C3$ and $C4$, and constants $c1$ and $c2$ depending on ϵ, υ, $\max_g df x_g$ and $\max_g df z_g$, such that whenever:
Assumption 4

$$C1(1 + N_0^2)L_n^2 logG \leq c_1 \quad \text{et} \quad C_1 logG \leq \lambda \leq \frac{c_1}{(1 + N_0^2)L_n^2} \tag{4.14}$$

The following probability inequality holds:

$$\mathbb{P}\left[d^2\left(\eta_{\hat{\theta}_\lambda}; \eta_{\theta^0}\right) \geq c_2 \frac{(1 + N_0)\lambda}{n}\right] \leq C_2\left[log(n) exp\left(-\frac{\lambda}{C_3}\right) + exp\left(-\frac{1}{C_4 L_n^2}\right)\right] \tag{4.15}$$

with $\eta()$.
This result follows from arguments similar to those used by [26]. An overview of the proof is presented in the Appendix.
For the asymptotic implications, assume that ϵ, υ, $\max_g df x_g$ and $\max_g df z_g$ are held fixed as n tends to infinity and that G remains much larger than $logn$. Let λ be substantially equal to $logG$, i.e., it is of order $logG$. When, for instance, the

number of non-zero group effects N_0 satisfies $N_0 = O(1)$ and $L_n^2 = O(\frac{1}{logG})$, we then find the quasi-parametric rate.

$$d^2\left(\eta_{\hat{\theta}_\lambda}, \eta_{\theta^0}\right) = O_p\left(\frac{logG}{n}\right) \tag{4.16}$$

With $L_n^2 = O\left(\frac{1}{n}\right)$, the maximum growth rate for N_0 is $N_0 = O\left(\sqrt{\frac{n}{logG}}\right)$, and when N_0 is exactly of this order, we obtain the rate

$$d^2\left(\eta_{\hat{\theta}_\lambda}, \eta_{\theta^0}\right) = O_p\left(\sqrt{\frac{log(G)}{n}}\right) \tag{4.17}$$

Remark 4.1. The result can be improved by replacing N_0 with the number of non-zero coefficients of the 'best' approximation of 0, which is the approximation that balances the estimation error and the approximation error.

Remark 4.2. A similar consistency result can be obtained for the Group Lasso in Gaussian regression.

5. Simulation study

We use the performance of Group Lasso for Logistic Regression with Zero Inflation on simulated datasets. The calculations were performed using the R software. There is no R package that handles ZIBer and Group Lasso simultaneously. We will therefore proceed with a pragmatic approach. We use the grpreg package developed by Patrick Breheny, Yaohui Zeng and Ryan Kurth for Group Lasso, and the MASS package developed by Brian Ripley, in collaboration with other contributors. The MASS package is part of the R project and is widely used for statistical analyses, including regression methods, discriminant analysis and other advanced functions. The dplyr package was created by Hadley Wickham, a highly influential statistician and R software developer. It forms part of the tidyverse ecosystem, a collection of packages designed for data science in R. The grpreg function is part of the grpreg package in R, designed for regularised regression with grouped variables. It supports various regularisation methods such as group Lasso, group MCP and group SCAD. It is suitable for linear, logistic, Poisson and Cox regression models. We evaluate the asymptotic properties of two-stage estimators with Group Lasso on each logistic and inflation component separately in finite dimensions through simulations. We also use Elastic-net for comparison.

5.1. **Study design.** :

We compare the ZIBer model's finite-sample performance under two regularization methods: Group Lasso y Group Lasso [16] and Elastic-net [31]. In zero-inflated logistic regression, the response exhibits an excess of zeros relative to a standard Bernoulli model; group-wise variable selection is important here for reducing model complexity and improving prediction [5]. As for variable selection by penalising the log-likelihood, we use regularisation methods such as Group Lasso and Elastic-net. For this task, we study the properties of the ZIBer model

estimators in finite dimensions, and an application to real-world data will be carried out through simulations using the R programming language.
We simulate $p = 100$, $p = 160$ and $p = 240$ predictors for the logistic regression part and $q = 100$, $q = 35$ and $q = 240$ predictors for the zero-infation part, using sample sizes of $n = 500$ and finally $n = 1000$. We ran Group Lasso for the logistic regression and for the zero-inflated regression on these datasets. The responses are simulated using the logistic regression model with zero inflation ($Y \sim ZIBer(n, p, \pi)$) where $p = \frac{e^{X\beta}}{1+e^{X\beta}}$ and $\pi = \frac{e^{Z\gamma}}{1+e^{Z\gamma}}$ with $\beta = (\beta^1, \ \ldots, \beta^G)$ and $\gamma = (\gamma^1, \ldots, \gamma^{G'})$ with, respectively, g groups having non-zero coefficients among the G groups and g' groups having non-zero coefficients among the G' groups. We estimate the parameters γ and β for the probability of a zero (π) and the probability of success (p), respectively. We then use the datasets to evaluate the performance of the fitted model according to a specific loss function. The parameter vector $\theta = (\beta, \gamma)^T$ is finally recalibrated to adjust the empirical risk $r = (r_1, r_2)$ to the desired level.

