Statistics and Its Interface Volume 17 (2024) 27–37

Copy number variation detection based on constraint least squares

Xiaopu Wang, Xueqin Wang, Aijun Zhang, and Canhong Wen*

Copy number variations (CNVs) are a form of structural variation of a DNA sequence, including amplification and deletion of a particular DNA segment on chromosomes. Due to the huge amount of data in every DNA sequence, there is a great need for a computationally fast algorithm that accurately identifies CNVs. In this paper, we formulate the detection of CNVs as a constraint least squares problem and show that circular binary segmentation is a greedy approach to solving this problem. To solve this problem with high accuracy and efficiency, we first derived a necessary optimality condition for its solution based on the alternating minimization technique and then developed a computationally efficient algorithm named AMIAS. The performance of our method was tested on both simulated data and two real-world applications using genomic data from diagnosed primal glioblastoma and the HapMap project. Our proposed method has competitive performance in identifying CNVs with high-throughput genotypic data.

Keywords and phrases: Alternating minimization induced active set, Change point detection, Circular binary segmentation, HapMap.

1. INTRODUCTION

Copy number variations (CNVs) are a form of structural variation of the DNA sequence, including amplification and deletion of a particular DNA segment on chromosomes (Freeman et al., 2006). CNVs have been shown to be associated with a wide collection of diseases, including autism (Sebat et al., 2007), Alzheimer's (Brouwers et al., 2012), schizophrenia (Ingason et al., 2011; Vacic et al., 2011), and cancer (Shlien and Malkin, 2009).

Array comparative genomic hybridization (aCGH) has been widely used in genome-wide identification of CNVs. The aCGH technology compares the copy number of a differentially labeled case sample with a normal reference and returns $\log_2$ intensity ratio measurements. A value greater than zero indicates a possible gain in the DNA sequence, while a value less than zero suggests possible losses.

In recent years, various algorithms have been developed to identify CNVs for aCGH data. A general approach is based on the idea of binary segmentation (Vostrikova, 1981), which looks at a single point at each step and can be viewed as a greedy approach with low computational complexity. One of its most popular variants is circular binary segmentation (CBS, Olshen et al., 2004; Venkatraman and Olshen, 2007), which searches for two change points at a time and works well for multiple change point detection. CBS divides the genome into regions of equal copy number using maximum likelihood ratio statistics or t statistics. Another popular variant is the wild binary segmentation (WBS, Fryzlewicz, 2014, which randomly draws a collection of sub samples and chooses the one with the maximum statistic. Thus, WBS can detect change points even for very short spacings between change points or very small jump magnitudes (Fryzlewicz, 2014). Contrary to forward detection methods such as CBS, the backward detection (BWD, Jun Shin et al., 2020) works from the opposite extreme in that every single position is assumed to be a change point and then two neighbor ring groups are merged into one repeatedly. BWD works well in identifying short signals in lengthy noise sequences.

Another general approach formalizes the detection of CNVs as a penalized least squares problem, a technique first considered by Yao (1988). Yao (1988) proposed the Schwartz information criterion (SIC) to estimate the number of change points, but this approach brings a heavy computational burden. To alleviate the computational cost, Tibshirani et al. (2005) and Tibshirani and Wang (2008) replaced the penalty term with the fused lasso penalty and transformed the problem into convex optimization. Unlike the binary segmentation approach, the fused lasso regression method encourages both sparsity and smoothness in the copy number estimate simultaneously. Other approaches include PennCNV (Wang et al., 2007) (based on hidden Markov models), and the screening and ranking algorithm and its variant (Niu and Zhang, 2012; Xiao et al., 2014).

In this paper, we formulate the detection of CNVs as a constraint least squares problem and show that CBS is a greedy approach to solving this problem. Then we derive a necessary optimality condition for its solution based on the alternating minimization technique. A computationally fast algorithm named Alternating Minimization Induced Active Set (AMIAS) is provided to obtain an accurate solution for the constraint least squares problem. Our proposed algorithm is compared with previous methods using extensive

*Corresponding author.

simulated and real-world data, showing that it has competitive performance.

The rest of the article proceeds as follows. In Section 2, we describe a constraint least squares framework and develop the AMIAS algorithm. Section 3 is devoted to simulated numerical studies, and Section 4 analyzes two empirical datasets. We conclude the article in Section 5. Proofs of the main theorems and technical lemmas are provided in the Appendix.

2. MODEL AND METHODOLOGY

2.1 Constraint least squares framework

Suppose we have an aCGH data set from n probes. Let $Y_i, i = 1, \ldots, n$ be the observed $\log_2$ intensity ratio measurement at probe i. Then Y_i is assumed to be the sum of the true hidden signal β_i and random error ϵ_i,

$$Y_i = \beta_i + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where ϵ_i is assumed to have mean 0 and variance σ^2. Here the parameter $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ quantifies the copy number levels with $\beta_i = 0$ corresponding to the standard copy number 2. Since CNVs are rare in the whole genome, the true parameter β^* is expected to be piecewise constant with the majority of values equal to 0 and a few segments with positive values (duplication) and negative values (deletion). To recover $\boldsymbol{\beta}^*$, one could solve the following constraint optimization problem

$$\begin{aligned} &\min_{\beta} \frac{1}{2}\sum_{i=1}^{n}(Y_i - \beta_i)^2, \\ &\text{s.t.} \sum_{i=1}^{n} |\beta_i|_0 \le s_1, \ \sum_{i=2}^{n} |\beta_i - \beta_{i-1}|_0 \le s_2, \end{aligned} \tag{2}$$

where $|x|_0 = 1$ if $x \neq 0$, and $|x|_0 = 0$ if $x = 0$. Here s_1 controls the overall sparsity of CNV regions (the number of nonzero values in β), and s_2 controls the number of CNV alterations (the number of change points between the adjacent probes).

We can rewrite problem (2) in matrix and vector form as follows:

$$\min_{\beta} \frac{1}{2}\|Y - \beta\|_2^2, \quad \text{s.t.} \|\beta\|_0 \le s_1, \ \|D\beta\|_0 \le s_2, \tag{3}$$

where $\|x\|_0 = \sum_{i=1}^{n} |x_i|_0$ is the L_0 norm, and D is an $(n-1) \times n$ matrix with zeros entries everywhere except for 1 in the diagonal and -1 in the super-diagonal.

Due to the use of the non convex and noncontinuous L_0 constraint, it is computationally expensive to directly solve problem (3) by exhaustively searching over all possible choices of selections. Optimization-based methods that are computationally friendlier have been proposed as surrogates for problem (3), among which the convex relaxation technique is a popular choice. For example, with the norm $\|x\|_0$ replaced by the convex norm $\|x\|_1 = \sum_{i=1}^{n} |x_i|$, the corresponding minimizer is named the fused lasso method in Tibshirani and Wang (2008). However, like other L_1-norm based methods, the fused lasso tends to overshrink large coefficients, which leads to biased estimates and an over-selection problem (Behrendt and Schweikert, 2021; Zhao and Yu, 2006). Non convex penalties such as the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li, 2001) and the minimax concave penalty (MCP) (Zhang, 2010) can be used to remedy the bias issue by rewriting problem (3) as a regularization formulation (Xiu et al., 2019). Rather than using the relaxed penalties, we propose to directly solve the problem (3). Unlike the L_1 penalty or other non convex penalties, the L_0 penalty enables us to directly control the number of change-points, which is more explicit than using the tuning parameters as in the works of Tibshirani and Wang (2008) and Xiu et al. (2019).

