Contents

1 Introduction **3**

2 Dimension Reduction **4**

2.1 Sliced Inverse Regression 4

2.2 Sliced Average Variance Estimation 5

3 Modification for SAVE **7**

3.1 Identity 7

3.2 Estimation 9

4 Permutation Tests **10**

5 Simulations **12**

5.1 Example 1 13

5.2 Example 2 15

5.3 Example 3 17

6 Conclusion **19**

Dimension Reduction for Censored Regression Data via Sliced Average Variance Estimation

Advisor: Dr. Heng-Hui Lue   Student: Shih-Chia Liu

Department of Statistics, Tunghai University, Taiwan

hhlue@thu.edu.tw

g934701@thu.edu.tw

Abstract

With censoring, high-dimensional regression becomes much more complicated. Since censoring can cause severe bias in estimation, modification to adjust such bias is needed to be made. Under the conditionally independent censoring condition, we propose the modification of sliced average variance estimation (SAVE) for estimating the joint effective dimension reduction (e.d.r.) space of lifetime and censoring time. Several simulation examples are reported and comparisons are made with the sliced inverse regression method of Li, Wang and Chen (1999).

1. Introduction

With censoring, there are several methods and techniques have been suggested, including censored linear regression, the Cox model and many others. In this article, we shall show how censored regression data can be analyzed without assuming the functional from a priori and offer methods of finding the data structure.

For these difficulties, we consider the dimension reduction model as given in Li (1992):

$$Y = g(\beta_1' \mathbf{x}, \cdots, \beta_k' \mathbf{x}, \varepsilon) \tag{1.1}$$

where $\beta_j' s$ are unknown vectors to be estimated from data, **x** is a p-dimensional predictor vector, g is a completely unknown link function and the error term ε is independent of **x**. The space spanned by these $\beta_j' s$ is called the e.d.r. (effective dimension reduction) space and any vector in this space is referred to as an e.d.r. direction and $\nu' \mathbf{x}$, a linear combination of **x**, is called the e.d.r. variate.

To incorporate censoring into the dimension reduction framework, let

$$\begin{aligned} Y^0 &= \text{the true lifetime (unobservable)}, \\ C &= \text{the censoring time}, \\ \delta &= \text{the censoring indicator; } \delta = 1, \text{if } Y^0 \leq C \text{ and } \delta = 0, \text{otherwise}, \\ Y &= \min\{Y^0, C\}, \text{the observed time}. \end{aligned}$$

We assume that

$$Y^0 \text{ follows model (1.1);} \tag{1.2}$$

$$\text{Conditional on } \mathbf{x},\ C \text{ is independent of } Y^0. \tag{1.3}$$

The observed sample consists of n i.i.d. observations, $(Y_i, \mathbf{x}_i, \delta_i),\ i = 1, \ldots, n$ from the distribution of $(Y, \mathbf{x}, \delta)$.

In section 2, we introduce some dimension reduction methods. Since censoring can cause severe bias in estimation, modification to adjust such bias is needed to be made. We propose the modification of sliced average variance estimation under the condition (1.3) in section 3. In section 4, we want to make inference about the dimension of e.d.r. space and propose a permutation test to do it. Several simulation examples are reported and comparisons are made with the sliced inverse regression method in section 5.

2. Dimension Reduction

2.1. Sliced Inverse Regression

Li (1991) proposed a data-analytic tool, sliced inverse regression (SIR), for dimension reduction. Instead of regressing the univariate output variable Y against $\mathbf{x}$, the method explores the simplicity of the inverse view, $\mathbf{x}$ against Y.

Denote the inverse regression curve by $\eta(y) = E(\mathbf{x}|Y = y)$, the population version of SIR is based on the following eigenvalue decomposition:

$$\Sigma_\eta \alpha_i = \tau_i \Sigma_{\mathbf{x}} \alpha_i,$$

$$\tau_1 \geq \cdots \geq \tau_p,$$

where $\Sigma_\eta = cov(E(\mathbf{x}|Y = y))$ and $\Sigma_{\mathbf{x}} = cov(\mathbf{x})$.

The justification for using the first k eigenvectors α_i with nonzero eigenvalue to estimate the e.d.r. direction (Lemma 3.1 of Li, 1991).