$$r_1 = \frac{1}{n}\sum_{i=1}^{n} min\{p(x_i), 1 - p(x_i)\} \tag{5.1}$$

$$r_2 = \frac{1}{n}\sum_{i=1}^{n} min\{\pi(z_i), 1 - \pi(z_i)\} \tag{5.2}$$

We determine the optimal regularisation parameter using cross-validation. From this, we determine the model with the parameter vectors β_{r_1} and γ_{r_2}. Next, we calculate the hits that is, the number of relevant variables correctly identified—the false positives that is, the number of non-significant variables selected as relevant—and the degrees of freedom that is, the total number of variables selected in the model. Finally, we estimate the performance of the selected model by calculating several metrics such as bias and RMSE to evaluate and interpret the performance of the prediction model (ZIBerRM). We ran Group Lasso without excess zeros (logistic part) and Group Lasso with excess zeros (inflation part) on these datasets.

Three examples are considered in the simulations; we use a random design matrix where the predictors are simulated according to a uniform distribution to ensure bounded predictors. There were 100 replicates for each set of parameters described below. These parameters were chosen to have a different number of effective predictors and different levels of zero inflation.

Example 1: : let's take $p = 100$ predictors in the logistic component:

$$\begin{array}{llll} \text{Group 1} & X_i = U_1 + \epsilon_i & \text{pour} & 1 \leq i \leq 10 \\ \text{Group 2} & X_i = U_2 + \epsilon_i & \text{pour} & 11 \leq i \leq 20 \\ \text{Group 3} & X_i = U_i & \text{for} & 21 \leq i \leq 30 \\ \vdots & \vdots & \vdots & \vdots \\ \text{Group 10} & X_i = U_i & \text{for} & 91 \leq i \leq p \end{array}$$

with $U_1 \sim U([0,1])$, $U_2 \sim U([0,1])$, $U_i \sim U([-0.1,0.1])$ and $\epsilon_i \sim U([0,0.01])\forall i \in \{1,\ldots,p\}$ and $q = 100$ predictors in the zero-inflated component:

$$\begin{array}{llll} \text{Group 1} & Z_i = U_3 & \text{for} & 1 \leq i \leq 10 \\ \text{Group 2} & Z_i = U_4 & \text{for} & 11 \leq i \leq 20 \\ \text{Group 3} & Z_i = U_i + \epsilon_i & \text{pour} & 21 \leq i \leq 30 \\ \vdots & \vdots & \vdots & \vdots \\ \text{Group 10} & Z_i = U_i + \epsilon_i & \text{pour} & 91 \leq i \leq q \end{array}$$

with $U_3 \sim U([0,0.5])$, $U_4 \sim U([0.5,1])$, $U_i \sim U([-0.5,0.5])$ et $\epsilon_i \sim U([0,0.001])\forall i \in \{1,\ldots,q\}$. The covariates in the first two blocks are highly correlated (~ 0.8) and there are weak correlations between the blocks. The coefficients are:

$$\beta = (\overbrace{1.10,\ldots,1.10}^{10},\overbrace{-0.32,\ldots,-0.32}^{10},\overbrace{-0.36,\ldots,-0.36}^{10},\overbrace{\overbrace{0,\ldots,0}^{10},\ldots,\overbrace{0,\ldots,0}^{10}}^{10})$$

$$\gamma = (\overbrace{0.30,\ldots,0.30}^{10},\overbrace{-0.48,\ldots,-0.48}^{10},\overbrace{0.44,\ldots,0.44}^{10},\overbrace{\overbrace{0,\ldots,0}^{10},\ldots,\overbrace{0,\ldots,0}^{10}}^{10})$$

The zero inflation rate is approximately 55%.

Example 2: : we have doubled the number of ineffective predictors from the examples above, and the parameters have been set to achieve approximately 29% zero inflation.

$$\beta = (\overbrace{1.50,\ldots,1.50}^{20},\overbrace{-0.22,\ldots,-0.22}^{20},\overbrace{-0.25,\ldots,-0.25}^{20},\overbrace{0.30,\ldots,0.30}^{20},\overbrace{\overbrace{0,\ldots,0}^{20},\ldots,\overbrace{0,\ldots,0}^{20}}^{8})$$

$$\gamma = (\overbrace{-1.05,\ldots,-1.05}^{20},\overbrace{0.45,\ldots,0.45}^{20},\overbrace{-0.30,\ldots,-0.30}^{20},\overbrace{-0.36,\ldots,-0.36}^{20},\overbrace{\overbrace{0,\ldots,0}^{20},\ldots,\overbrace{0,\ldots,0}^{20}}^{8})$$

Example 3: : the parameters are chosen to achieve approximately 21% zero inflation.

