

# ODTWARZANIE ZNAKU PRZY POMOCY LASSO, THRESHOLDED LASSO I THRESHOLDED BASIS PURSUIT

ANNA KOPSZAK

## 1. WSTĘP

W praktycznej statystyce nierzadko zdarza się, że liczba parametrów  $p$  jest większa od liczby obserwacji  $n$ . Typowych przykładów takiego zjawiska dostarczają badania genetyczne, w których często zbiór danych opisuje poziomy ekspresji tysięcy genów dla zaledwie kilku obserwacji, i które najczęściej służą do znalezienia genów odpowiedzialnych za konkretne medyczne cechy. Kolejnym przykładem jest problem odtworzenia punktów nieciągłości dla kawałkami ciągłej funkcji  $f$  w sytuacji, gdy obserwacja zawiera zakłócenia. Można ten problem opisać wzorem  $y = X\beta_0 + \epsilon$ , w którym niezerowy współczynnik przy  $\beta_0$  odpowiada punktowi nieciągłości funkcji  $f$ . W obu tych przykładach dążymy do tego, by poprawnie odtworzyć nośnik  $\beta_0$ .

Badając wielowymiarowy model liniowy

$$y = X\beta_0 + \epsilon,$$

gdzie  $X$  jest macierzą rozmiaru  $n \times p$  oraz  $\epsilon \sim N(0, \sigma^2 I_n)$  jest elementem zakłócającym, powinniśmy poczynić pewne dodatkowe założenia, aby estymacja parametrów była możliwa. O macierzy  $X$  zakładamy, że jest pełnego rzędu  $n$ . Dodatkowo zakładamy że wektor współczynników  $\beta_0 \in \mathbb{R}^p$  jest rzadki (innymi słowy zakładamy, że większość jego elementów jest równa 0). W modelu regresji, gdzie każda kolumna  $X_j \in \mathbb{R}^n$  macierzy  $X$  odpowiada potencjalnemu predyktorowi, zaś  $y$  jest zmienną zależną, rzadkość oznacza, że tylko niewielka część spośród predyktorów jest istotna dla predykcji  $y$ .

Rozważmy wielowymiarowy model liniowy

$$Y = X\beta + \epsilon,$$

gdzie  $X = (X_1 | \dots | X_p)$  jest macierzą planu rozmiaru  $n \times p$ ,  $n \leq p$ , a  $\epsilon$  jest nieznanym wektorem współczynników regresji. Wektor znaku  $\beta$  to  $S(\beta) = (S(\beta_1), \dots, S(\beta_p)) \in \{-1, 0, 1\}^p$ , gdzie dla  $x \in \mathbb{R}$  mamy  $S(x) = \mathbf{1}_{x>0} - \mathbf{1}_{x<0}$ . Naszym głównym celem jest odtworzenie  $S(\beta)$ . Jest to problem nieco ogólniejszy niż odtworzenie wyłącznie nośnika  $\text{supp}(\beta)$ . Znak  $\beta$  możemy estymować przy pomocy znaku dobrze znanego estymatora LASSO.

$$\hat{\beta}^L := \operatorname{argmin}_{b \in \mathbb{R}} \sum_{i=1}^p |b_i|$$

- (1) Znak  $\beta^{(r)}$  jest niezmienny, to znaczy istnieje wektor znaku  $s^0 \in \{-1, 0, 1\}^p$ , taki że dla każdego  $r \in \mathbb{N}$  zachodzi  $S(\beta^{(r)}) = s_0$ .
- (2) Zachodzi następująca równość  $\lim_{r \rightarrow +\infty} \min\{|\beta_i^{(r)}| : i \in \text{supp}(s_0)\} = +\infty$ .

(3) Istnieje  $q > 0$ , takie że  $\forall r \in \mathbb{N}$ ,  $\frac{\min\{|\beta_i^{(r)}| : i \in \text{supp}(s_0)\}}{\|\beta^{(r)}\|_\infty} \geq q$ .

