



HAL
open science

HoPSIR: Homogeneous Penalization of Sliced Inverse Regression

Stéphane Girard, Hadrien Lorenzo

► **To cite this version:**

Stéphane Girard, Hadrien Lorenzo. HoPSIR: Homogeneous Penalization of Sliced Inverse Regression. SFdS 2023 - 54èmes Journées de Statistique de la Société Française de Statistique, Jul 2023, Bruxelles, Belgium. pp.1-6. hal-04174522

HAL Id: hal-04174522

<https://inria.hal.science/hal-04174522>

Submitted on 1 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

HOPSIR: HOMOGENEOUS PENALIZATION OF SLICED INVERSE REGRESSION

Stéphane Girard ¹ & Hadrien Lorenzo ²

¹ *Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.*

stephane.girard@inria.fr

² *Univ. Bordeaux, CNRS, Bordeaux INP, IMB, UMR 5251, F-33400, Talence, France,*

hadrien.lorenzo@ensc.fr

Résumé. En régression, les approches purement paramétriques nécessitent un modèle parfois complexe à mettre en place. Inversement, les méthodes non-paramétriques souffrent lorsque la dimension de la variable explicative augmente puisqu'alors les points de données sont isolés les uns des autres. Les approches semi-paramétriques ont été proposées afin d'allier les bénéfices des deux approches. La méthode SIR (Sliced Inverse Regression en anglais pour Régression Inverse par Tranches en français) est une d'entre elles, la partie paramétrique permettant une réduction de dimension. En grande dimension, SIR n'est cependant plus applicable car elle nécessite l'inversion de la matrice de covariance empirique. Différentes approches ont été proposées afin de pallier cette limitation technique mais aucune n'a intégré sa solution via un modèle statistique, ce que propose ce travail. Au travers d'une classe de fonctions particulières, les fonctions homogènes de degré positif, nous introduisons une famille de lois *a priori* qui permet de construire une version pénalisée de SIR par maximisation de la loi *a posteriori*. Cette approche montre un excellent comportement sur simulations par comparaisons aux approches actuelles.

Mots-clés. Régression, réduction de dimension, pénalisation, méthodes semi-paramétriques.

Abstract. In regression, purely parametric approaches require a model that is sometimes complex to set up. Conversely, non-parametric methods suffer when the dimension of the covariate increases as the data points are isolated from each other. Semi-parametric approaches have been proposed to combine the benefits of both approaches. The SIR method is one of them, the parametric part allowing a reduction of dimension. In high dimension, however, SIR is no longer applicable as it requires the inversion of the empirical covariance matrix. Different approaches have been proposed to overcome this technical limitation but none of them has integrated its solution via a statistical model, which is precisely what is proposed in this work. Through a particular class of functions, the homogeneous functions of positive degree, we introduce a family of prior distributions which allows to build a penalized version of SIR by maximizing the posterior distribution. This approach shows an excellent behaviour on simulations compared to current approaches.

Keywords. Regression, dimension reduction, penalization, semi-parametric methods.

1 Introduction

Let be a supervised problem trying to predict a continuous response Y from a p -dimensional variable X , also continuous. For this analysis, the expert has at his/her disposal a data set of size n , where all observations are assumed to be independent and identically distributed. In the case where p becomes large, the ordinary least squares (OLS) method is not feasible because it requires the inversion of the empirical covariance matrix, which is ill-conditioned in this case. This aspect is even more pronounced if the context is non-linear, it is then necessary to estimate a larger number of parameters, which makes the problem more sensitive to small samples. Instead of regressing Y on the set of p components of X , it is common to project X onto a lower-dimensional subspace that is of interest in explaining Y . This subspace must therefore concentrate all the information contained in X that is associated with Y , this space is called a d.r.s. (for dimension-reduction subspace). Thus, by working in this central space, the information useful for the explanation of Y relative to the independent information of Y is maximal compared to the other spaces in which the analysis can be conducted. We will assume in the following that the central subspace is unidimensional.

Many solutions have been devised in order to estimate the central subspace. Among them the SIR method (for Sliced Inverse Regression) [DL91] or the SAVE method (for Sliced Average Variance Estimation) [Coo00]. Also, and more recently, [CFG17] have used the model of [Coo07] to propose a maximum likelihood estimator in the context of a Student distributed additive error.