For the linear regression model without the constraint on $D\beta$, Huang et al. (2018) proposed an L_0 construction approach to a least squares problem with L_0 penalty on the coefficient parameter and developed a support detection and root finding algorithm to solve the optimization problem. Wen et al. (2017) studied the constrained maximum likelihood problem and proposed a primal-dual active set algorithm for solving it. Our formulation (4) can be viewed as an extension of the algorithm proposed by Huang et al. (2018) and Wen et al. (2017) to the fused detection problem with an additional penalty on $D\beta$. However, it is not straight forward to generalize their proposed algorithm to solve problem (3) since the term $D\beta$ entangles all elements in β. To remedy the problem, we introduce an auxiliary variable $v = D\beta$, and reformulate the problem (4) to its augmented Lagrangian form:

$$\begin{aligned} L(\beta, v, u) = &\frac{1}{2}\|Y - \beta\|_2^2 + \lambda_1\|\beta\|_0 + \lambda_2\|v\|_0 \\ &+ \frac{\rho}{2}\|D\beta - v\|_2^2 + u^T(D\beta - v), \end{aligned} \tag{4}$$

where $u \in \mathcal{R}^{n-1}$ is a dual variable corresponding to the linear constraint $v = D\beta$, and parameters λ_1, λ_2, and ρ all have positive values. As in the constrained version, parameters λ_1 and λ_2 encourage sparsity in β and shrink the differences between neighboring coefficients in β toward zero, respectively. The introduction of v separates the parameters needing to be estimated, making the problem computational feasible. To be specific, let $(\beta^\diamond, v^\diamond, u^\diamond)$ be a coordinate-wise solution to (4). Next, we present the necessary optimality conditions for $(\beta^\diamond, v^\diamond, u^\diamond)$ with the help of an alternating minimization technique. The proof appears in the Appendix.

Theorem 1. *Let $\boldsymbol{N}^\rho = (N_1^\rho, \ldots, N_n^\rho)$ with $N_j^\rho = 1 + 2\rho$ for $j = 2, \ldots, n-1$, and $N_j^\rho = 1 + \rho$ for $j = 1, n$. Then, for the*

solution $(\beta^\diamond, v^\diamond, u^\diamond)$*, we have:*

$$\begin{aligned}\beta_j^\diamond &= \begin{cases}\beta_j^\diamond + d_j^\diamond, & \text{if } |\beta_j^\diamond + d_j^\diamond| > \sqrt{2\lambda_1/N_j^\rho}\\ 0, & \text{otherwise,}\end{cases}\\ v_j^\diamond &= \begin{cases}v_j^\diamond + u_j^\diamond/\rho, & \text{if } |v_j^\diamond + u_j^\diamond/\rho| > \sqrt{2\lambda_2/\rho},\\ 0, & \text{otherwise,}\end{cases}\end{aligned} \tag{5}$$

where $d_j^\diamond$ *is the* j*-th element of* $d^\diamond = (y - D^T u^\diamond - \beta^\diamond)/\boldsymbol{N}^\rho$.

Note that $d^\diamond$ can be viewed as a dual variable of $\beta^\diamond$. From (5), we conclude that

$$\begin{cases}\beta_j^\diamond = 0, & \text{if } |\beta_j^\diamond + d_j^\diamond| \le \sqrt{2\lambda_1/N_j^\rho},\\ d_j^\diamond = 0, & \text{if } |\beta_j^\diamond + d_j^\diamond| > \sqrt{2\lambda_1/N_j^\rho}.\end{cases} \tag{6}$$

Define an active set $\mathcal{A} = \{j : |\beta_j^\diamond + d_j^\diamond| > \sqrt{2\lambda_1/N_j^\rho}\}$ and an inactive set $\mathcal{I} = \{j : |\beta_j^\diamond + d_j^\diamond| \le \sqrt{2\lambda_1/N_j^\rho}\}$. Then, we have $d_{\mathcal{A}}^\diamond = 0$ and $\beta_{\mathcal{I}}^\diamond = 0$. Thus on the active set $\mathcal{A}$, only the dual variable $d^\diamond$ is 0, and the primal variable $\beta^\diamond$ is different from 0. In contrast, on the inactive set $\mathcal{I}$, the primal variable $\beta^\diamond$ is 0, and the dual variable $d^\diamond$ may or may not be 0. From the above discussion, set $\mathcal{A}$ is actually the active set for sparsity in $\beta^\diamond$; that is, $\mathcal{A} = \{j : \beta_j^\diamond \neq 0\}$.

Similarly, we have the following equation for v:

$$\begin{cases}v_j^\diamond = 0, & \text{if } |v_j^\diamond + u_j^\diamond/\rho| \le \sqrt{2\lambda_2/\rho},\\ u_j^\diamond = 0, & \text{if } |v_j^\diamond + u_j^\diamond/\rho| > \sqrt{2\lambda_2/\rho}.\end{cases} \tag{7}$$

Define an active set $\mathcal{B} = \{j : |v_j^\diamond + u_j^\diamond/\rho| > \sqrt{2\lambda_2/\rho}\}$ and an inactive set $\mathcal{J} = \{j : |v_j^\diamond + u_j^\diamond/\rho| \le \sqrt{2\lambda_2/\rho}\}$. Then, we have $u_{\mathcal{B}}^\diamond = 0$ and $v_{\mathcal{J}}^\diamond = 0$. Thus on the active set $\mathcal{B}$, only the dual variable $u^\diamond$ is 0, and the primal variable $v^\diamond$ is not 0. On the inactive set $\mathcal{J}$, the primal variable $v^\diamond$ is 0, and the dual variable $u^\diamond$ component may or may not be 0. These facts suggest that set $\mathcal{B}$ is the active set for sparsity in $v^\diamond = D\beta^\diamond$; that is, $\mathcal{B} = \{j : v_j^\diamond \neq 0\}$. Each element in $\mathcal{B}$ corresponds to a change point at which the copy number changes.

2.2 Relationship with CBS

The basic ingredient of the CBS is the likelihood ratio test statistic defined by

$$Z_{ij} = \sqrt{\frac{(j-i)(n-j+i)}{n}}\left(\frac{S_j - S_i}{j-i} - \frac{S_n - S_j + S_i}{n-j+i}\right), \tag{8}$$

where $S_i = Y_1 + \cdots + Y_n$, $i = 1, \ldots, n$, is the partial sum. The CBS algorithm computes Z_{ij} and then takes $Z_C = \max_{1\le i<j\le n}|Z_{ij}|$ to be the first change-point candidate, whose significance is to be judged against a certain criterion. Once it is considered significant, the interval $[1, n]$ is split, and the same procedure is applied recursively to identify all the change points. The following lemma provides further insight into the relationship between Z_{ij} and $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$.