## 2. ODTWARZANIE ZNAKU PRZY POMOCY ESTYMATORA LASSO

**Definicja 1.** Dla ustalonego  $b \in \mathbb{R}^p$  określmy  $I := \{i \in 1, \dots, p \mid b_i \neq 0\}$  oraz niech  $X_I$  i  $X_{\bar{I}}$  będą macierzami o kolumnach  $(X_i)_{i \in I}$  i  $(X_i)_{i \notin I}$ , odpowiednio. Wektor  $b$  spełnia *warunek niereprezentowalności (IC)*, jeżeli  $\ker(X_I) = 0$  oraz  $\|X_{\bar{I}}' X_I (X_I' X_I)^{-1} S(b_I)\|_\infty \leq 1$ .

**Uwaga.** IC jest warunkiem koniecznym, aby prawdopodobieństwo odtworzenia  $S(\beta)$  było duże.

**Definicja 2.** Wektor  $b \in \mathbb{R}^p$  jest identyfikowalny względem normy  $L_1$ , jeżeli zachodzi implikacja

$$(X\gamma = Xb \wedge \gamma \neq b) \Rightarrow (\|\gamma\|_1 > \|b\|_1).$$

Poniższe twierdzenie pozwala określić górną granicę prawdopodobieństwa odtworzenia  $S(\beta)$  przy pomocy estymatora LASSO przy założeniu, że zachodzi IC.

**Twierdzenie 1.** Niech  $I := \text{supp}(\beta)$  oraz  $X_I$  i  $X_{\bar{I}}$  będą macierzami o kolumnach  $(X_i)_{i \in I}$  i  $(X_i)_{i \notin I}$ , odpowiednio. Załóżmy, że  $\ker(X_I) = 0$  i określmy

$$\zeta_{X,\lambda,S(\beta)} := X_{\bar{I}}' X_I (X_I' X_I)^{-1} S(b_I) + \frac{1}{\lambda} X_{\bar{I}}' (Id - X_I (X_I' X_I)^{-1} X_I') \epsilon.$$

Prawdziwe jest następujące ograniczenie górne prawdopodobieństwa odtworzenia znaku

$$\mathbb{P}(S(\hat{\beta}_L(\lambda)) = S(\beta)) \leq \mathbb{P}(\|\zeta_{X,\lambda,S(\beta)}\|_\infty \leq 1).$$

Ponadto, niech  $(\beta^r)$  oznacza wektor z  $\mathbb{R}^p$  spełniający założenie (1). Jeżeli  $s^0$  jest identyfikowalny (ang. identifiable) względem normy  $L_1$ , to otrzymujemy następujące asymptotyczne oszacowania

$$\limsup_{r \rightarrow +\infty} \mathbb{P}(S(\hat{\beta}^L(\lambda, r)) = s^0) \leq \mathbb{P}(\|\zeta_{X,\lambda,s^0}\|_\infty \leq 1)$$

oraz

$$\limsup_{r \rightarrow +\infty} \mathbb{P}(S(\hat{\beta}^L(\lambda, r)) = s^0) \geq \mathbb{P}(\|\zeta_{X,\lambda,s^0}\|_\infty < 1).$$

Niech  $\mathbb{P}(\|\zeta_{X,\lambda,s^0}\|_\infty < 1) = \gamma$  i  $\mathbb{P}(\|\zeta_{X,\lambda,s^0}\|_\infty \leq 1) = \bar{\gamma}$ . Znak niezerowych elementów  $\beta^{(r)}$  jest dobrze odtworzony z prawdopodobieństwem zbiegającym do 1 i błąd pierwszego rodzaju (FWER) jest kontrolowany na poziomie  $1 - \gamma$ . Mianowicie zachodzą oszacowania

$$\limsup_{r \rightarrow +\infty} \mathbb{P}(\forall i \in I \ S(\hat{\beta}_i^L(\lambda, r)) = s_i^0) = 1,$$

$$\limsup_{r \rightarrow +\infty} \mathbb{P}(\exists i \notin I \ S(\hat{\beta}_i^L(\lambda, r)) \neq 0) \leq 1 - \gamma,$$

$$\limsup_{r \rightarrow +\infty} \mathbb{P}(\exists i \notin I \ S(\hat{\beta}_i^L(\lambda, r)) \neq 0) \geq 1 - \gamma.$$

Jeśli  $\beta$  nie spełnia IC, to estymator znaku LASSO (ang. LASSO sign estimator)  $S(\hat{\beta}^L(\lambda))$  nie może odtworzyć  $S(\beta)$  z dużym prawdopodobieństwem. Poniższe twierdzenie pokazuje, że LASSO po dodaniu odpowiedniego progu (ang. threshold) może odtworzyć  $S(\beta)$ , jeżeli tylko niezerowe elementy  $\beta$  są odpowiednio duże oraz zachodzi IC.