In the case of the SIR model, the direction of the central subspace is in fact aligned with an eigen-vector associated with the largest eigenvalue of the matrix $\Sigma^{-1}\Gamma$ where $\Sigma = \text{cov}(X)$ and $\Gamma = \text{cov}(\mathbb{E}(X|Y))$. Thus, the inversion of the empirical estimate of Σ is not possible in our context. The objective of the regularization is to approach the ill-conditioned solution by slightly biasing the considered estimators. Many regularization solutions of the SIR method have been proposed. Let us quote the SIR-QZ method [CLS14] or a Ridge-based SIR approach [LY08] method for which [BMGG08] proved that the only solution of the proposed optimization problem is the null solution. Also, [BT09] proposes an elegant algorithm based on a rewriting of the considered optimization. This algorithm uses the conjugate gradient method on the Grassmann manifold [EAS98]. The solution proposed by [BMGG09] introduces a Gaussian prior to regularize the estimators.

Variable selection in predictive analysis improves models through three effects. First, by making the models more interpretable. Secondly, by eliminating from the estimates the variations due to low sampling, thus improving the predictive capacities of the model. Finally, and in comparison to other regularization approaches, by physically removing some variables from the dataset, the computations are done faster. Some solutions performing variable selection have been proposed and a non-exhaustive list of methodologies is proposed in [GLS22]. Let us note the approach proposed by [LZL19] and denoted Lasso-SIR by the authors and in the following. The Lasso-SIR model uses the eigen-vector associated with the largest eigenvalue of the matrix Γ , denoted by $\hat{\eta}$ in the associated paper, to propose as a non-regularized solution $\Sigma^{-1}\hat{\eta}$ and as regularized solutions a vector close to this direction. As discussed before, the SIR estimator looks for an eigen-vector associated with the largest

eigenvalue of the matrix $\Sigma^{-1}\Gamma$. Then, if the problem is well-conditioned, the non regularized Lasso-SIR solution and the SIR solution do not match unless Σ is proportional to the identity matrix.

2 SIR method as a maximum likelihood estimator

The notation $M \succ 0$ or equivalently $M \in \mathcal{S}_{++}^p$, resp. $M \succeq 0$ or equivalently $M \in \mathcal{S}_+^p$, denotes the $p \times p$ symmetric positive definite (resp. semi-definite) matrices. The operators $\|\cdot\|_2$ and $\|\cdot\|_1$ correspond to the Euclidean norm and the 1-norm respectively.

2.1 An inverse regression model

Let us consider the single-index inverse regression model introduced in [Coo07]:

$$X = \mu + \alpha(Y)Vb + \varepsilon, \quad (1)$$

where μ and b are nonrandom \mathbb{R}^p -vectors, ε is a centered \mathbb{R}^p -Gaussian random vector, independent of Y , with covariance matrix $\text{cov}(\varepsilon) = V \succ 0$ and $\alpha : \mathbb{R} \rightarrow \mathbb{R}$ is a nonrandom function. The unknown parameters to be estimated are (μ, V, b) and the function $\alpha(\cdot)$. Here, we focus on projection estimators, *i.e.* the function $\alpha(\cdot)$ is expanded as a linear combination of basis functions $s_h(\cdot)$, $h \in \llbracket 1, H \rrbracket$ such as

$$\alpha(\cdot) = \sum_{h=1}^H c_h s_h(\cdot), \quad (2)$$

where the coefficients c_h are unknown whereas the number H of basis functions is supposed to be known. Introducing $c = (c_1, \dots, c_H)^\top$ and $s(\cdot) = (s_1(\cdot), \dots, s_H(\cdot))^\top$, model (1) can be rewritten as

$$X = \mu + s^\top(Y)cVb + \varepsilon, \quad \text{with } \varepsilon \sim \mathcal{N}(0, V). \quad (3)$$