Lemma 1. *Define* $\mathcal{F}_{i,j}^{1,n}$ *to be the set of circular piecewise-constant functions supported on* $[1, n]$ *with change points at* i *and* j*. We have*

$$arg\max_{i,j}|Z_{ij}| = arg\min_{i,j}\min_{f_{i,j}^{1,n}\in\mathcal{F}_{i,j}^{1,n}}\sum_{t=1}^{n}(Y_t - f_{i,j}^{1,n})^2. \tag{9}$$

Lemma 1 shows that maximizing $|Z_{ij}|$ is equivalent to the least squares fit of a piecewise-constant function with one or two change-points to $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$. Furthermore, the optimization problem in (9) can be reformulated as the following penalized least squares problem with the L_0 penalty,

$$\min_\beta \sum_{t=1}^{n}(Y_t - \beta_t)^2 \quad \text{subject to} \quad \sum_{t=1}^{n}|\beta_i - \beta_{i+1}|_0 \le 2. \tag{10}$$

Thus CBS iteratively solves the L_0 constraint least squares problem with the count of change points no more than 2. The process of the CBS algorithm declaring a change is to check whether $|Z_{i,j}|$ is large enough, which is equivalent to checking whether the difference of the estimated β in two separated segments is large enough. In this regard, we can see that CBS is a special case of our method, which we name it the fused L_0 approach.

When the change points are estimated to be i and j, then the CBS is applied cursively on the segments $[1, i]$, $[i+1, j]$, and $[j+1, n]$. We will see next that the fused L_0 method is optimal compared to the CBS algorithm in terms of the sum of squared errors (SSE). Note that

$$\begin{aligned}&\min_{f_6^{1,n}\in\mathcal{F}_6^{1,n}}\sum_{t=1}^{n}(Y_t - f_6^{1,n})^2\\ &\le \min_{i',j'}\min_{f_{i',j'}^{1,i}\in\mathcal{F}_{i,j}^{1,i}}\sum_{t=1}^{i}(Y_t - f_{i',j'}^{1,i})^2\\ &\quad + \min_{i',j'}\min_{f_{i',j'}^{i+1,j}\in\mathcal{F}_{i,j}^{i+1,j}}\sum_{t=i+1}^{j}(Y_t - f_{i',j'}^{i+1,j})^2\\ &\quad + \min_{i',j'}\min_{f_{i',j'}^{j+1,n}\in\mathcal{F}_{i,j}^{j+1,n}}\sum_{t=j+1}^{n}(Y_t - f_{i',j'}^{j+1,n})^2,\end{aligned} \tag{11}$$

where $\mathcal{F}_6^{1,n}$ is the set of piecewise-constant function supported on $[1, n]$ with at most six change points. The left-hand side of the above inequality is equivalent to the penalized least squares problem with L_0 penalty,

$$\min_\beta \sum_{t=1}^{n}(Y_t - \beta_t)^2 \quad \text{subject to} \quad \sum_{t=1}^{n}|\beta_i - \beta_{i+1}|_0 \le 6. \tag{12}$$

The right-hand side of (11) is the SSE with change points detected by the CBS algorithm. Thus, the SSE of the fused L_0 method is always smaller than the CBS algorithm's SSE. This relationship proves that CBS can be viewed as a top-down sequential greedy algorithm to solve the constraint least squares problem.

2.3 Algorithm

From the discussion in Section 2.1, the solution to problem (4) has complementary supports in the primal-dual variable pairs. The two active sets $\mathcal{A}$ and $\mathcal{B}$ play a crucial role in solving the fused L_0 problem; indeed, if $\mathcal{B}$ or $\mathcal{J}$ is known a priori, we may estimate $u^\diamond$ on $\mathcal{J}$ by

$$u^\diamond_{\mathcal{J}} = (D_{\mathcal{J}} D_{\mathcal{J}}^\top)^{-1} D_{\mathcal{J}} Y, \tag{13}$$

where $D_{\mathcal{J}}$ denotes the sub matrix with the row index in $\mathcal{J}$. Then if $\mathcal{A}$ is known a priori, the primal-dual variable $\beta^\diamond$ on $\mathcal{A}$ can be estimated by

$$\beta^\diamond_{\mathcal{A}} = Y_{\mathcal{A}} - D_{\mathcal{A}}^\top u^\diamond. \tag{14}$$

Thus, the key ingredient in our algorithm is the determination of sets $\mathcal{A}$ and $\mathcal{B}$. Firstly, we can see that there are three parameters in sets $\mathcal{A}$ and $\mathcal{B}$. The parameter ρ can be viewed as a learning rate, for which a larger value is better; here, we set $\rho = n^2$. The parameters λ_1 and λ_2 are two real numbers and related to the number of change points and nonzero points respectively. Take the active set $\mathcal{B}^\diamond$ as an example. Let $\Delta_j = |v_j^\diamond + \frac{1}{\rho} u_j^\diamond|$, $j = 1, \ldots, n-1$ and denote the order statistic $\Delta_1, \ldots, \Delta_{n-1}$ by decreasing order as $\Delta_{[1]}, \ldots, \Delta_{[n-1]}$. If parameter $\sqrt{2\lambda_1}$ lies in $[\Delta_{[k]}, \Delta_{[k+1]})$, the active set $\mathcal{B}^\diamond$ stays the same. Thus, we can choose $\sqrt{2\lambda_1} = \Delta_{[k]}$ and transform the problem of selecting λ_1 into determining the integer k. Similarly, we can replace λ_2 with an integer s. Note that k and s have the explicit meaning of the number of segments and number of nonzero probes, respectively. Thus we can develop an iterative procedure for updating $\mathcal{A}$ and $\mathcal{B}$ for fixed k and s and search over a two-dimensional grid of possible integers to find an optimal pair of parameters.

Suppose at the m-th iteration with the current estimate $\mathcal{A}^m$ and $\mathcal{B}^m$, we can estimate u^m and β^m by (13)–(14) and derive d^m and v^m as discussed above. Then we can update the active sets by

$$\mathcal{B}^{m+1} = \left\{ j : \Delta_j^m \geq \Delta_{[k]}^m \right\}, \quad \mathcal{J}^{m+1} = \left\{ j : \Delta_j^m < \Delta_{[k]}^m \right\},$$
$$\mathcal{A}^{m+1} = \left\{ i : \Gamma_i^m \geq \Gamma_{[s]}^m \right\}, \quad \mathcal{I}^{m+1} = \left\{ i : \Gamma_i^m < \Gamma_{[s]}^m \right\},$$

where $\Delta_j^m = |v_j^m + \frac{1}{\rho} u_j^m|$, $j = 1, \ldots, n-1$ and $\Gamma_i^m = \sqrt{N_i^\rho} |\beta_i^m + d_i^m|$, $i = 1, \ldots, n$.