$$\beta = (\overbrace{\overbrace{2.20,\ldots,2.20}^{15},\ldots,\overbrace{2.20,\ldots,2.20}^{15}}^{4},\overbrace{\overbrace{-0.72,\ldots,-0.72}^{15},\ldots,\overbrace{-0.72,\ldots,-0.72}^{15}}^{3},\overbrace{\overbrace{0,\ldots,0}^{15},\ldots,\overbrace{0,\ldots,0}^{15}}^{9})$$

$$\gamma = (\overbrace{\overbrace{-2.01,\ldots,-2.01}^{15},\ldots,\overbrace{-2.01,\ldots,-2.01}^{15}}^{4},\overbrace{\overbrace{0.88,\ldots,0.88}^{15},\ldots,\overbrace{0.88,\ldots,0.88}^{15}}^{3},\overbrace{\overbrace{0,\ldots,0}^{15},\ldots,\overbrace{0,\ldots,0}^{15}}^{9})$$

5.2. **Results.** The results of the eight simulations are shown in Tables 3, 4 and 5 for sample sizes $n = 500$ and $n = 1000$ respectively. For each configuration

(sample size and zero inflation proportion) of the simulation parameters, we calculate: the mean bias, the root mean square error (RMSE), mean hit and mean false positive for $\hat{\beta}_i$ for all $i \in \{1, ..., p\}$ and $\hat{\gamma}_i$ for all $i \in \{1, ..., q\}$.

- $\triangleright$: p is the number of covariates in the logistic part and q is the number of covariates in the zero inflation part.
- $\triangleright$: sEN is the number of significant covariates in the logistic part and s' is the number of significant covariates in the zero-inflation part for the Elastic-Net method.
- $\triangleright$: sGL is the number of significant covariates in the logistic part and s' is the number of significant covariates in the zero inflation part for the Group Lasso method.
- $\triangleright$: G is the number of groups in the logistic part and G' is the number of groups in the zero inflation part.
- $\triangleright$: G^* is the number of non-zero groups in the logistic part; $G^{*'}$ is the number of non-zero groups in the zero inflation part.
- $\triangleright$: v is the size of the non-zero groups in the logistic part. v' is the size of the non-zero groups in the zero-inflation part.

λ	Example 1	Example 2	Example 3
Elastic-Net logistic part			
λ_{min}	0.01843	0.01612	0.00219
λ_{max}	0.03222	0.02818	0.05694
Elastic-Net zero inflation part			
λ_{min}	0.01777	0.01816	0.00198
λ_{max}	0.03412	0.03484	0.03903
Groupe Lasso partie logistique			
λ_{min}	0.01575	0.01292	0.00077
λ_{max}	0.02285	0.03682	0.07181
Group lasso zero inflation part			
λ_{min}	0.01513	0.01663	0.00861
λ_{max}	0.03261	0.04121	0.06977

TABLE 3. The tuning parameters (λ) for the five examples $n = 1000$

Examples	Example 1	Example 2	Example 3
Logistic part			
p	100	240	240
sEN	24	240	195
sGL	10	20	220
G	10	12	16
v	10	20	15
G^*	01	01	11
	Mean Hit (%) (Average success rate)		
Elastic-net	99.95833	99.87468	99.94609
Group Lasso	99.99083	100	99.99962
			Mean False positive (%)
Elastic-net	16.31579	3.60109	36.59794
Group Lasso	27.10448	80.10452	37.29742
			Biais
Elastic-net	-0.16267	-0.18317	-0.68121
Group Lasso	-0.12979	-0.18916	-0.68503
			RMSE
Elastic-net	0.00097	0.00095	0.00068
Group Lasso	0.00079	0.00044	0.00022
Zero inflation part			
q	100	240	240
sEN'	25	240	210
sGL'	10	240	90
G'	10	12	16
v	10	20	15
$G^{*'}$	01	12	06
	Mean Hit (%) (Average success rate)		
Elastic-net	90.01327	60.12029	89.91758
Group Lasso	100	100	99.99962
			Mean False positive (%)
Elastic-net	14.73333	3.60109	40.59794
Group Lasso	37.21951	82.32489	38.23102
			Biais
Elastic-net	-0.00009	-0.00092	0.00037
Group Lasso	-0.00009	-0.00002	-0.00381
			RMSE
Elastic-net	0.00037	0.00066	0.00193
Group Lasso	0.00008	0.00042	0.00019

TABLE 4. Simulation results for the three examples for $n = 500$

Group Lasso appears to perform better than Elastic-net at including relevant predictors in the model, particularly when there are strong intra-group correlations. Elastic-net tends to select fewer variables from among the influential ones