Lemma 2.1. *Under the linear condition and the dimension reduction assumption (1.1) holds, then for any y, $\Sigma_{\mathbf{x}}^{-1}(\eta(y) - E(\mathbf{x}))$ falls into the e.d.r. space.*

To implement SIR on the data $(Y_i, \mathbf{x}_i),\ i = 1, \ldots, n$, we follow Li (1991). First we partition $y_i's$ into H slices and construct the slice mean $\bar{\mathbf{x}}_h$:

$$\bar{\mathbf{x}}_h = \frac{1}{n_h} \sum_{\mathbf{x}_i \in I_h} \mathbf{x}_i,$$

where n_h is the number of cases falling into I_h. Then the sample covariance matrix is

$$\hat{\Sigma}_\eta = \sum_{h=1}^{H} \frac{n_h}{n} (\bar{\mathbf{x}}_h - \bar{\mathbf{x}})(\bar{\mathbf{x}}_h - \bar{\mathbf{x}})'.$$

The sample version of SIR is based on the following eigenvalue decomposition:

$$\hat{\Sigma}_\eta \hat{\alpha}_i = \hat{\tau}_i \hat{\Sigma}_{\mathbf{x}} \hat{\alpha}_i,$$

where $\hat{\Sigma}_x$ is the sample covariance matrix, $\hat{\tau}_1 \geq \cdots \geq \hat{\tau}_p$.

2.2. Sliced Average Variance Estimation

Cook and Weisberg (1991) proposed sliced average variance estimation (SAVE) to overcome the inability of SIR to detect certain types of nonlinear regression relationships.

Consider standardization of the predictors. If $\Sigma_{\mathbf{x}} = cov(\mathbf{x})$ is positive definite, taking $\mathbf{z} = \Sigma_{\mathbf{x}}^{-1/2}(\mathbf{x} - E(\mathbf{x}))$, then the population version of SAVE is based on the following eigenvalue decomposition:

$$Mb_i = \lambda_i b_i,$$

$$\lambda_1 \geq \cdots \geq \lambda_p,, \tag{2.1}$$

where

$$M = E(I_p - cov(\mathbf{z}|Y = y))^2, \tag{2.2}$$

λ_i and b_i denote the eigenvalues and eigenvectors of M for $i = 1, \ldots, p$. Let k denote the number of nonzero eigenvalues, the eigenvectors corresponding to the k positive eigenvalues are estimated the e.d.r. directions. Under the linearity and constant covariance conditions, SAVE estimates e.d.r. directions (Cook and Weisberg 1991, Cook and Lee 1999).

Construct the sample mean $\bar{\mathbf{x}}$ and covariance matrix $\hat{\Sigma}_{\mathbf{x}}$ and then form the sample standardized predictors $\hat{\mathbf{z}}_i = \hat{\Sigma}_{\mathbf{x}}^{-1/2}(\mathbf{x}_i - \bar{\mathbf{x}}),\ i = 1, \ldots, n$. Divide the observed rang of response Y into H slices, then form

$$\hat{M} = \sum_{h=1}^{H} \frac{n_h}{n}(I_p - \hat{V}_h)^2,$$

where $\hat{V}_h$ is the sample version of $cov(\mathbf{z}|Y = y)$ for slice h. The sample version of SAVE is based on the following eigenvalue decomposition:

$$\hat{M}\hat{b}_i = \hat{\lambda}_i \hat{\Sigma}_{\mathbf{z}} \hat{b}_i,$$

$$\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p.$$

The j^{th} sample SAVE predictor can be constructed as

$$S_j = \hat{b}_j' \mathbf{z}_i,\ j = 1, \ldots, p,\ i = 1, \ldots, n,$$

where $\hat{b}_j$ is the j^{th} eigenvector of $\hat{M}$ corresponding to the j^{th} nonzero eigenvalue.

3. Modification for SAVE

Because the true lifetime Y^0 is unobservable, the key observations come from an identity derived in section 3.1, which relates the conditional expectation to the observed time Y and δ. We develop an method under the conditional independence (1.3) and LDC. In section 3.2, we used this identities to modify by a suitable weighting function for offsetting the censoring bias in estimating.

3.1. Identity

We known that $cov(\mathbf{z}|Y) = E(\mathbf{z}\mathbf{z}'|Y) - E(\mathbf{z}|Y)E(\mathbf{z}'|Y)$. Consider $0 = t_1 < t_2 < \cdots < t_H < \infty = t_{H+1}$. Let $\mathbf{m}_j^1 = E\{\mathbf{z}|Y^0 \in [t_j, t_{j+1})\}$ be the expected value of **z** in j^{th} slice. Li, Wang and Chen (1999) proved the following equality

$$\mathbf{m}_j^1 = \frac{E\{\mathbf{z}\mathbf{1}(Y^0 \in [t_j, t_{j+1}))\}}{P\{Y^0 \in [t_j, t_{j+1})\}} = \frac{E\{\mathbf{z}\mathbf{1}(Y^0 \geq t_j)\} - E\{\mathbf{z}\mathbf{1}(Y^0 \geq t_{j+1})\}}{E\{\mathbf{1}(Y^0 \geq t_j)\} - E\{\mathbf{1}(Y^0 \geq t_{j+1})\}}, \tag{3.1}$$

where

$$E\{\mathbf{z}\mathbf{1}(Y^0 \geq t)\} = E\{\mathbf{z}\mathbf{1}(Y \geq t)\} + E\{\mathbf{z}\mathbf{1}(Y < t, \delta = 0)\}w(Y, t, \mathbf{z}), \tag{3.2}$$

and

$$E\{\mathbf{1}(Y^0 \geq t)\} = E\{\mathbf{1}(Y \geq t)\} + E\{\mathbf{1}(Y < t, \delta = 0)\}w(Y, t, \mathbf{z}). \tag{3.3}$$

For more details, see Li, Wang and Chen, (1999).