We call this the Alternating Minimization Induced Active Set (AMIAS) algorithm for solving the constraint least squares problem (4). AMIAS is presented in pseudocode form below.

AMIAS algorithm

1. Specify the number of change points k, the cardinality s of the active set for sparsity, and the maximum number of iterations $m_{\max}$. Let $\rho = n^2$. Initialize $\mathcal{I}^0 = \phi$, $\mathcal{A}^0 = (\mathcal{I}^0)^c$ and $\mathcal{B}^0 = \phi$, $\mathcal{J}^0 = (\mathcal{B}^0)^c$, where ϕ is the empty set.
2. For $m = 0, 1, 2, \ldots, m_{\max}$, do

   (2.a) Determine u^m and v^m by

   $$u^m_{\mathcal{B}^m} = 0, \quad \text{and} \quad u^m_{\mathcal{J}^m} = (D_{\mathcal{J}^m} D_{\mathcal{J}^m}^\top)^{-1} D_{\mathcal{J}^m} Y,$$
   $$v^m_{\mathcal{J}^m} = \mathbf{0}, \quad \text{and} \quad v^m_{\mathcal{B}^m} = D_{\mathcal{B}^m}(Y - D^\top u^m).$$

   (2.b) Determine β^m and d^m by

   $$\beta^m_{\mathcal{I}^m} = 0, \quad \text{and} \quad \beta^m_{\mathcal{A}^m} = Y_{\mathcal{A}^m} - D^\top_{\mathcal{A}^m} u^m,$$
   $$d^m_{\mathcal{A}^m} = 0, \quad \text{and} \quad d^m_{\mathcal{I}^m} = \left(Y_{\mathcal{I}^m} - D^\top_{\mathcal{I}^m} u^m\right) / \boldsymbol{N}^\rho_{\mathcal{I}^m}.$$

   (2.c) Determine $\Delta_j^m = |v_j^m + \frac{1}{\rho} u_j^m|$, $j = 1, \ldots, n-1$ and $\Gamma_i^m = \sqrt{N_i^\rho} |\beta_i^m + d_i^m|$, $i = 1, \ldots, n$.

   (2.d) Update the active and inactive sets by

   $$\mathcal{B}^{m+1} = \left\{ j : \Delta_j^m \geq \Delta_{[k]}^m \right\},$$
   $$\mathcal{J}^{m+1} = \left\{ j : \Delta_j^m < \Delta_{[k]}^m \right\},$$
   $$\mathcal{A}^{m+1} = \left\{ i : \Gamma_i^m \geq \Gamma_{[s]}^m \right\},$$
   $$\mathcal{I}^{m+1} = \left\{ i : \Gamma_i^m < \Gamma_{[s]}^m \right\}.$$

   (2.e) If $\mathcal{A}^{m+1} = \mathcal{A}^m$ and $\mathcal{B}^{m+1} = \mathcal{B}^m$, then stop, else $m = m + 1$ and return to steps (2.a)–(2.d).
3. Output $(u, v) = (u^m, v^m)$ and $(\beta, d) = (\beta^m, d^m)$.

In the AMIAS Algorithm, all computational steps require at most $O(n)$ operations except for the computation of the dual variable u in (2.a). This step involves matrix inverse calculation, which is computationally expensive, especially when n is large. To avoid the calculation of the inverse matrix $(D_{\mathcal{J}^m} D_{\mathcal{J}^m}^\top)^{-1}$, we here adopt the efficient algorithm proposed in Wen et al. (2023). In particular, for a pre-specified number of change points k, assume that $\mathcal{B}^m$ separates the indices $\{1, \ldots, n\}$ into $k+1$ blocks with size $\{n_1, \ldots, n_{k+1}\}$. The matrix $D_{\mathcal{J}^m} D_{\mathcal{J}^m}^\top$ is a block diagonal matrix with $k+1$ blocks and its i-th main-diagonal sub matrix defined by

$$B_{n_i} = \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix} \in \mathbf{R}^{n_i \times n_i}.$$

Then the inverse of the matrix $D_{\mathcal{J}^m} D_{\mathcal{J}^m}^\top$ is given by

$$D_{\mathcal{J}^m} D_{\mathcal{J}^m}^\top = \begin{bmatrix} B_{n_1}^{-1} & 0 & \dots & 0 \\ 0 & B_{n_2}^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_{n_{k+1}}^{-1} \end{bmatrix}, \tag{15}$$

where $B_{n_i}^{-1} = (b_{pq})_{n_i \times n_i}$ with $b_{pq} = \min\{p,q\}(n_i + 1 - \max\{p,q\})/(n_i+1)$. Therefore, the computational cost for the matrix inverse is $O(\sum_{i=1}^{k+1} n_i^2)$ and the overall computational cost of AMIAS algorithm is $O(n + \sum_{i=1}^{k+1} n_i^2)$. For example, the AMIAS algorithm costs $O(n^2)$ when the number of change points is fixed. When the number of change points is $O(n/\log(n))$ and they are equally spaced, the computational cost of the AMIAS algorithm is $O(n + \log(n))$, a linear computational cost.

Two parameters, i.e., k and s, need to be tuned in practical applications, which would need a two-dimensional grid search for an optimal choice if approached naively. To alleviate the computation cost in the tuning parameter determination, we develop a two-step procedure with change point detection in the first step and zero-segment removal in the second step. In this way, there is no need to tune s, and thus the computational time is much reduced. To be specific, we set $\lambda_1 = 0$ in (4) and derive an estimate of β in the first step. Based on the current estimate, we remove segments with the smallest β estimate sequentially. For each removal, we calculate the sparsity Schwartz Information Criterion (sSIC) defined as follows,

$$sSIC(\beta) = n \log\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \beta_i)^2\right) + 2q\log(n), \tag{16}$$

where $q = \#\{\text{non zero coefficient block in } \beta\}$ represents the complexity of the model, with larger values indicating greater complexity. Then, an optimal β estimate with the minimum sSIC value is chosen. We remark that if we discard the penalty in sparsity in β, it reduces to the standard SIC proposed by Yao (1988) for multiple change point detection. The two-step AMIAS algorithm is outlined as follows.