Examples	Example 1	Example 2	Example 3
Logistic part			
p	100	240	240
sEN	41	220	210
sGL	10	100	120
G	10	12	16
v	10	20	15
G^*	01	05	06
	Mean Hit (%) (Average success rate)		
Elastic-net	99.00271	77.97527	80.06381
Group Lasso	99.99999	100	100
			Mean False positive (%)
Elastic-net	18.49153	6.18421	33.46724
Group Lasso	29.22841	87.42763	35.14837
			Biais
Elastic-net	-0.00009	-0.00092	0.00037
Group Lasso	-0.00007	-0.00009	-0.00101
			RMSE
Elastic-net	-0.00017	-0.00105	0.00073
Group Lasso	-0.00011	-0.00048	-0.00572
Zero inflation part			
q	100	240	240
sEN'	53	240	210
sGL'	30	240	75
G'	10	12	16
v	10	20	15
$G^{*'}$	03	12	5
	Mean Hit (%) (Average success rate)		
Elastic-net	92.98973	72.05654	95.91129
Group Lasso	100	99.99981	100
			Mean False positive (%)
Elastic-net	19.36174	6.18421	33.46742
Group Lasso	29.10446	87.99052	37.82814
			Biais
Elastic-net	-0.00009	-0.00085	0.00034
Group Lasso	-0.00005	-0.00008	0.00099
			RMSE
Elastic-net	0.00017	0.00078	0.000074
Group Lasso	0.00002	0.00014	0.000016

TABLE 5. Simulation results for the threze examples for $n = 1000$

(Elastic-net selects only certain variables from groups of highly correlated predictors) than Group Lasso in the case of highly correlated covariates and when the size or number of non-zero groups is large. Group Lasso succeeds in including the truly significant groups in most cases. Conversely, the Group Lasso estimator tends to include more irrelevant covariates than Elastic-net, particularly when the number or size of non-influential groups is large. We might expect this result because the Group Lasso estimator does not include a single variable, but rather groups of variables (when one of the covariates is included in the model, all others belonging to the same group are also included in the model). Thus, Group Lasso selects models that are larger than the true model.
However, when Group Lasso selects a number of covariates equal to the number of significant covariates, it is highly likely that the correct groups are included. We also measured the performance of Elastic-net and Group Lasso in terms of estimation error and prediction error. Group Lasso appears to perform better than Elastic-net in terms of estimation error and prediction error in most cases, and the improvement is particularly significant for prediction error.
We can also note that the performance of both estimators decreases as (G_n, G'_n) and $(G^*_n, G^{*\prime}_n)$ increase. In conclusion, the Group Lasso estimator appears to perform better than Elastic-net by including hits in the model and in terms of prediction and estimation error when the covariates are structured into groups, and particularly in the case of strong correlations within the groups.

6. Real data application

In this section, we apply the Group Lasso-penalised ZIBer model to data from the *National Medical Expenditure Survey 1987–1988* (NMES1988) [4] (a large-scale cross-sectional study conducted to assess the demand for medical care in the United States). This dataset is particularly well-suited to the zero-inflated Bernoulli model due to the high proportion of non-utilisation of care (zeros) in the medical consumption indicators.

6.1. **Data description.** The distinctive feature of Group Lasso lies in the definition of group structures. In this application, the $p + q$ explanatory variables are partitioned into $G + G'$ thematic groups, allowing for the selection of coherent blocks rather than isolated variables. We will classify these 19 variables into 10 groups of 5 (making a total of 50 slots). To classify these 19 variables into 10 groups of 5 (making a total of 50 slots), it is necessary to repeat certain variables. The parameters $\widehat{\beta}$ (Bernoulli component) and $\widehat{\gamma}$ (zero inflation component) are estimated using an EM (Expectation-Maximisation) algorithm. At each M-step, we minimise the negative log-likelihood function penalised by Group LASSO (2.9). The regularisation parameter $\lambda = (\lambda_1, \lambda_2)$ is determined by 10-fold cross-validation (10-fold CV), by maximising the predictive log-likelihood.

6.2. **Results.** We fit a penalised ZIBer model to the data using the Group LASSO method, including all covariates. We then select the variables at a significance threshold of 5%. We present the estimate, the standard error (s.e.) and the

significance level (labelled: not significant (NS), significant (S) or highly significant (SVS)) of the Wald nullity test for each parameter, for the null hypothesis $H_0 : \theta_i = 0$. We run the following model:

$$logit(p_i) = \beta X \quad \text{and} \quad logit(\pi_i) = \gamma Z. \tag{6.1}$$

The results are shown in the table 6.2.

6.3. **Discussion.** The results in Table 6.2 illustrate a structural advantage of Group Lasso when applied to the ZIBer model: entire thematic clusters 'health-care utilization,' 'basic demographic profile,' and 'socio-economic status' are retained or discarded as coherent blocks, rather than being split apart as individual dummy variables would be under an ordinary Lasso. This is clinically more interpretable, since these indicators jointly describe a single risk profile for the patient rather than a set of unrelated covariates.
Furthermore, this approach enhances the model's parsimony in terms of the inflation structure. Indeed, simulations show that the variables "Health Insurance Coverage" and "Vulnerability Factors" are the first to be significant, suggesting that these factors exert a positive influence on the probability of non-utilization of non-medical care.
Finally, the method ensures greater estimation stability. By imposing simultaneous selection on the coefficient vectors β and γ, Group Lasso mitigates the recurring numerical instability in standard ZIBer models, where a variable might be deemed significant in one component whilst generating excessive noise in the other.
An interesting approach for future work would be to compare these results with a Bayesian approach, and also to examine the problem of variable selection in cases where there is missing data in the covariates