**Twierdzenie 2.** *Niech  $X$  będzie macierzą w pozycji ogólnej (ang. general position), taką że  $\text{rank}(X) = n$  i niech  $\hat{\beta}$  będzie estymatorem LASSO albo BPDN z ustaloną wartością  $\lambda > 0$  lub z ustalonym parametrem regularyzacji  $R \geq 0$ . Jeżeli  $S(\beta)$  jest identyfikowalny względem normy  $L_1$ , to estymator znaku otrzymany dzięki thresholded LASSO albo thresholded BPDN nie może odtworzyć  $S(\beta)$ . Dla każdego  $\epsilon \in \mathbb{R}^n$  albo niektóre ujemne współrzędne  $\beta$  nie są poprawnie estymowane jako ujemne przez  $\hat{\beta}(\epsilon)$ , albo niektóre dodatnie współrzędne  $\beta$  nie są poprawnie estymowane jako dodatnie przez  $\hat{\beta}(\epsilon)$ .*

*Ponadto, zachodzi następujący asymptotyczny warunek. Niech  $\beta^{(r)}$  będzie wektorem w  $\mathbb{R}^p$  spełniającym założenie (1). Jeżeli  $s^0$  jest identyfikowalny względem normy  $L_1$  dla każdego ustalonego  $\epsilon \in \mathbb{R}^n$  i wystarczająco dużego  $r > r_0(\epsilon)$ , to estymator  $\hat{\beta}(\epsilon, r)$  oddziela ujemne współrzędne  $\beta^{(r)}$  od dodatnich współrzędnych  $\beta^{(r)}$ .*

Twierdzenie 2 potwierdza, że nie możemy odtworzyć  $S(\beta)$  z estymatorem znaku otrzymanym dzięki LASSO albo BPDN, kiedy wektor  $\beta$  nie jest identyfikowalny względem  $L_1$ . Jeśli wektor  $\beta$  jest identyfikowalny względem normy  $L_1$ , to Twierdzenie 2 sugeruje, że  $S(\beta)$  może być odtworzone poprzez uzyskanie estymatorów znaku przy pomocy thresholded LASSO albo thresholded BPDN.

### 3. PORÓWNANIE ESTYMATORÓW ZNAKU

Twierdzenie 2 mówi, że estymatory znaku dostarczone przez LASSO albo thresholded BPDN pozwalają odtworzyć  $S(\beta)$  pod warunkiem, że jest spełniony IC. Innym sposobem, aby odtworzyć  $S(\beta)$ , jest estymator znaku otrzymamy przy pomocy adaptive LASSO [3]. Jeżeli wagi dla adaptive LASSO opierają się na estymatorze dostatecznym  $\beta$ , to możemy otrzymać spójny estymator przy założeniach słabszych niż IC.

**Układ 1.** Niech  $X$  będzie macierzą  $n \times p$ , taką że  $n = 100$ ,  $p = 300$ , której elementy są generowane przez niezależne obserwacje ze standardowego rozkładu normalnego  $N(0, 1)$ . Rozkład wektorów znaku jest jednostajny na  $\{u \in \{-1, 0, 1\}^p : \#(\text{supp}(u)) = k\}$ .

**Układ 2.** Niech  $X$  będzie macierzą  $n \times p$ , taką że  $n = 100$ ,  $p = 300$ , której elementy są generowane przez niezależne obserwacje z wielowymiarowego rozkładu normalnego  $N(0, \Gamma)$ , gdzie  $\Gamma_{i,i} = 1$  dla  $i \in \{1, \dots, p\}$  oraz  $\Gamma_{i,j} = 0.9$  dla  $i \neq j$ . Rozkład wektorów znaku jest jednostajny na  $\{u \in \{-1, 0, 1\}^p : \#(\text{supp}(u)) = k\}$ .