Two-step AMIAS algorithm for constraint least squares problem

1. Specify the maximum number of change points K and the maximum number of iterations $m_{\max}$. Let $\rho = n^2$.
2. For $k = 1, 2, \dots, K$, do
   (2.1) Initialize $\mathcal{B}^0 = \phi$ and $\mathcal{J}^0 = (\mathcal{B}^0)^c$, where ϕ is the empty set.
   (2.2) For $m = 0, 1, 2, \dots, m_{\max}$, do
   (a) Determine u^m and v^m by
   $$u^m_{\mathcal{B}^m} = 0, \text{ and } u^m_{\mathcal{J}^m} = (D_{\mathcal{J}^m} D_{\mathcal{J}^m}^\top)^{-1} D_{\mathcal{J}^m} Y,$$
   $$v^m_{\mathcal{J}^m} = \mathbf{0}, \text{ and } v^m_{\mathcal{B}^m} = D_{\mathcal{B}^m}(Y - D^\top u^m).$$
   (b) Determine $\Delta_j^m = |v_j^m + \frac{1}{\rho} u_j^m|, j = 1, \dots, n-1$.
   (c) Update the active and inactive sets by
   $$\mathcal{B}^{m+1} = \left\{ j : \Delta_j^m \geq \Delta_{[k]}^m \right\}, \ \mathcal{J}^{m+1} = (\mathcal{B}^{m+1})^c.$$
   (d) If $\mathcal{B}^{m+1} = \mathcal{B}^m$, then stop, else $m = m+1$ and return to steps (a)–(c).
   (e) Output $(u(k), v(k)) = (u^m, v^m)$ and $(\beta(k), d(k)) = (\beta^m, d^m)$.
3. Arrange the absolute values of the jump sizes in the $K+1$ segments of $\beta(k)$ in descending order. Denote the ordered segment sets by $S_{(1)}, S_{(2)}, \dots, S_{(K+1)}$, and their sizes by $s_k = |S_{(k)}|$, $k = 1, 2, \dots, K+1$. For $k = 1, 2, \dots, K+1$, calculate
   (a) $\tilde{\beta}(k)_j = \beta(k)_j$ when $j \in S_{(1)}, \dots, S_{(k)}$,
   (b) $\tilde{\beta}(k)_j = 0$ for $j \in S_{(k+1)}, \dots, S_{(K+1)}$.
4. Determine the optimal tuning parameters k^* and $s^* = \sum_{k=1}^{k^*} s_k$ such that $\tilde{\beta}(k^*)$ has the smallest sSIC value. Output $\beta = \tilde{\beta}(k^*)$.

Remark 1. *The two-step AMIAS algorithm provides a simplified strategy for determining the optimal values of tuning parameters k and s in the AMIAS algorithm. As said before, the naive approach requires a two-dimensional grid search, say $\{1, 2, \dots, K\} \times \{1, 2, \dots, S\}$, for choosing an optimal pair for k and s. In the two-step AMIAS algorithm, we only need to search for the best k in $\{1, 2, \dots, K\}$ and determine optimal s by that optimal k. This strategy can reduce the computational cost of finding the tuning parameters substantially.*

3. SIMULATION STUDIES

In this section, we investigate the finite-sample performance of the fused L_0 method and compare it to other competitors. The competing methods include the CBS method proposed by Olshen et al. (2004), the wild binary segmentation (WBS) method introduced by Fryzlewicz (2014) for multiple change point detection, and the fused lasso method proposed by Tibshirani and Wang (2008). In general, the default settings are employed for each method, following their respective proposals.

Given an estimator $\hat{\beta}$ of β, we measure the estimation error by two metrics: the mean squared error (MSE) and the mean absolute deviation (MAD):

$$\text{MSE}(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_i - \beta_i)^2, \quad \text{MAD}(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}|\hat{\beta}_i - \beta_i|.$$

Here n is the number of observations, and $\hat{\beta}_i$ is the estimate of parameter β_i for the i-th observation.

We also report the number of change points detected, denoted by $\hat{q}$. To assess the accuracy of our change point detection, we use the scaled Hausdorff distance defined as

$$(17)\quad d_H = \frac{1}{n}\max\left\{\max_j \min_k |\tau_j - \hat{\tau}_k|, \max_k \min_j |\hat{\tau}_j - \tau_k|\right\},$$

where $\hat{\tau}_k$, $k = 1, \ldots, \hat{q}$ and τ_j, $j = 1, \ldots, q$ are the estimated and true change-point locations, respectively. We can see that $0 \le d_H \le 1$, and a value of d_H close to 0 indicates the good performance of the estimator. We consider the model in (1) with $n = 10{,}000$ and the true signal β having ten non zero blocks of values (2.56, −3.47, 3.02, 3.26, −3.92, −3.12, 1.74, 3.05, −3.09, −3.69) and sizes (35, 18, 79, 62, 51, 27, 84, 32, 26, 19), respectively. Four scenarios, (S1)–(S4), are considered to evaluate the robustness of the fused L_0 method in identifying change points if the assumption of homogenous variance in error term or the normal distribution assumption are violated. In particular, the errors are drawn from standard normal distribution $N(0, 1)$ in scenarios (S1)–(S2). In scenarios (S1) and (S2), the parameter σ is set to be 1 and 2, respectively. These values are chosen to represent a situation with low noise and a situation with low signal strength, respectively. In scenario (S3), the parameter σ varies across different positions such that $\sigma = 2$ when $i = 3001, \ldots, 4000$ and $i = 7001, \ldots, 8000$, and $\sigma = 1$ otherwise. We consider the noise distribution as a t-distribution with degree of freedom of 3 multiplied by a factor 0.5 in scenario (S4). We simulate data from (S1)–(S4), generating 100 replications for each scenario, and record summary results. Figure 1 shows an example of the simulated data from (S1)–(S4).

Table 1 summarizes the mean and standard deviation of the five metrics. Among all methods, cghFLasso produces a significantly larger number of change points on average in all scenarios, indicating its tendency to over-select. This phenomenon is common to methods using the L_1-type penalty (Bühlmann and Van De Geer, 2011).

In scenario (S1), where errors come from a Gaussian distribution with a small and homogeneous variance, WBS has the best performance in terms of all measures except for MAD, followed by the CBS and fused L_0 methods. Compared with CBS and WBS, the fused L_0 method yields a slightly larger yet comparable average value in terms of d_H, particularly when considering the large values in the standard deviations. Additionally, the fused L_0 method bears the smallest MAD among all methods. Therefore, our proposed method yields comparable results to CBS and WBS when the data contains Gaussian error and a high signal.