The weight function can be further expressed as

$$w(t', t, \mathbf{z}) = \frac{S^0(t|\mathbf{z})}{S^0(t'|\mathbf{z})} = exp\{-\Lambda(t', t|\mathbf{z})\} \tag{3.4}$$

where for $t' < t$,

$$\Lambda(t', t|\mathbf{z}) = E\left\{\frac{\mathbf{1}(t' < Y < t, \delta = 1)}{S_Y(Y|\mathbf{z})}\middle|\mathbf{z}\right\},$$

$$S^0(t|\mathbf{z}) = P\{Y^0 \geq t|\mathbf{z}\}$$

$$= \text{conditional survival for } Y^0 \text{ given } \mathbf{z},$$

$$S_Y(\cdot|\mathbf{z}) = \text{the conditional survival function of } Y \text{ conditional on } \mathbf{z}.$$

Let $\mathbf{m}_j^2 = E\{zz^{'}|Y^0 \in [t_j, t_{j+1})\}$. Using the same argument, we obtain

$$\mathbf{m}_j^2 = \frac{E\{\mathbf{z}\mathbf{z}^{'}\mathbf{1}(Y^0 \geq t_j)\} - E\{\mathbf{z}\mathbf{z}^{'}\mathbf{1}(Y^0 \geq t_{j+1})\}}{E\{\mathbf{1}(Y^0 \geq t_j)\} - E\{\mathbf{1}(Y^0 \geq t_{j+1})\}}, \tag{3.5}$$

where

$$E\{\mathbf{z}\mathbf{z}^{'}\mathbf{1}(Y^0 \geq t)\} = E\{\mathbf{z}\mathbf{z}^{'}\mathbf{1}(Y \geq t)\} + E\{\mathbf{z}\mathbf{z}^{'}\mathbf{1}(Y < t, \delta = 0)\}w(Y, t, \mathbf{z}). \tag{3.6}$$

Plugging (3.1)-(3.6) into (2.2), then we can conduct the eigenvalue decomposition (2.1).

3.2. Estimation

We describe the estimates for $\mathbf{m}_j^1$ and $\mathbf{m}_j^2$ as follows: By (3.1), (3.2) and (3.5), (3.6),

$$\hat{\mathbf{m}}_j^1 = \frac{\hat{E}\{\mathbf{z}\mathbf{1}(Y^0 \geq t_j)\} - \hat{E}\{\mathbf{z}\mathbf{1}(Y^0 \geq t_{j+1})\}}{\hat{P}\{Y^0 \geq t_j\} - \hat{P}\{Y^0 \geq t_{j+1}\}}, \tag{3.7}$$

$$\hat{\mathbf{m}}_j^2 = \frac{\hat{E}\{\mathbf{z}\mathbf{z}'\mathbf{1}(Y^0 \geq t_j)\} - \hat{E}\{\mathbf{z}\mathbf{z}'\mathbf{1}(Y^0 \geq t_{j+1})\}}{\hat{P}\{Y^0 \geq t_j\} - \hat{P}\{Y^0 \geq t_{j+1}\}}, \tag{3.8}$$

$$\hat{E}\{\mathbf{z}\mathbf{1}(Y^0 \geq t)\} = n^{-1}\sum_{i:\ Y_i \geq t}^{n} \mathbf{z}_i + n^{-1}\sum_{i:\ Y_i < t,\ \delta_i = 0}^{n} \mathbf{z}_i \hat{w}(Y_i, t, \mathbf{z}_i), \tag{3.9}$$

$$\hat{E}\{\mathbf{z}\mathbf{z}'\mathbf{1}(Y^0 \geq t)\} = n^{-1}\sum_{i:\ Y_i \geq t}^{n} \mathbf{z}_i\mathbf{z}_i' + n^{-1}\sum_{i:\ Y_i < t,\ \delta_i = 0}^{n} \mathbf{z}_i\mathbf{z}_i' \hat{w}(Y_i, t, \mathbf{z}_i), \tag{3.10}$$

$$\hat{P}\{Y^0 \geq t\} = \sharp\{i:\ Y_i \geq t\}/n + n^{-1}\sum_{i:\ Y_i < t,\ \delta_i = 0}^{n} \hat{w}(Y_i, t, \mathbf{z}_i). \tag{3.11}$$