7. Conclusion

We studied the zero-inflated logistic model in high-dimensional settings and proposed a group Lasso estimator for β and γ when covariates are naturally structured into groups and the true parameter is group-sparse; only a few groups being relevant to explaining the response. Under assumptions on the sparsity of β and γ, the correlation structure between covariate groups, and the choice of tuning parameter, we established consistency and an asymptotic Gaussian approximation for the resulting estimator.
When all groups have size one, our framework reduces to the ordinary Lasso for zero-inflated logistic models; we further extended the comparison to the Elastic-net penalty, evaluating its performance against Group Lasso on simulated data. Our results show an improvement in both variable selection and prediction error for Group Lasso relative to Elastic-net when covariates are structured into groups. We also used simulation to illustrate the effect of the total number of groups, the number of non-zero groups, and group size on Group Lasso's performance; the estimator performs particularly well in high-dimensional, group-sparse settings with strong within-group correlation, though it requires prior knowledge of the group structure its main practical limitation.
On the NMES1988 data, the Group Lasso approach offers better structural interpretability than Elastic-net while maintaining comparable predictive performance (as measured by RMSE), and confirms that a multidimensional health-status profile is the principal driver of variation in medical-care utilization in this population.

Group	Parameter	Variable	Estimate	Std. Error	Wald test
Bernoulli component					
Group 1	$\hat{\beta_{1n}}$	visits	-0.372435	0.060609	VS
	$\hat{\beta_{2n}}$	nvisits	0.277489	0.036426	VS
	$\hat{\beta_{3n}}$	ovisits	0.313113	0.048290	VS
	$\hat{\beta_{4n}}$	novisits	0.109707	0.005722	S
	$\hat{\beta_{5n}}$	emergency	0.177876	0.012395	VS
Group 2	$\hat{\beta_{6n}}$	hospital	0.000000	-	NS
	$\hat{\beta_{7n}}$	emergency	0.000000	-	NS
	$\hat{\beta_{8n}}$	visits	0.000000	-	NS
	$\hat{\beta_{9n}}$	ovisits	0.000000	-	NS
	$\hat{\beta_{10n}}$	nvisits	0.000000	-	NS
Group 3	$\hat{\beta_{5n}}$	health	0.000000	-	NS
	$\hat{\beta_{6n}}$	chronic	0.000000	-	NS
	$\hat{\beta_{7n}}$	adl	0.000000	-	NS
	$\hat{\beta_{8n}}$	age	0.000000	-	NS
	$\hat{\beta_{10n}}$	health	0.000000	-	NS
Group 4	$\hat{\beta_{11n}}$	age	0.218615	0.027832	S
	$\hat{\beta_{12n}}$	gender	-0.001999	0.005304	S
	$\hat{\beta_{13n}}$	afam	0.795283	0.012941	S
	$\hat{\beta_{14n}}$	married	-2.656241	0.005934	VS
	$\hat{\beta_{15n}}$	region	-1.404549	0.005347	VS
Group 5	$\hat{\beta_{16n}}$	income	0.027560	0.051146	S
	$\hat{\beta_{17n}}$	employed	-0.554917	0.025491	VS
	$\hat{\beta_{18n}}$	school	0.707816	0.003192	VS
	$\hat{\beta_{19n}}$	insurance	-1.456849	0.006147	VS
	$\hat{\beta_{20n}}$	medicaid	-2.816104	0.010934	VS
Zero component					
Group 6	$\hat{\gamma_{1n}}$	insurance	0.006777	0.035115	VS
	$\hat{\gamma_{2n}}$	medicaid	-0.007812	0.051089	VS
	$\hat{\gamma_{3n}}$	employed	-0.008276	0.131608	S
	$\hat{\gamma_{4n}}$	income	0.062487	0.016657	VS
	$\hat{\gamma_{5n}}$	health	-0.006143	0.008212	VS
Group 7	$\hat{\gamma_{6n}}$	chronic	0.227574	0.002355	VS
	$\hat{\gamma_{7n}}$	adl	-0.107657	0.001824	VS
	$\hat{\gamma_{8n}}$	medicaid	0.024505	0.074243	S
	$\hat{\gamma_{9n}}$	afam	0.208744	0.010042	S
	$\hat{\gamma_{10n}}$	age	-0.010646	0.060848	VS
Group 8	$\hat{\gamma_{11n}}$	region	0.000000	-	NS
	$\hat{\gamma_{12n}}$	insurance	0.000000	-	NS
	$\hat{\gamma_{13n}}$	income	0.000000	-	NS
	$\hat{\gamma_{14n}}$	gender	0.000000	-	NS
	$\hat{\gamma_{15n}}$	school	0.000000	-	NS
Group 9	$\hat{\gamma_{16n}}$	employed	0.000000	-	NS
	$\hat{\gamma_{17n}}$	school	0.000000	-	NS
	$\hat{\gamma_{18n}}$	married	0.000000	-	NS
	$\hat{\gamma_{19n}}$	age	0.000000	-	NS
	$\hat{\gamma_{20n}}$	income	0.000000	-	NS
Group 10	$\hat{\gamma_{21n}}$	age	0.000000	-	NS
	$\hat{\gamma_{22n}}$	gender	0.000000	-	NS
	$\hat{\gamma_{23n}}$	region	0.000000	-	NS
	$\hat{\gamma_{24n}}$	health	0.000000	-	NS
	$\hat{\gamma_{25n}}$	insurance	0.000000	-	NS