The final set of parameters is $\theta := (\mu, V, c, b) \in \Theta := \mathbb{R}^p \times \mathcal{S}_{++}^p \times \mathbb{R}^H \times \mathbb{R}^p$. Let $\mathcal{D}_n := \{(X_i, Y_i), i \in \llbracket 1, n \rrbracket\}$ be a set of independent random variables with the same distribution as (X, Y) defined in (1). Let us denote by \mathcal{S} the $n \times H$ matrix such that $\mathcal{S}_{ih} = s_h(Y_i) - \bar{s}_h$ and \mathcal{X} the $n \times p$ matrix such that $\mathcal{X}_{ij} = (X_i)_j - \bar{X}_j$, $i \in \llbracket 1, n \rrbracket$, $j \in \llbracket 1, p \rrbracket$ and $h \in \llbracket 1, H \rrbracket$, where

$$\bar{s} = \frac{1}{n} \sum_{i=1}^n s(Y_i) \quad \text{and} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i. \quad (4)$$

Introducing the following empirical covariance matrices:

$$M = \frac{1}{n} \mathcal{S}^\top \mathcal{X}, \quad W = \frac{1}{n} \mathcal{S}^\top \mathcal{S} \quad \text{and} \quad \Sigma = \frac{1}{n} \mathcal{X}^\top \mathcal{X},$$

the likelihood of \mathcal{D}_n is given by

$$\mathcal{L}(\mathcal{D}_n | \theta) \propto |V|^{-n/2} \exp \left(-\frac{n}{2} \text{tr}(\Sigma V^{-1}) - \frac{n}{2} (c^\top W c) (b^\top V b) + n c^\top M b \right) \quad (5)$$

$$\cdot \exp \left(-\frac{n}{2} (\mu - \bar{X} + \bar{s}^\top c V b)^\top V^{-1} (\mu - \bar{X} + \bar{s}^\top c V b) \right), \quad (6)$$

where $|\cdot|$ is the determinant, see [BMGG09, Lemma 1]. The maximum likelihood estimator (MLE) of (μ, V, c, b) is established by [BMGG09, Proposition 1], in particular it is proved that \hat{b} is the eigenvector associated with the largest eigenvalue of $\Sigma^{-1}M^\top W^{-1}M$. See [CFG17] for an adaptation to the case where ε is Student distributed. These results rely on the assumption that Σ is of full rank which may not be the case in practice, especially if the dimension p is large. The introduction of Gaussian priors is then investigated in [BMGG09] and it is shown that the maximum *a posteriori* estimator of b can be calculated in closed form, even when Σ is singular.

2.2 A particular and useful case, the SIR model

Suppose the range Y is partitioned in $H + 1$ non-overlapping slices denoted by Slice_h and consider the associated basis functions defined by $s_h(\cdot) = \mathbf{1}\{\cdot \in \text{Slice}_h\}$, $h \in \llbracket 1, H + 1 \rrbracket$, where $\mathbf{1}\{\cdot\}$ is the indicator function. Let n_h be the number of Y_i in Slice_h for all $h \in \llbracket 1, H + 1 \rrbracket$. In that context, $M^\top W^{-1}M = \Gamma$ where Γ is the estimation of the covariance matrix of the regression curve $\text{Cov}(\mathbb{E}(X|Y))$ defined by

$$\Gamma = \sum_{h=1}^H \frac{n_h}{n} (\bar{X}_h - \bar{X})(\bar{X}_h - \bar{X})^\top \text{ with } \bar{X}_h = \frac{1}{n_h} \sum_{Y_i \in \text{Slice}_h} X_i. \quad (7)$$

As a consequence, the MLE of b coincides with the SIR estimator [BMGG09, Corollary 1].

3 Homogeneous Penalization of SIR (HoPSIR)

Throughout, $\pi(\cdot)$ denotes a prior distribution while $p(\cdot)$ denotes a posterior distribution.