By (3.7) and (3.8), construct

$$\hat{M}^0 = \sum_{j=1}^{H} \hat{p}_j\{I_p - [\hat{\mathbf{m}}_j^2 - \hat{\mathbf{m}}_j^1 \hat{\mathbf{m}}'^1_j]\}^2,$$

$$\hat{p}_j = \hat{P}\{Y^0 \geq t_j\} - \hat{P}\{Y^0 \geq t_{j+1}\}.$$

Now, we can conduct the eigenvalue decomposition to find the SAVE directions

$$\hat{M}^0 \hat{b}_i^0 = \hat{\lambda}_i \hat{\Sigma}_{\mathbf{z}} \hat{b}_i^0,$$

$$\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p. \tag{3.12}$$

To estimate $w(t', t, \mathbf{z})$, we needed using the smoothing technique. We only consider the kernel smoothing method here for simplicity. Let $K_p(\cdot)$ be a kernel function on R^p and h_n be the bandwidth in each coordinate. We shall assume that $h_n = o(1)$ and nh_n^p tends to infinity. For some one-dimensional kernel function $K(\cdot)$, it is common for $K_p(\cdot)$

to take a product form, $K_p(z_1, \cdots, z_p) = K(z_1) \cdots K(z_p)$. Our kernel estimate is defined by

$$\hat{\Lambda}(t', t|\mathbf{z}) = \frac{n^{-1} \sum_{i:\ t'<Y_i<t,\ \delta_i=1}^{n} (\hat{S}_Y(Y_i|\mathbf{z}_i))^{-1} h_n^{-p} K_p(h_n^{-1}(\mathbf{z}_i - \mathbf{z}))}{\hat{f}(\mathbf{z})}, \tag{3.13}$$

$$\hat{S}_Y(Y_i|\mathbf{z}_i) = \frac{n^{-1} \sum_{j:\ Y_j>Y_i}^{n} h_n^{-p} K_p(h_n^{-1}(\mathbf{z}_j - \mathbf{z}))}{\hat{f}(\mathbf{z}_i)}, \tag{3.14}$$

$$\hat{f}(\mathbf{z}) = n^{-1} \sum_{i}^{n} h_n^{-p} K_p(h_n^{-1}(\mathbf{z}_i - \mathbf{z})). \tag{3.15}$$

However, the kernel smoothing only works well in the low-dimensional case. Thus, we must reduce the dimensionality first before applying the kernel smoothing technique in estimating the weight function.

4. Permutation Tests

In this section, we want to make inference about the dimension of the e.d.r. space. Issues encountered when making inference about the dimension may weaken some methods in practice. The asymptotic distributions of the various test statistics are generally linear combinations of chi-squared variables with unknown coefficients, reducing to simper chi-squared distributions with normal predictors. Even with a simple chi-squared asymptotic distribution, these may be concern about the accuracy of the approximation it provides.

Li (1991) proposed a chi-squared test for determining the number of the significant e.d.r. directions obtained by SIR. Cook and Weisberg (1991) propose simple permutation tests that can bypass some of these issues and generalize the permutation tests.

Let $U = [u_j]$ be the $p \times p$ matrix of eigenvectors u_j of the population kernel matrix

M, and assume $S(M) = S_{Y|\mathbf{z}}$ the e.d.r. space. Denote $d_{Y|\mathbf{z}}$ be the dimension of the e.d.r. space. Consider testing the hypothesis that $d_{Y|\mathbf{z}} \leq m$ versus $d_{Y|z} > m$. Partition $U = (U_1, U_2)$ where U_1 is $p \times m$ so that under the null hypothesis $S(U_1) \supseteq S_{Y|\mathbf{z}}$. Cook and Yin (2001) proposed the following proposition for constructing a permutation test and for inference on $d_{Y|\mathbf{z}}$.

Proposition 4.1. *Let U be constructed as indicated previously.*

(a) If $(Y, U_1'\mathbf{z}) \perp U_2'\mathbf{z}$ then $S(U_1)$ is a D.R.S. (dimension reduction subspace) for the regression Y on $\mathbf{z}$. Therefore $dim(S(U_1)) \geq d_{Y|\mathbf{z}}$.

(b) Assume that $U_1'\mathbf{z} \perp U_2'\mathbf{z}$. Then $S(U_1)$ is a D.R.S. for the regression of Y on $\mathbf{z}$ if and only if $(Y, U_1'\mathbf{z}) \perp U_2'\mathbf{z}$.

(c)Assume that $U_1'\mathbf{z} \perp U_2'\mathbf{z}|Y$. If $U_2'\mathbf{z} \perp Y$ then $S(U_1)$ is a D.R.S. for the regression of Y on $\mathbf{z}$.