TABLE 6. Analysis of health data using the Group Lasso regularization method, with a 25% zero inflation proportion

8. Appendix

8.1. **Proof of lemma 2.1.** As a reminder, the ZIBer model using the lasso group can be defined as follows:

Regression part: :

$$logit(p_i) = \beta^T x_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_G x_{iG} := \beta_0 + d_i \tag{8.1}$$

Inflation part (classification): :

$$logit(\pi_i) = \gamma^T z_i = \gamma_0 + \gamma_1 z_{i1} + \cdots + \gamma_{G'} z_{iG'} := \gamma_0 + d'_i \tag{8.2}$$

We will first eliminate the intercept. Let $\beta_1, \ldots, \beta_G$ and $\gamma_1, \ldots, \gamma_{G'}$ be two fixed parameters. To obtain an estimate of the intercept, we must minimise the following logarithmic likelihood function:

$$\begin{aligned} f(\theta_0) &= -\sum_{i=1}^{n} \left[Jlog\left(e^{\gamma_0+d'_i} + (1+e^{\beta_0+d_i})^{-1}\right) - log\left(1+e^{\gamma_0+d'_i}\right)\right] \\ &+ \left[(1-J)\left(y_i\beta_0 x_0 + y_i d_i - log(1+e^{\beta_0+d_i})\right)\right] \end{aligned} \tag{8.3}$$

avec $J := 1_{\{y=0\}}$.
The derivative of $f(\theta_0)$ is:
Partial derivative with respect to β_0:

$$\begin{aligned} f'(\theta_0) &= -\sum_{i=1}^{n} \left[-J_i \frac{\left(1+e^{\beta_0+d_i}\right)^{-2}}{e^{\gamma_0+d'_i} + (1+e^{\beta_0+d_i})^{-1}} \right] \\ &+ \left[(1-J_i)\left(y_i x_i - \frac{e^{\beta_0+d_i}}{1+e^{\beta_0+d_i}} \right) \right] \end{aligned} \tag{8.4}$$

Partial derivative with respect to γ_0:

$$f'(\gamma_0) = -\sum_{i=1}^{n} \left[J_i \left(\frac{e^{\gamma_0+d'_i}}{e^{\gamma_0+d'_i} + (1+e^{\beta_0+d_i})^{-1}} - \frac{e^{\gamma_0+d'_i}}{1+e^{\gamma_0+d'_i}} \right) \right] \tag{8.5}$$

It follows that

$$\lim_{\beta_0 \longmapsto +\infty} f'(\beta_0) = n - \sum_{i=1}^{n} y_i x_i > 0 \tag{8.6}$$

$$\lim_{\gamma_0 \longmapsto +\infty} f'(\gamma_0) = n > 0 \tag{8.7}$$

$$\lim_{\beta_0 \longmapsto -\infty} f'(\beta_0) = -\left[\sum_{i=1}^{n} y_i x_i - J_i \frac{1}{1+e^{\gamma_0+d'_i}} \right] < 0 \tag{8.8}$$

$$\lim_{\gamma_0 \longmapsto -\infty} f'(\gamma_0) = 0 \tag{8.9}$$

Furthermore, the function f is continuous and strictly increasing. Therefore, there exists a unique pair $(\beta_0^\star, \gamma_0^\star) \in \mathbb{R}^2$ such as $f(\theta_0^\star) = 0$. By the implicit function theorem, the corresponding functions $\beta_0^\star(\beta_1, \ldots, \beta_G)$ and $\gamma_0^\star(\gamma_1, \ldots, \gamma_{G'})$ are continuously differentiable. By replacing β_0 and γ_0 in $S_\lambda(\theta)$ by functions $\beta_0^\star(\beta_1, \ldots, \beta_G)$ and $\gamma_0^\star(\gamma_1, \ldots, \gamma_{G'})$ and by applying the theory of duality, we can rewrite (2.9) as a constrained optimisation problem $\sum_{i=1}^{G} \|\beta_g\|_2 + \sum_{i=1}^{G'} \|\beta_g\|_2 \leq t$

for a certain $t > 0$. This is a problem of optimising a continuous function on a compact set, so the minimum is reached.