3.1 Some multivariate and matrix-variate distributions

Matrix Generalized Inverse Gaussian distribution. Let us introduce the Matrix Generalized Inverse Gaussian distribution [But98, FB16] (\mathcal{MGIG} in the following) of parameters $(\Phi, \Psi) \succeq 0$ and $\lambda > 0$. Then the $\mathcal{MGIG}(\Phi, \Psi, \lambda)$ density writes $\forall W \succ 0$:

$$f_{\mathcal{MGIG}}(W|\Phi, \Psi, \lambda) = a_p(\Phi, \Psi, \lambda)^{-1} |W|^{\lambda-(p+1)/2} \exp\left(-\frac{1}{2} \text{tr}(\Phi W^{-1} + \Psi W)\right) \mathbf{1}\{W \succ 0\}, \quad (8)$$

with $a_p(\Phi, \Psi, \lambda) = |\Phi/2|^{-\lambda} B_\lambda(\Psi\Phi/4)$, and where B_λ is the matrix Bessel function of the second kind defined through

$$B_\lambda\left(\frac{1}{4}\Psi\Phi\right) = \left|\frac{\Psi}{2}\right|^{-\lambda} \int_{\mathcal{S}_{++}^p} |S|^{-\lambda-(p+1)/2} \exp\left(-\frac{1}{2} \text{tr}(\Phi S^{-1} + \Psi S)\right) dS.$$

This distribution is defined, for a fixed value for λ , on the domain of variation of Φ and Ψ :

$$\left. \begin{array}{ll} \{\Phi \succeq 0, \Psi \succ 0\} & \text{if } \lambda \geq p/2 \\ \{\Phi \succ 0, \Psi \succ 0\} & \text{if } -(p-1)/2 \leq \lambda < p/2 \\ \{\Phi \succ 0, \Psi \succeq 0\} & \text{if } \lambda < -(p-1)/2 \end{array} \right\} \quad (p \geq 2).$$

As a particular case, $\mathcal{W}^{-1}(\Psi, \nu)$ is the inverse-Wishart distribution and corresponds to the case $\{\Phi = 0, \Psi \succ 0, \lambda \geq p/2\}$. The parameter λ is classically abandoned in favor of $\nu = -2\lambda$ and its density is given by

$$\begin{aligned} f_{\mathcal{W}^{-1}}(W|\Psi \succ 0, \nu = -2\lambda) &= f_{\mathcal{MGIG}}(W|\Phi = 0, \Psi \succ 0, \lambda = -\nu/2), \\ &= \frac{|\Psi|^{\nu/2}}{2^{p\nu/2}\Gamma_p(\nu/2)} |W|^{-\frac{\nu+p+1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\Psi W^{-1})\right), \end{aligned}$$

where $\Gamma_p(\cdot)$ is the multivariate gamma function, see [GCSR95] for further details. The special case $p = 1$ corresponds to the generalized inverse Gaussian distribution, $\mathcal{GIG}(\phi, \psi, \lambda)$ in the following. Its form corresponds to (8) with norming constant

$$a_1(\phi, \psi, \lambda) = 2K_\lambda(\sqrt{\phi\psi})(\phi/\psi)^{\lambda/2}$$

where K_λ is the modified Bessel function of the second order of parameter λ . The domains of variation here writes

$$\begin{array}{ll} \{\phi \geq 0, \psi > 0\} & \text{if } \lambda > 0 \\ \{\phi > 0, \psi > 0\} & \text{if } \lambda = 0 \\ \{\phi > 0, \psi \geq 0\} & \text{if } \lambda < 0. \end{array}$$

Also, the Gamma distribution noted $\text{Gamma}(r, \delta)$ with shape $r > 0$ and rate $\delta > 0$ is a particular case of \mathcal{MGIG} distribution with $p = 1$ and $\{\phi \geq 0, \psi > 0, \lambda > 0\}$. Indeed if $\kappa \sim \text{Gamma}(r > 0, \delta > 0)$, its density, $\forall \kappa > 0$ writes

$$f_{\text{Gamma}}(\kappa|r, \delta) = f_{\mathcal{MGIG}}(\kappa|\Phi = 0, \Psi = 2\delta, \lambda = r) = \frac{\kappa^{r-1}}{\Gamma(r)} \delta^r \exp(-\delta\kappa).$$

Multivariate distributions based on homogeneous functions. Let us denote by $S^{p-1} = \{v \in \mathbb{R}^p, \|v\|_2 = 1\}$ the unit sphere in \mathbb{R}^p . Let $\alpha \neq 0$. For all positive α -homogeneous function $\varphi : \mathbb{R}^p \rightarrow \mathbb{R}^+$, let us consider the distribution associated with the density defined for all $x \in \mathbb{R}^p$ by

$$\pi_\varphi(x) = \frac{1}{\Gamma(1+p/\alpha)} \frac{1}{I_p(\varphi)} \exp(-\varphi(x)), \quad \text{with } I_p(\varphi) = \int_{S^{p-1}} (\varphi(u))^{-p/\alpha} du,$$

and where $\Gamma(\cdot)$ is the gamma function.