Application of Proposition 4.1 to test the hypothesis that $d_{Y|\mathbf{z}} \leq m$ in practice involves four general steps.

1. Compute the sample kernel matrix $\hat{M}$ for SIR or SAVE and from the matrices of its eigenvectors $\hat{U}_1 = (\hat{u}_1, \ldots, \hat{u}_m)$ and $\hat{U}_2 = (\hat{u}_{m+1}, \ldots, \hat{u}_p)$.

2. Construct the vectors of sample principal predictors $\hat{V}_{1i} = \hat{U}_1'\hat{\mathbf{z}}_i$ and $\hat{V}_{2i} = \hat{U}_2'\hat{\mathbf{z}}_i, i = 1, \ldots, n$.

3. Randomly permute the indices i of the $\hat{V}_{2i}$ to obtain the permuted set $\hat{V}_{2i}^*$.

4. Construct the test statistic $\hat{\Lambda}_m^*$ based on the original data Y_i, $\hat{V}_{1i}$ along with the permuted data $\hat{V}_{2i}^*,\ i = 1, \ldots, n$.

After repeating the steps 3, 4 a number of times, and computing the permutation *P*-value. Repeating steps 1-4 for $m = 0, \ldots, p-1$ gives the series of *P*-values. This simple test can be able to be quite useful in practice. In next section, we used this procedure with SAVE for making inference about the dimension of the e.d.r. space.

5. Simulations

In order to find d.r. directions we propose the following two-stage procedure:

1. Apply double slicing on (Y, δ) and find the joint e.d.r. directions, $\hat{b}_{di}$. Let $\hat{B}_r = (\hat{b}_{d1}, \ldots, \hat{b}_{dr})$ be the first r significant directions.

2. Apply the r-dimensional kernel smoothing on $\hat{B}_r'\mathbf{z}$, and use that for estimating the weight function $\hat{w}$,

$$\hat{w}(t', t|\mathbf{z}) = exp\{-\hat{\Lambda}(t', t|\mathbf{z})\}, \tag{5.1}$$

where

$$\hat{\Lambda}(t', t|\mathbf{z}) = \frac{n^{-1}\sum_{i:\ t'<Y_i<t,\ \delta_i=1}^{n}(\hat{S}_Y(Y_i|\mathbf{z}_i))^{-1}h_n^{-r}K_r(h_n^{-1}(\hat{B}_r(\mathbf{z}_i - \mathbf{z})))}{\hat{f}(\mathbf{z})}, \tag{5.2}$$

$$\hat{S}_Y(Y_i|\mathbf{z}_i) = \max\left\{\frac{n^{-1}\sum_{j:\ Y_j>Y_i}^{n} h_n^{-r}K_r(h_n^{-1}(\hat{B}_r(\mathbf{z}_j - \mathbf{z})))}{\hat{f}(\mathbf{z}_i)}, c\right\}, \tag{5.3}$$

$$\hat{f}(\mathbf{z}) = n^{-1}\sum_{i}^{n} h_n^{-r}K_r(h_n^{-1}(\hat{B}_r(\mathbf{z}_i - \mathbf{z}))). \tag{5.4}$$

Note that prevent $\hat{S}_Y(Y_i|\mathbf{z}_i)$ from being too small, make use of a constant c to replace it (set to 0.05 in following examples). Apply (3.7)∼(3.11) to estimate the true lifetime e.d.r.

directions after estimating the weight function. To illustrate how this strategy works, we report some simulation studies.

5.1. Example 1

Take $p = 6$ and let $\tilde{\beta}_1 = (1, -1, 0, 0, 0, 0)'$ and $x = (x_1, \ldots, x_6)'$ be generated from the standard normal distribution. Suppose

$$Y^0 = |x_1 - x_2| + \sigma_1 \varepsilon_1,$$

$$C = 0.5 + \sigma_2 \varepsilon_2 \ \text{ for } x_3 > 0,$$

$$= 10 \ \text{ otherwise},$$

where $\sigma_1 = \sigma_2 = 0.1$. Here $\varepsilon_1, \ \varepsilon_2$ are normal random variables. Generate 300 cases. There are 104 cases censored in the data set. Applying double slicing with the number of slices equal to 5 for the censored and the uncensored groups. The first two SAVE directions, (-.72,.69,-.05,.02,-.06,.03)′ and (.02,.11,.99,-.05,.01,-.01)′, are close to the joint d.r. directions.