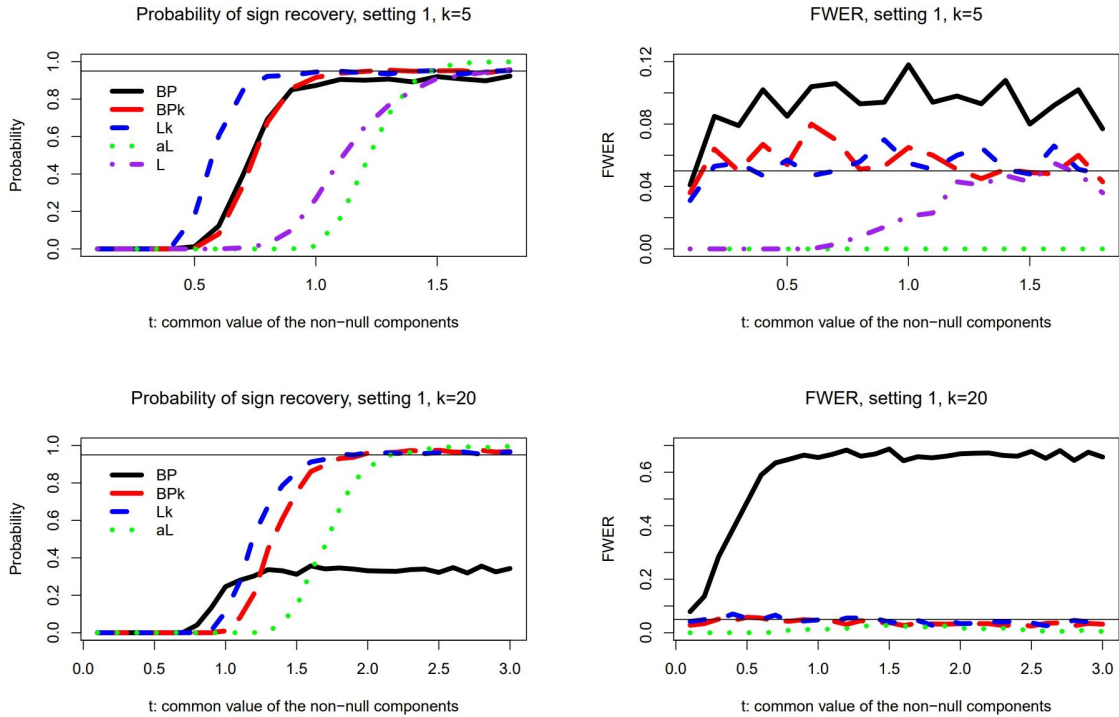
**Układ 3.** Układ 3 różni się od Układu 2 tylko tym, że rozkład wektorów znaku jest jednostajny na  $\{u \in \{0, 1\}^p : \#(\text{supp}(u)) = k\}$ .

Niedawno opracowana metoda knockoff [2] pozwala kontrolować False Discovery Rate. Osiągamy to dzięki dodaniu do macierzy planu dodatkowych zmiennych. W tym zestawieniu korzystamy z knockoffów przedstawionych w [2], żeby przybliżyć wartość progu, który pozwala kontrolować FWER na ustalonym poziomie. To podejście sprawdza się, kiedy  $X$  jest macierzą Gaussowską z rozkładem niezmiennym ze względu na permutacje kolumn.