In scenario (S2), where a weak signal is considered, CBS performs the best, as evidenced by the smallest average value of MSE and d_H. Nevertheless, our fused L_0 method has a considerably small MAD value. Similarly, in scenario (S3), where errors have heterogeneous variance, CBS performs the best, followed by the fused L_0 and WBS methods. Notably, WBS tends to identify slightly more change

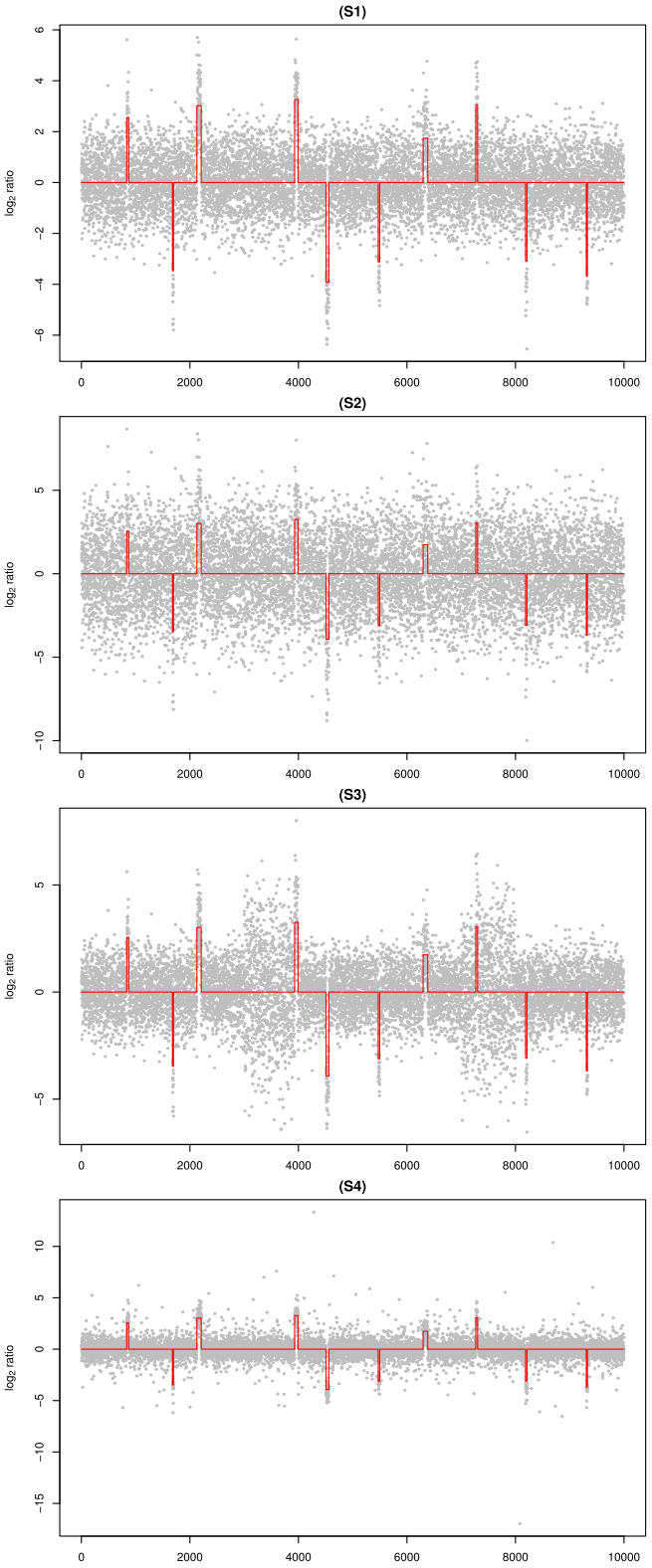

Figure 1. An example using simulated data from scenarios (S1)–(S4).

points on average and has the largest d_H among all methods except for cghFLasso. WBS's success might be because WBS works even for very small jump magnitudes and treats random jumps in segments with higher variance as change points. Again, our fused L_0 does a reasonably good job in this scenario.

When the errors are distributed from a t distribution with $df = 3$ (as observed in scenario (S4)), the performance of WBS deteriorates drastically, as evidenced by it having the largest MSE and the second-largest number of identified change points. CBS has a tendency of under-selecting change points, giving rise to a low value in the number of change points ($\hat{q}$). It should be noted that not detecting a true change point is much more serious than over-selecting by identifying one unnecessary point. Our fused L_0 method identifies an average of 22 change points and gives the lowest d_H among all methods. In terms of estimation accuracy, while CBS yields the smallest MSE, our fused L_0 method has the smallest MAD. This means that the fused L_0 method performs toward the top based on the overall evaluation of estimation and change point detection accuracy.

Compared with scenario (S1), where a Gaussian error is considered, enormous observations occasionally occur in scenario (S4), as demonstrated in Figure 1. These enormous observations can be considered as 'spikes' within a very small region, with magnitudes greater than the original signal. Our approach aims to solve a constrained least squares problem, with constraints imposed on the sparsity of both β and $D\beta$. By constraining $D\beta$, we can treat the 'spikes' as change points with very short spacings between change points, effectively limiting their impact on the least squares loss. The average number of detected change points from our method is 22, which is slightly larger than the true value of 20. In contrast, the cghFLasso method detects a much larger number of change points due to the use of an L_1-type constraint, which can cause bias and over-selection problems similar to other variable selection methods based on the L_1-norm (Behrendt and Schweikert, 2021; Zhao and Yu, 2006).

Overall, the fused L_0 method has comparable and stable performance with different error distributions, which suggests that our proposal is more suitable for practical application than previous methods.

Table 1. Performance of the different methods on detecting change points in the simulated datasets. Within each column, the mean and the standard deviations (inside parenthesis) are recorded

	MSE	MAD	$\hat{q}$	$d_H(\times 10^2)$
(S1) Gaussian error sd=1				
CBS	0.01 (0.00)	0.03 (0.01)	20 (1)	0.29 (0.95)
WBS	0.01 (0.00)	0.03 (0.01)	20 (0)	0.07 (0.48)
cghFlasso	0.02 (0.00)	0.08 (0.01)	97 (33)	6.28 (2.15)
Fused L_0	0.03 (0.01)	0.02 (0.01)	20 (0)	0.35 (0.93)
(S2) Gaussian error sd=2				
CBS	0.03 (0.01)	0.07 (0.01)	20 (1)	0.80 (2.20)
WBS	0.04 (0.01)	0.07 (0.01)	19 (1)	4.01 (5.38)
cghFlasso	0.07 (0.01)	0.17 (0.01)	141 (29)	7.55 (0.84)
Fused L_0	0.09 (0.02)	0.05 (0.01)	18 (2)	7.33 (6.41)
(S3) mixed error				
CBS	0.01 (0.01)	0.04 (0.01)	21 (1)	0.56 (1.22)
WBS	0.02 (0.01)	0.04 (0.01)	21 (3)	1.26 (1.98)
cghFlasso	0.06 (0.01)	0.15 (0.02)	288 (58)	7.64 (0.90)
Fused L_0	0.04 (0.01)	0.03 (0.01)	21 (1)	1.20 (2.41)
(S4) t-distributed error				
CBS	0.04 (0.08)	0.04 (0.03)	18 (5)	7.60 (15.26)
WBS	0.11 (0.10)	0.04 (0.01)	41 (7)	6.33 (1.43)
cghFlasso	0.06 (0.04)	0.08 (0.02)	47 (14)	5.02 (4.05)
Fused L_0	0.05 (0.09)	0.02 (0.01)	21 (1)	1.28 (1.92)

4. REAL DATA ANALYSES

In this section, we apply the aforementioned methods to two array CGH datasets. The first data set contains synthesis data from two patients with diagnosed primal glikblastoma (GBM) and can be downloaded from R package **cghFLasso** (Tibshirani and Wang, 2008). It was constructed such that both local duplications and a large region of deletion exist in the same chromosome; for further information see Tibshirani and Wang (2008).