The family of distributions, parameterized with the α -homogeneous function φ is closed to the exponential family distribution. It differs from this since it must verifies the homogeneous property, let us denote it the homogeneous exponential family, $\mathcal{HE}(\alpha, p)$ in the following. In that context the centered multivariate Gaussian distribution is indeed an element of $\mathcal{HE}(\alpha, p)$

but any non centered multivariate Gaussian distribution is not. The following Table 1 gathers different usual distributions which are members of \mathcal{HE} .

Name of the distribution	Definition Set	$\varphi(x)$	Associated parameters	α
Multivariate centered Gaussian	\mathbb{R}^p	$x^\top C^{-1}x/2$	$C \in \mathcal{S}_{++}^p$	2
Centered Laplace	\mathbb{R}^p	$\ x\ _1/L$	$L > 0$	1

Table 1: Some usual distributions in the $\mathcal{HE}(\alpha, p)$ family

3.2 A new family of priors

The following prior is considered on $\theta = (\mu, V, c, b, \kappa)$:

$(\mu, V, c, b, \kappa) | (\Psi, \nu, f, g) \sim \mathcal{U}(\mathbb{R}^p) \otimes \mathcal{W}^{-1}(\Psi, \nu) \otimes \pi(c | \kappa, f, g) \otimes \pi(b | \kappa, f, g) \otimes \mathcal{U}(\mathbb{R}^+ \setminus \{0\})$, with prior parameters $\lambda := (\Psi, \nu, f, g) \in \Lambda := \mathcal{S}_{++}^p \times \llbracket p+1, \infty \rrbracket \times \mathcal{HE}(\alpha_f, H) \times \mathcal{HE}(\alpha_g, p)$. Here, $\mathcal{U}(E)$ is the uniform distribution on the set E , this distribution being improper when E is not bounded. The prior distributions of $c | \kappa, f, g$ and $b | \kappa, f, g$ are based on the multivariate distribution defined in Subsection 3.1 with respectively $\varphi_c : c \mapsto \kappa \alpha_g f(c)$ and $\varphi_b : b \mapsto \kappa \alpha_f g(b)$ leading to

$$\begin{aligned} \pi(c | \kappa, f, g) &= \pi_{\varphi_c}(c) = \frac{(\kappa \alpha_g)^{H/\alpha_f}}{\Gamma(1 + H/\alpha_f)} \frac{1}{I_H(f)} \exp(-\kappa \alpha_g f(c)), \\ \pi(b | \kappa, f, g) &= \pi_{\varphi_b}(b) = \frac{(\kappa \alpha_f)^{p/\alpha_g}}{\Gamma(1 + p/\alpha_g)} \frac{1}{I_p(g)} \exp(-\kappa \alpha_f g(b)), \end{aligned}$$

where $\kappa > 0$, $f \in \mathcal{HE}(\alpha_f, H)$, $g \in \mathcal{HE}(\alpha_g, p)$ and with $\alpha_f, \alpha_g > 0$. The complete prior on the parameter $\theta = (\mu, V, c, b, \kappa)$ can be equivalently rewritten as:

$$\begin{aligned} \pi(\theta | \lambda) &= \pi(\mu) \pi(V | \Psi, \nu) \pi(c | \kappa, f, g) \pi(b | \kappa, f, g) \pi(\kappa) \\ &= \frac{\kappa^{\frac{H}{\alpha_f} + \frac{p}{\alpha_g} - 1} \alpha_g^{\frac{H}{\alpha_f}} \alpha_f^{\frac{p}{\alpha_g}} |V|^{-\frac{\nu+p+1}{2}}}{\Gamma(1 + H/\alpha_f) I_H(f) \Gamma(1 + p/\alpha_g) I_p(g)} \exp\left(-\frac{\text{tr}(\Psi V^{-1})}{2} - \kappa \alpha_g f(c) - \kappa \alpha_f g(b)\right). \end{aligned} \tag{9}$$