Table 1. Mean and standard deviation for the eigenvalues and the coefficients in the leading direction for Example 1 by modified SAVE in 100 runs

eigenvalues	.76(.15)	.31(.06)	.22(.04)	.18(.03)	.14(.02)	.11(.02)
$\hat{b}_1$	.692(.06)	-.693(.06)	-.021(.11)	.001(.08)	-.003(.08)	.001(.08)

With the weight adjustment and the smoothing method, we apply modified SAVE for estimating the true lifetime directions $\beta_1 = \tilde{\beta}_1/\|\tilde{\beta}_1\|$ =(.71,-.71,0,0,0,0)′. By the permutation test we can get the permutation P-values (.00,.31,.65,.83,.94,.64).

Table 2. Performance of modified SAVE

and modified SIR of Example 1 with 100 runs

	Modified SAVE	Modified SIR
$\|\hat{cos}(\theta)\|$	.979(.02)	.097(.15)
P-value	.002(.01)	.832(.22)

The results from 100 runs are summarized in Table 1 and Table 2 which report statistics related to eigenvalues, $\hat{b}_1$ and $|\hat{cos}(\theta)|$, where θ denote the angle between the estimate and β_1. The mean of $\hat{b}_1$ is very close to β_1 and the mean of $|\hat{cos}(\theta)|$ is close to 1.

Table 3. $|\hat{cos}(\theta)|$ of modified SAVE and modified

SIR as the number of dimensions increases

under Example 1 with 100 runs

p	Modified SAVE	Modified SIR
6	.9799(.0175)	.1043(.1407)
10	.9505(.0711)	.0817(.1568)
15	.9174(.0725)	.1039(.1526)
20	.8219(.1855)	.0933(.1500)

5.2. Example 2

Let $\beta_1 = (1, 0, 0, 0, 0, 0)'$ and $x = (x_1, \cdots, x_6)$. $x_i \sim U(0, \theta),\ i = 1, 2$ and $x_3 \sim x_5$ from the standard normal distribution. The true survival time Y^0 and the censoring time C are generated from

$$Y^0 = \alpha\lambda_1(\lambda_1 x_1)^{\alpha-1} e^{-(\lambda_1 x_1)^\alpha} + \sigma\epsilon_1,$$

$$C = \alpha\lambda_2(\lambda_2 x_2)^{\alpha-1} e^{-(\lambda_2 x_2)^\alpha} + \sigma\epsilon_2.$$

Take $\lambda_1 = 0.3$, $\lambda_2 = 0.4$, $\alpha = 3$, $\sigma = 0.05$ and $\theta = 6$. We obtain 300 independent observations of (Y, δ); among them, 181 cases are censored.

First, we apply SAVE with double slicing procedure on Y and δ give two significant directions. The first two SAVE directions, (-.99,.05,.04,.07,.11,-.05)′ and (.05,.98,.06,-.14,.01,-.15)′, are close to the joint d.r. directions.

Table 4. Mean and standard deviation for the eigenvalues and the coefficients in the leading directions for Example 2 by modified SAVE in 100 runs

eigenvalues	.68(.11)	.29(.07)	.21(.03)	.16(.02)	.13(.02)	.09(.02)
$\hat{b}_1$	.945(.04)	-.017(.15)	.001(.16)	.021(.13)	-.007(.15)	-.006(.14)

We use these two directions with weight adjustment to apply modified SAVE for estimating the lifetime directions β_1. The permutation P-values are given by (.00,.37,.50,.56,.99,.95). In this example, we summarized the results from 100 runs in Tables 4 and Table 5. Our estimate is very close to the true lifetime direction.

Table 5. Performance of modified SAVE

and modified SIR of Example 2 with 100 runs

	Modified SAVE	Modified SIR
$\lvert \hat{cos}(\theta) \rvert$	.945(.04)	.325(.26)
P-value	.001(.01)	.730(.28)