The second data set comes from the HapMap project (http://www.hapmap.org/), which catalogued SNPs in four of the major ethnic human populations. We downloaded the raw CEL file of 270 individuals from https://www.mathworks.com and obtained the log R ratio values following the preprocessing procedure. Here we use data from one individual as an illustrative example.

The estimators for the GBM data from different methods are illustrated in Figure 2. We can see that while all methods successfully identify both the local duplications and the large region of deletions, WBS seems to be very sensitive to outlier measurements. Compared with cghFLasso, fused L_0 has cleaner and smoother solutions for segmentations, especially in the large region. The fused L_0 approach has estimates similar to those from CBS except for the left region, where a weak alteration signal was detected by fused L_0.

For the HapMap data, we first look at the number of change points found by the tested methods. Again, cghFLasso detected too many change points, a total of 3492. The number of change points detected by our fused L_0 method is 106, which lies between those by CBS (53 change points) and WBS (180 change points). Next, we evaluate the overlapping proportion of the change points identified by our proposal and the other three methods by the Venn diagram in Figure 4. From Figure 4, we can see that 13 common change points are detected by all methods, which indicates these are true change points.

To gain further insight into the differences in these four methods, we plotted the log R ratio estimate from each method in Figure 3. All methods except cghFLasso produce

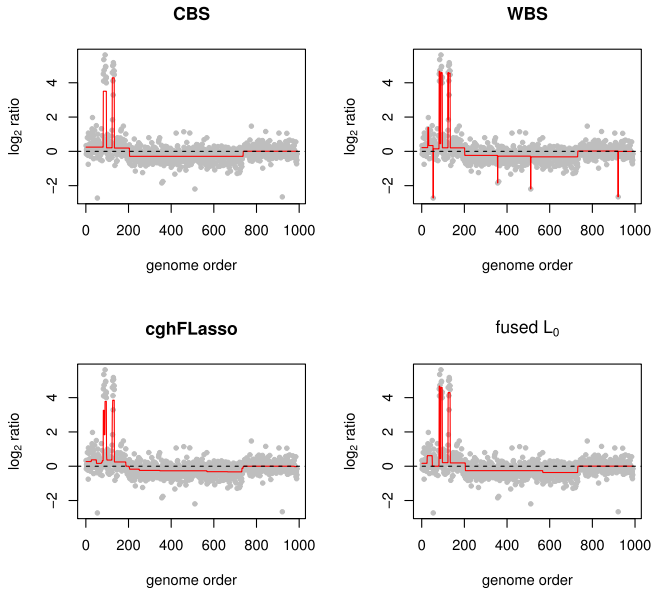

Figure 2. Array CGH profile of the GBM data. The red line in each panel represents the estimated copy number of a particular method, whose name is shown on the top of each sub figure.

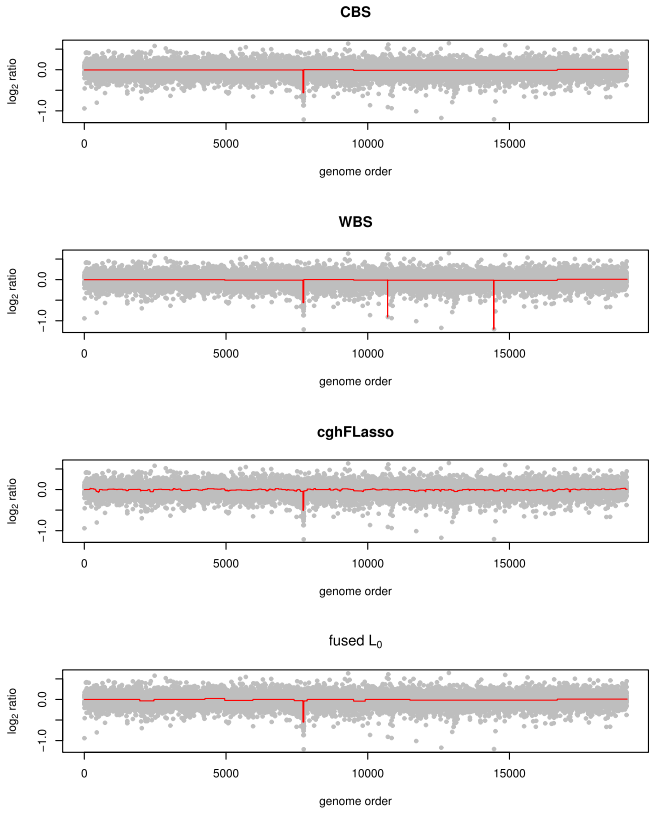

Figure 3. Array CGH profile of the HapMap data. The red line in each panel represents the estimated copy number of a particular method, whose name is shown on the top of each sub figure.

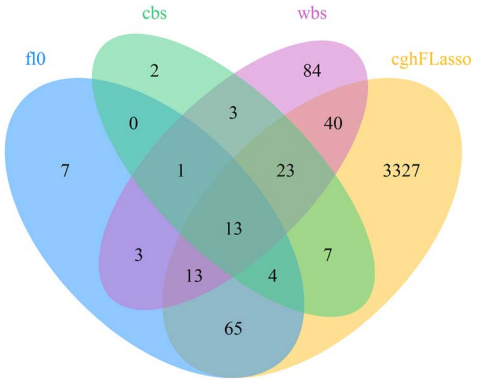

Figure 4. Venn diagram of the detected change points by different methods using the HapMap data.

smooth solutions. The estimate from cghFLasso fluctuates around zero with small jump size, which means that cghFLasso might not control false positives for change point detection in this specific data. WBS has a tendency to pick outliers as change points, as shown in Figure 3. While fused L_0 performs similarly to CBS, it identified more changes with weak jump size, as shown in the left side of the sub figures. This confirms that CBS tends to miss weak signals (Fryzlewicz, 2014).

5. CONCLUSION

In this paper, we propose a novel CNV detection method by formulating the problem as a constraint least squares problem based on the L_0 norm. An efficient algorithm is developed by deriving a necessary optimality condition of its solution. We evaluate the finite-sample performance of the proposal via numerical simulation studies and two analyses of real-world data.

The proposed method can be extended to the whole-genome sequencing data obtained from next-generation sequencing technology. Xi et al. (2011) developed an accurate and efficient CNV detection algorithm via minimizing the Bayesian information criterion (BIC). The algorithm iteratively identifies and merges the most similar pair of bins, using the BIC as the merging and stopping criterion. The BIC can be formulated as penalized log-likelihood function with L_0 penalty on the model complexity, similar to the model we introduce here. The study of how to extend our methods to the whole-genome sequencing data is beyond the scope of the current paper but constitutes an intriguing topic for further research.

ACKNOWLEDGMENTS

The authors would like to thank the editor, associate editor, and referees for their careful comments and useful

suggestions that significantly improved the paper. Wang's research is partially supported by the National Natural Science Foundation of China grants No. 12231017, 72171216, 71921001, and 71991474, and the National Key R&D Program of China No. 2022YFA1003803. Zhang's research is supported in part by Hong Kong General Research Fund (No. 17306519). Wen's research is partially supported by the National Science Foundation of China (No. 12171449) and Fundamental Research Funds for the Central Universities (No. WK3470000027).