3.3 Posterior distributions

From Bayes theorem, the posterior distribution of \mathcal{D}_n given (μ, V, c, b, κ) is proportional to the likelihood multiplied by the associated prior. Combining (5) and (9), it thus follows:

$$\begin{aligned} p(\theta | \mathcal{D}_n, \lambda) &\propto \mathcal{L}(\mathcal{D}_n | \theta) \pi(\theta | \lambda) \\ &\propto |V|^{-\frac{\nu+p+n+1}{2}} \exp\left(-\frac{1}{2} \text{tr}((n\Sigma + \Psi) V^{-1}) - \frac{1}{2} n(c^\top Wc)(b^\top Vb)\right) \\ &\quad \cdot \exp(nc^\top Mb - \kappa \alpha_g f(c) - \kappa \alpha_f g(b)) \\ &\quad \cdot \exp\left(-\frac{n}{2} (\mu - \bar{X} + \bar{s}^\top cVb)^\top V^{-1} (\mu - \bar{X} + \bar{s}^\top cVb)\right), \\ &\quad \cdot \kappa^{\frac{H}{\alpha_f} + \frac{p}{\alpha_g} - 1}. \end{aligned} \tag{10}$$

One might recognize known proper distributions of the conditional posteriors on the parameters μ , V , c , b and κ . The authors will present their developments on this subject and propose an estimation algorithm of the maxima *a posteriori*. The performance of the proposed HoPSIR method will be compared to state of the art regularizations methods [LZL19, LY08, LCTS21].

References

- [BMGG08] C. Bernard-Michel, L. Gardes, and S. Girard. A note on sliced inverse regression with regularizations. *Biometrics*, 64(3):982–984, 2008.
- [BMGG09] C. Bernard-Michel, L. Gardes, and S. Girard. Gaussian regularized sliced inverse regression. *Statistics and Computing*, 19(1):85–98, 2009.
- [BT09] W. Bian and D. Tao. Manifold regularization for SIR with rate root- n convergence. *Advances in Neural Information Processing Systems*, 22:117–125, 2009.
- [But98] R. W. Butler. Generalized inverse Gaussian distributions and their Wishart connections. *Scandinavian Journal of Statistics*, 25(1):69–75, 1998.
- [CFG17] A. Chiancone, F. Forbes, and S. Girard. Student sliced inverse regression. *Computational Statistics & Data Analysis*, 113:441–456, 2017.
- [CLS14] R. Coudret, B. Lique, and J. Saracco. Comparison of Sliced Inverse Regression approaches for underdetermined cases. *Journal de la Société Française de Statistique*, 155(2):72–96, 2014.
- [Coo00] R. D. Cook. SAVE: a method for dimension reduction and graphics in regression. *Communications in Statistics - Theory and Methods*, 29:2109–2121, 2000.
- [Coo07] R. D. Cook. Fisher Lecture: Dimension Reduction in Regression. *Statistical Science*, 22(1):1 – 26, 2007.
- [DL91] N. Duan and K.-C. Li. Slicing regression: a link-free regression method. *The Annals of Statistics*, 19:505–530, 1991.
- [EAS98] A. Edelman, T. Arias, and S. Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.
- [FB16] F. Fazayeli and A. Banerjee. The matrix generalized inverse Gaussian distribution: Properties and applications. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 648–664. Springer, 2016.
- [GCSR95] A. Gelman, J. Carlin, H. Stern, and D. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

- [GLS22] S. Girard, H. Lorenzo, and J. Saracco. Advanced topics in Sliced Inverse Regression. *Journal of Multivariate Analysis*, 188:104852, 2022.
- [LCTS21] H. Lorenzo, O. Cloarec, R. Thiébaud, and J. Saracco. Data-driven sparse partial least squares. *Statistical Analysis and Data Mining*, 2021.
- [LY08] L. Li and X. Yin. Sliced inverse regression with regularizations. *Biometrics*, 64(1):124–131, 2008.
- [LZL19] Q. Lin, Z. Zhao, and J. Liu. Sparse Sliced Inverse Regression via lasso. *Journal of the American Statistical Association*, 114(528):1726–1739, 2019.