Poniżej będziemy stosować następującą notację

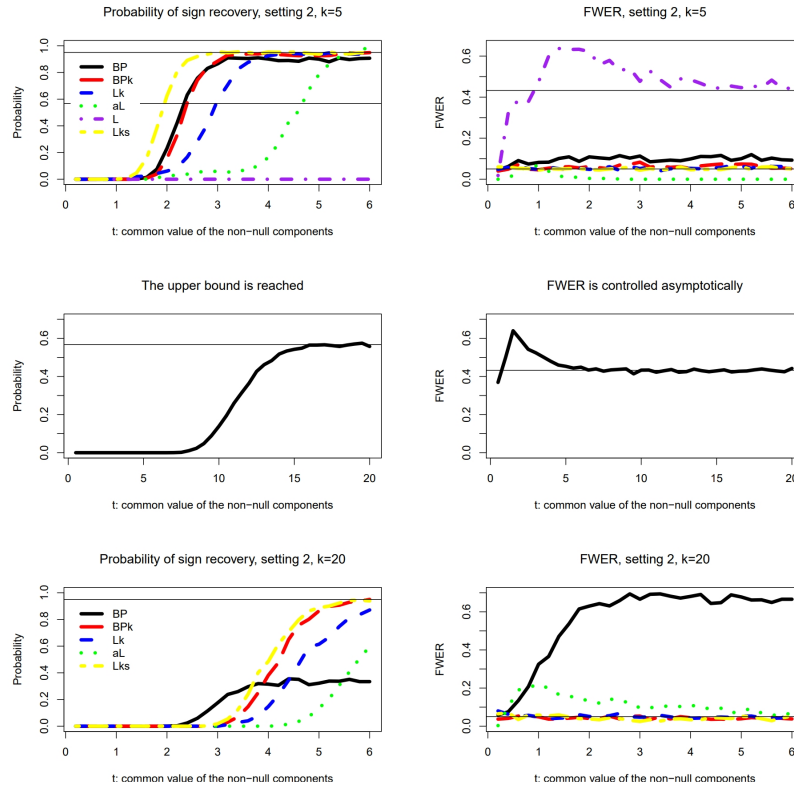
- **L** - estymator LASSO,
- **aL** - estymator adaptive LASSO,
- **BP** - estymator thresholded Basis Pursuit,
- **BPk** - estymator thresholded Basis Pursuit; próg ustalamy dzięki metodzie knockoff,
- **Lk** - estymator thresholded LASSO;  $\lambda = \lambda_{AMP}$  ustalamy dzięki metodzie knockoff,
- **Lks** - estymator thresholded LASSO;  $\lambda = 0.5\lambda_{AMP}$  ustalamy dzięki metodzie knockoff.

Poniższe wykresy opisują FWER i prawdopodobieństwo odtworzenia  $S(\beta)$  dla każdego z estymatorów znaku, kiedy  $X$  jest macierzą planu jak w Układzie 1. Wykresy po lewej stronie pokazują prawdopodobieństwo odtworzenia  $S(\beta)$  w postaci funkcji zmiennej  $t$ , gdzie  $t$  jest wartością bezwzględną niezerowych elementów  $\beta$ . Wykresy po prawej stronie ilustrują FWER jako funkcję  $t$ . Estymatorem, który odtwarza  $S(\beta)$  z największym prawdopodobieństwem, jest thresholded LASSO. Wszystkie estymatory znaku odtwarzają prawidłowo  $S(\beta)$  z prawdopodobieństwem bliskim 0.95, kiedy  $t$  jest duże.



Poniższe wykresy opisują FWER i prawdopodobieństwo odtworzenia  $S(\beta)$  dla każdego z estymatorów znaku, kiedy  $X$  jest macierzą planu jak w Układzie 2. Wykresy po lewej stronie pokazują prawdopodobieństwo odtworzenia  $S(\beta)$  w postaci funkcji zmiennej  $t$ , gdzie  $t$  jest wartością bezwzględną niezerowych elementów  $\beta$ . Wykresy po prawej stronie ilustrują FWER jako funkcję  $t$ . Poziome linie  $y = 0.55$  i  $y = 0.45$  reprezentują odpowiednio średnią wartość ograniczenia górnego prawdopodobieństwa odtworzenia znaku i FWER odpowiadające LASSO.

Możemy zauważyć, że ograniczenie górne jest osiągnięte i FWER jest kontrolowane, kiedy  $t$  jest bardzo duże, tak jak pokazują wykresy na środku. Estymatory znaku (poza estymatorem LASSO) odtwarzają  $S(\beta)$  z prawdopodobieństwem bliskim 0.95, kiedy  $t$  jest duże.



## LITERATURA

- [1] Patrick J. C. Tardivel i Małgorzata Bogdan, *On the sign recovery by LASSO, thresholded LASSO and thresholded Basis Pursuit Denoising*.
- [2] R. F. Barber i E. J. Candès, *Controlling the false discovery rate via knockoffs*.
- [3] Hui Zou, *The adaptive lasso and its oracle properties*.