8.2. **Proof of Proposition 3.1.** We first show that the minima of S_λ are attained for each group. For $g = 0$, this follows from the proof of Lemma 2.1. The case $g \geq 1$ corresponds to the minimisation of a continuous function on a compact set, so the minimum is attained. We now show that step (3) minimises the convex function $S_\lambda(\theta_g)$ for $g \geq 1$. Since $S_\lambda(\theta_g)$ is not differentiable everywhere, we invoke subdifferential calculus . The subdifferential of S_λ with respect to θ_g is the set

$$\partial S_\lambda(\beta_g) = \{-X_g^T(y - p_\beta) + \lambda e, \quad e \in E(\beta_g)\} \tag{8.10}$$

$$\partial S_\lambda(\gamma_g) = \{-X_g^T(y - \pi_\gamma) + \lambda e, \quad e \in E(\gamma_g)\} \tag{8.11}$$

$$\mathbb{E}(\beta_g) = \{e \in \mathbb{R}^{df_g} : \quad e_g = s(df_g)\frac{\beta_g}{\|\beta\|_2} \quad \text{if} \quad \beta_g \neq 0 \quad \text{and} \quad \|e_g\|_2 \leq s(df_g) \quad \text{if} \quad \beta_g = 0\} \tag{8.12}$$

$$\mathbb{E}(\gamma_g) = \{e \in \mathbb{R}^{df_g} : \quad e_g = s(df_g)\frac{\gamma_g}{\|\gamma\|_2} \quad \text{si} \quad \gamma_g \neq 0 \quad \text{and} \quad \|e_g\|_2 \leq s(df_g) \quad \text{si} \quad \gamma_g = 0\} \tag{8.13}$$

The parameter vector θ_g minimises $S_\lambda(\theta_g)$ if and only if $0 \in \partial S_\lambda(\theta_g)$, every boundary point of the sequence $\{\hat{\theta}^{(t)}\}_{t\geq 0}$ is a stationary point of the convex function S_λ, and therefore a minimum point. It should be noted that $s(df_g) = \sqrt{df_g}$.

8.3. **Proof of Proposition 3.2.** Setting $s(df_g) = \sqrt{df_g}$, the statement follows directly from Theorem 4.1(e) of [25]. We must show that $-H^{(t)}$ is bounded above and non-zero. The Hessian of the negative log-likelihood function is $N = \sum_{i=1}^n p_\beta(xi)(1 - p_\beta(xi))x_i x_i^T \leq \frac{1}{4}X^T X$ et $N' = \sum_{i=1}^n \pi_\gamma(zi)(1 - \pi_\gamma(zi))z_i z_i^T \leq \frac{1}{4}Z^T Z$ in the sense that $N - \frac{1}{4}X^T X$ and $N' - \frac{1}{4}Z^T Z$ are negative semidefinite. For the block matrices N_{gg} and N'_{gg} corresponding to the gth predictor, it follows from block orthogonalisation that $N_{gg} \leq \frac{n}{4}I_{df_g}$ and consequently $\max\{diag(N_{gg})\} \leq \frac{n}{4}$. An upper bound on $-H^{(t)}$ is therefore always guaranteed. The lower bound is imposed by the choice of $H_{gg}^{(t)}$. By the choice of line search, we ensure that $\alpha^{(t)}$ is bounded from above and thus Theorem 4.1(e) of [25] can be applied.

References

1. Antoniadis, A., & Fan, J. (2001). Regularization of wavelet approximations (with discussion). *Journal of the American Statistical Association*, 96(455), 939–967. https://doi.org/10.1198/016214501753208942.
2. Bakin, S. (1999). *Adaptive Regression and Model Selection in Data Mining Problems* (Doctoral dissertation, Australian National University). https://openresearch-repository.anu.edu.au/items/85e1d66b-ad82-4fb3-a0bb-1b09820c635b
3. Cai, T. T. (2001). Discussion of "Regularization of wavelet approximations" (auths. Antoniadis, A. and Fan, J.). *Journal of the American Statistical Association*, 96: 960–962. https://doi.org/10.1198/016214501753208951.