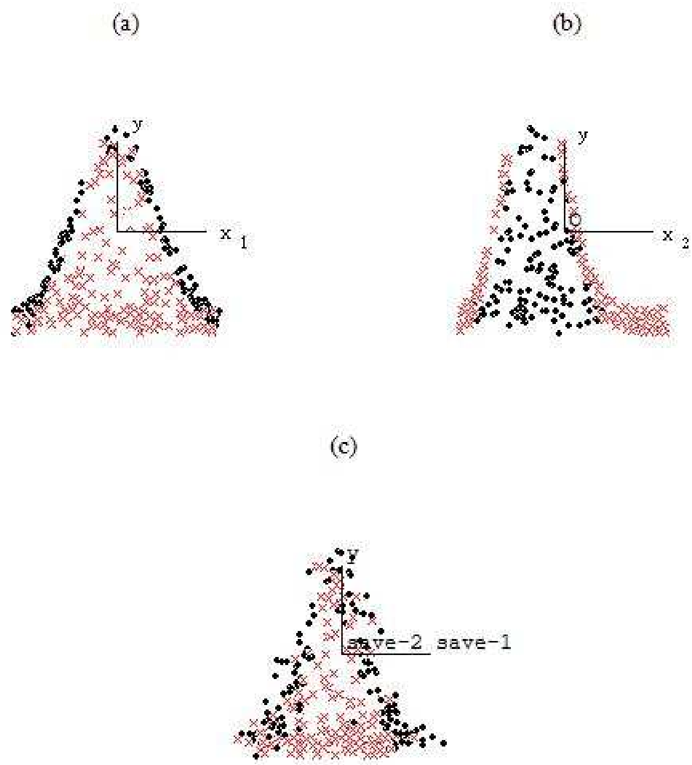

Figure 1. (a) The lifetime structure (b) The censoring pattern (c) Modified SAVE's view.

(● : true lifetime   × : censoring time)

Table 6. $|\hat{cos}(\theta)|$ of modified SAVE and modified

SIR as the number of dimensions increases

under Example 2 with 100 runs

p	Modified SAVE	Modified SIR
6	.9435(.0474)	.3392(.2546)
10	.8915(.0660)	.3451(.2356)
15	.8021(.1165)	.3038(.2281)
20	.6894(.1801)	.2802(.2091)

5.3. Example 3

Consider that the lifetime has two e.d.r. directions. In this example, we assume that the true survival time Y^0 follows an exponential distribution with the natural parameter equal to $\varepsilon^{2x_1^2}$ until time $\tau = 0.1$. From time τ on, the additional survival time follows the exponential distribution with the natural parameter $\varepsilon^{3x_2^2}$. More specifically, we assume

$$Y^* \sim \text{exponential with parameter } \varepsilon^{2x_1^2},$$

$$Y^{**} \sim \text{exponential with parameter } \varepsilon^{3x_2^2},$$

$$Y^0 = Y^*\mathbf{1}(Y^* < \tau) + (\tau + Y^{**})\mathbf{1}(Y^* > \tau).$$

The censoring time C follows an exponential distribution with parameter equal to $\varepsilon^{x_3^2}$.

Let $\beta_1 = (1, 0, 0, 0, 0, 0)'$ and $\beta_2 = (0, 1, 0, 0, 0, 0)'$. Again 300 independent observations of (Y, δ) are obtained. Among them, 100 cases are censored. The first three eigenvectors of the double slicing SAVE are (-.98,-.01,.18,.08,.03,-.06)′, (.18,.21,.94,.14,-.03,.14)′ and (.05,-.94,.17,.22,-.20,-.06)′. The eigenvalues are 1.14,.82,.47,.22,.21,.18.

Apply the modified SAVE, the first two significant directions are (.99,-.05,-.09,-.08,-.02,.05) and (.05,.94,.15,-.14,.24,.09). The eigenvalues are 1.30,.81,.26,.17,.16,.13.Now we see that only the first two eigenvectors stand out and the important variables x_1 and x_2 can be identified. Table 7 and Table 8 summarized results about $\hat{b}_1$, $\hat{b}_2$, $|c\hat{o}s(\theta_1)|$ and $|c\hat{o}s(\theta_2)|$ from 100 simulation runs. We can see that the means of $\hat{b}_1$ and $\hat{b}_2$ are very close to the true lifetime directions.

Table 7. Mean and standard deviation for the eigenvalues and the coefficients in the leading directions for Example 3 by SAVE with weight adjustment in 100 runs

eigenvalues	1.44(.21)	.65(.14)	.29(.05)	.21(.03)	.16(.03)	.12(.02)
$\hat{b}_1$	.977(.02)	-.006(.13)	-.010(.09)	.005(.08)	.007(.08)	.006(.08)
$\hat{b}_2$	.003(.13)	.952(.08)	-.035(.15)	.011(.12)	-.002(.13)	.001(.13)

In this example, the first two eigenvectors of the modified SIR are (0.38,-0.72,-0.21,-0.20,-0.40,0.30)′ and (-0.05,0.01,0.23,-0.75,0.48,0.38)′. We can see that the modified SIR failed to find the true lifetime directions.

Table 8. Performance of modified SAVE and modified SIR of Example 3 with 100 runs

	$\lvert c\hat{o}s(\theta_1)\rvert$	P-value	$\lvert c\hat{o}s(\theta_2)\rvert$	P-value
Modified SAVE	.977(.02)	.000(.00)	.952(.08)	.001(.01)
Modified SIR	.341(.26)	.997(.01)	.320(.21)	.999(.01)

Table 9. $|\hat{cos}(\theta)|$ of modified SAVE and modified SIR as the number of dimensions increases under Example 3 with 100 runs

	p	6	10	15	20
Modified SAVE	$\|\hat{cos}(\theta_1)\|$	.9771(.0235)	.9601(.0709)	.9495(.0189)	.9333(.0254)
	$\|\hat{cos}(\theta_2)\|$	.9504(.0438)	.8673(.1852)	.7412(.2352)	.6153(.2435)
Modified SIR	$\|\hat{cos}(\theta_1)\|$	.3105(.2533)	.2521(.2076)	.2314(.1886)	.1777(.1439)
	$\|\hat{cos}(\theta_2)\|$	.3380(.2346)	.2546(.1895)	.2097(.1561)	.1737(.1264)

6. Conclusion

Under the conditionally independent censoring condition, we have demonstrated how to extend the dimension reduction method of SAVE to censored data and made comparisons with SIR. In estimating the kernel matrix of SAVE, we introduce a weight function to modify.