4. Deb, P., & Trivedi, P. K. (1997). Demand for medical care by the elderly: a finite mixture approach. Journal of Applied Econometrics, 12(3), 313-336 https://doi.org/10.1002/(SICI)1099-1255(199705)12:3<313::AID-JAE440>3.0.CO;2-G
5. Diop, A., Diop, A., & Dupuy, J.-F. (2016). Simulation-based inference in a zero-inflated Bernoulli regression model. *Communications in Statistics - Simulation and Computation*, 45(2), 652–668. https://doi.org/10.1080/03610918.2014.950743
6. Djeniba, S., Adjabi, S., & Tatachak, A. (2025). Nonparametric relative error regression estimator under weak dependence. *Gulf Journal of Mathematics*, 17(1), 12–25. https://doi.org/10.56947/gjom.v19i2.2770
7. Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407–499. https://doi.org/10.1214/009053604000000067.
8. Foster, D. P., & Stine, R. A. (2004). Variable selection in data mining: Building a predictive model for bankruptcy. *Journal of the American Statistical Association*, 99(465), 303–313. https://doi.org/10.1198/016214504000000287.
9. Genkin, A., Lewis, D. D., & Madigan, D. (2004). Large-scale Bayesian logistic regression for text categorization. *Technometrics*, 49(3), 291–304. https://doi.org/10.1198/004017007000000245.
10. Greenland, S. (2000). Invited commentary: Variable selection versus shrinkage in the control of multiple confounders. *American Journal of Epidemiology*, 151(6), 523–529. https://doi.org/10.1093/aje/kwm353.
11. Kim, Y., Kim, J., & Kim, Y. (2006). Blockwise sparse regression. *Statistica Sinica*, 16(2), 375–390. https://www3.stat.sinica.edu.tw/statistica/j16n2/j16n23/j16n23.html
12. Krishnapuram, B., Carin, L., Figueiredo, M. A., & Hartemink, A. J. (2005). Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6), 957–968. https://doi.org/10.1109/TPAMI.2005.127
13. Lee, K. Y., et al. (2006). Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Statistical Methods in Medical Research*, 15(5), 461–479. https://doi.org/10.1191/0962280206sm429oa.
14. Lounici, K., Pontil, M., Van de Geer, S., & Tsybakov, A. B. (2011). Oracle inequalities and optimal inference under group sparsity. *The Annals of Statistics*, 39(4), 2164–2204. https://doi.org/10.1214/11-AOS896.
15. Lo, Fatimata., Ba, Demba. Bocar., & Diop, Aba. (2022). Maximum likelihood estimation in the generalized extreme value regression model for binary data. *Gulf Journal of Mathematics*, 12(2), 49–56. https://gjom.org/index.php/gjom/article/view/733
16. Meier, L., van de Geer, S., & Bühlmann, P. (2008). The group Lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1), 53–69. https://doi.org/10.1111/j.1467-9868.2007.00627.x
17. Nardi, Y., & Rinaldo, A. (2008). On the asymptotic properties of the group lasso estimator for linear models. *Electronic Journal of Statistics*, 2, 605–633. https://doi.org/10.1214/08-EJS200.
18. Osborne, M., Presnell, B., & Turlach, B. (2000). A new approach to variable selection in least squares problems. *IMA Journal of Numerical Analysis*, 20(3), 389–403. https://doi.org/10.1093/imanum/20.3.389.
19. Park, M. Y., & Hastie, T. (2006). Regularization path algorithms for detecting gene interactions. *Biostatistics*, 8(2), 296–305. https://doi.org/10.1093/biostatistics/kxm010
20. Ridout, M., Hinde, J., & Demétrio, C. G. (1998). Models for count data with many zeros. *Biometrical Journal*, 40(5), 579–595. https://doi.org<579::AID-BIMJ579>3.0.CO;2-O
21. Rosset, S. (2005). Following curved regularized optimization solution paths. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 17* (pp. 1153–1160). MIT Press. https://neurips.cc

22. B. Tarigan and S.A. Van De Geer (2006). Classifiers of support vector machine type with l_1 complexity regularization. *Bernoulli*, 12(6): 1045–1076. https://doi.org/10.3150/bj/1165269150
23. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
24. Tseng, P. (2001). Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3), 475–494. https://doi.org/10.1023/A:1017501703105.
25. Tseng, P., & Yun, S. (2006). A coordinate gradient descent method for nonsmooth separable minimization. *Mathematical Programming*, 117(1), 387–423. https://doi.org/10.1007/s10107-007-0170-0.
26. Van de Geer, S. (2003). Adaptive quantile regression. In M. Akritas & D. Politis (Eds.), *Recent Advances and Trends in Nonparametric Statistics* (pp. 235–250). Elsevier. ISBN (978-0-444-51378-6).
27. West, M., et al. (2001). Predicting the clinical status of human breast cancer by using gene expression profiles. *Proceedings of the National Academy of Sciences*, 98(20), 11462–11467. https://doi.org/10.1073/pnas.201162998.
28. Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49–67. https://doi.org/10.1111/j.1467-9868.2005.00532.x
29. Zheng, H., & Zhang, Y. (2007). Feature selection for high dimensional data in astronomy. *Advances in Space Research*, 41(12), 1960–1964. https://doi.org/10.1016/j.asr.2007.09.032
30. Zhao, P., & Yu, B. (2004). *Boosted lasso* (Technical Report No. 649). Department of Statistics, University of California, Berkeley. https://www.stat.berkeley.edu/~binyu/ps/blasso.pdf
31. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

[1] Department of Mathematics, University of Alioune Diop, Bambey, Sénégal.
Email address: ndoyethiat@gmail.com

[2] Department of Mathematics, University of Alioune Diop, Bambey, Sénégal.
Email address: aba.diop@uadb.edu.sn