The estimation of the weight function requires nonparametric smoothing. There are two stage options. Using the double slicing procedure to find the joint e.d.r. directions first. Applying the kernel smoothing technique in estimating the weight function. We can construct the sample kernel matrix of SAVE by the two-stage procedure then find the true lifetime directions. However, if the censoring time also has a dimension reduction structure, we can use the same procedure to study the censoring pattern.

When making inference about the dimensions, we suggest the simple permutation tests by Cook and Weisberg to do it. Using the permutation *P*-values to find the significant lifetime directions.

If the data have the left truncated and right censored structure, how to adjust such bias caused by truncation and censoring is very difficult to be made. This prospect merits further study.

Appendix

Derivation of (3.4). Under the conditionally independent (1.3), it implies that $S_Y(y|z) = S^0(y|x)S_C(y|z)$, where $S_C(y|z) = P(C > y|z)$. From the well-known relationship between survival function and cumulated hazards, $\Lambda(t', t|\mathbf{x})$ can be written as

$$\begin{aligned}
\Lambda(t', t|z) &= E\left\{ \frac{\mathbf{1}(t' < Y < t, \delta = 1)}{S_Y(Y|z)} \middle| z \right\} = E\left\{ \frac{\mathbf{1}(t' < Y < t, \delta = 1)}{S^0(Y^0|z)S_C(Y^0|z)} \middle| z \right\} \\
&= E\left\{ \frac{\mathbf{1}(t' < Y^0 < t)\mathbf{1}(Y^0 < C)}{S^0(Y^0|z)S_C(Y^0|z)} \middle| z \right\} \\
&= E\left\{ \frac{\mathbf{1}(t' < Y^0 < t)}{S^0(Y^0|z)S_C(Y^0|z)} E\{\mathbf{1}(Y^0 < C)|z, Y^0\} \middle| z \right\} \\
&= E\left\{ \frac{\mathbf{1}(t' < Y^0 < t)}{S^0(Y^0|z)} \middle| z \right\} \\
&= \int_{t' < Y^0 < t} \frac{1}{P(Y^0 \geq y^0|z)} f^0(y^0|z) dy^0 = \int_{t' < Y^0 < t} \frac{-1}{S^0(y^0|z)} dS^0(y^0|z) \\
&= -lnS^0(y^0|z)\Big|_{t'}^{t} = ln\frac{S^0(t'|z)}{S^0(t|z)}. \\
\Rightarrow w(t', t, z) &= \frac{S^0(t|z)}{S^0(t'|z)} = exp\{-\Lambda(t', t|z)\}.
\end{aligned}$$

References

Carroll, R. J. and Li, K. C. (1992). Measurement error regression with unknown link: dimension reduction and data visualization. *J. Amer. Statist. Assoc.* **87** 1040-1050.

Carroll, R. J. and Li, K. C. (1995). Binary regressors in dimension reduction models: a new look at treatment comparisons. *Statist. Sinica* **5** 667-688.

Chen, C. H. and Li, K. C. (1998). Can SIR be as popular as multiple linear regression? *Statist. Sinica* **8** 289-316.

Cook, R. D. (1998). *Regression graphics*. Wiley, New York.

Cook, R. D. and Lee, H. (1999). Dimension reduction in binary response regression. *J. Amer. Statist. Assoc.* **94** 1187-1120.

Cook, R. D. and Weisberg, S. (1991). Comment on ”Sliced inverse regression for dimension reduction,” by K. C. Li. *J. Amer. Statist. Assoc.* **86** 328-332.

Cook, R. D. and Weisberg, S. (1994). *An Introduction to Regression Graphics*. Wiley, New York.

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.* **86** 316-342.

Li, K. C. (2000). High dimensional data analysis via the sir/phd approach. Unpublished manuscript dated April 2000.

Li, K. C., Wang, J. L. and Chen, C. H. (1999). Dimension reduction for censored regression data. *Ann. Statist.* **27** 1-23.

Lue, H. H. (2004). Principal Hessian Directions for regression with measurement error. *Biometrika* **91** 409-423.