APPENDIX A. PROOF OF THEOREM 1

Proof. It can be seen that the Lagrangian function $L(\beta, v, u)$ in (4) is continuously differentiable with respect to u. Since $(\beta^\diamond, v^\diamond, u^\diamond)$ is an alternating solution of minimizing $L(\beta, v, u)$, we have the following equation:

$$\left.\frac{\partial L(\beta, v, u)}{\partial u}\right|_{(\beta^\diamond, v^\diamond, u^\diamond)} = D\beta^\diamond - v^\diamond = 0.$$

It reduces to

$$D\beta^\diamond - v^\diamond = 0. \tag{18}$$

Given $(\beta^\diamond, u^\diamond)$, the Lagrangian function $L(\beta^\diamond, v, u^\diamond)$ can be minimized coordinate-wise with respect to v. Consider the expression

$$l(v_j) \triangleq \lambda|v_j|_0 + \frac{\rho}{2}((D\beta^\diamond)_j - v_j)^2 + u_j^\diamond((D\beta^\diamond)_j - v_j),$$

which is continuously differentiable with respect to v_j except a critical point $v_j = 0$. When $v_j \neq 0$, minimizing $l(v_j)$ leads to

$$v_{\min} = (D\beta^\diamond)_j + \frac{1}{\rho}u_j^\diamond.$$

To check whether $v_{\min}$ is the global minimum, we only need to see whether the inequality $l(v_{\min}) \leq l(0)$ is satisfied. By simple calculation,

$$l(v_{\min}) - l(0) = \lambda - \frac{\rho}{2}\left((D\beta^\diamond)_j + \frac{1}{\rho}u^\diamond\right)^2.$$

Then, given $(\beta^\diamond, u^\diamond)$, the alternating minimum $v_j^\diamond$ is

$$v_j^\diamond = \begin{cases} (D\beta^\diamond)_j + u_j^\diamond/\rho, & \text{if } \left|(D\beta^\diamond)_j + u_j^\diamond/\rho\right| > \sqrt{2\lambda_2/\rho}, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

Given $(v^\diamond, u^\diamond)$, the Lagrangian $L(\beta, v^\diamond, u^\diamond)$ is (up to some constants not related to β):

$$\begin{aligned} &\frac{1}{2}\|Y - \beta\|_2^2 + \lambda_1\|\beta\|_0 + \frac{\rho}{2}\|D\beta - v^\diamond\|_2^2 + (u^\diamond)^T D\beta \\ \propto\ &\frac{1}{2}\left\{\beta^\top(I+\rho D^\top D)\beta - 2(y - D^\top(u^\diamond - \rho v^\diamond))^T\beta\right\} + \lambda_1\|\beta\|_0 \\ \propto\ &\frac{1}{2}\|(I+\rho D^\top D)^{1/2}\beta - (I+\rho D^\top D)^{-1/2}(y - D^\top(u^\diamond - \rho v^\diamond))\|_2^2 \\ &+ \lambda_1\|\beta\|_0 \end{aligned}$$

Let $\tilde{X} = (I + \rho D^\top D)^{1/2}$ and $\tilde{Y} = (I + \rho D^\top D)^{-1/2}(y - D^\top(u^\diamond - \rho v^\diamond))$. Then the minimization of $L(\beta, v^\diamond, u^\diamond)$ with respect to β is equivalent to

$$\min_\beta \frac{1}{2}\|\tilde{X}\beta - \tilde{y}\|_2^2 + \lambda_1\|\beta\|_0.$$

Then given $\beta_{-j}^\diamond = (\beta_1^\diamond, \ldots, \beta_{j-1}^\diamond, \beta_{j+1}^\diamond, \ldots, \beta_n^\diamond)$, the term $\beta_j^\diamond$ can be obtained by

$$\begin{aligned} &\beta_j^\diamond \in \operatorname*{argmin}_t \frac{1}{2}\|\tilde{X}_{-j}\beta_{-j}^\diamond - \tilde{Y} + \tilde{X}_j t\|_2^2 + \lambda_1|t|_0 \\ \Leftrightarrow\ &\beta_j^\diamond \in \operatorname*{argmin}_t \frac{1}{2}|\tilde{X}_j|^2 t^2 + t\tilde{X}_j^\top(\tilde{X}_{-j}\beta_{-j}^\diamond - \tilde{Y}) + \lambda_1|t|_0 \\ \Leftrightarrow\ &\beta_j^\diamond \in \operatorname*{argmin}_t \frac{1}{2}|\tilde{X}_j|^2\left(t + \frac{1}{|\tilde{X}_j|}\tilde{X}_j^\top(\tilde{X}_{-j}\beta_{-j}^\diamond - \tilde{Y})\right)^2 \\ &+ \lambda_1|t|_0 \\ \Leftrightarrow\ &\beta_j^\diamond \in \operatorname*{argmin}_t \frac{1}{2}|\tilde{X}_j|^2\left(t - \beta_j^\diamond + \frac{1}{|\tilde{X}_j|}\tilde{X}_j^\top(\tilde{X}\beta^\diamond - \tilde{Y})\right)^2 \\ &+ \lambda_1|t|_0. \end{aligned}$$

Let $d_j^\diamond = \tilde{X}_j^T(\tilde{Y} - \tilde{X}\beta^\diamond)/|\tilde{X}_j|$ and $d^\diamond = (d_1^\diamond, \ldots, d_n^\diamond)$. Since the term $|\tilde{X}_j|^2$ is the j-th element in the diagonal vector of $\tilde{X}^T\tilde{X} = (I + \rho D^T D)$, we have $|\tilde{X}_j|^2 = 1 + 2\rho$ for $j = 2, \ldots, n-1$, and $|\tilde{X}_j|^2 = 1 + \rho$ for $j = 1, n$. Denote $\boldsymbol{N}^\rho = (N_1^\rho, \ldots, N_n^\rho)$ with $N_j^\rho = 1 + 2\rho$ for $j = 2, \ldots, n-1$, and $N_j^\rho = 1 + \rho$ for $j = 1, n$. Then we have

$$\begin{aligned} d^\diamond &= \tilde{X}^T(\tilde{Y} - \tilde{X}\beta^\diamond)/\boldsymbol{N}^\rho \\ &= (y - D^T(u^\diamond - \rho v^\diamond) - (I + \rho D^T D)\beta^\diamond)/\boldsymbol{N}^\rho, \end{aligned} \tag{20}$$

and

$$\beta_j^\diamond = \begin{cases} \beta_j^\diamond + d_j^\diamond, & \text{if } |\beta_j^\diamond + d_j^\diamond| > \sqrt{2\lambda_1/N_j^\rho} \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

This, together with Equations (18) and (19), completes the proof. □

APPENDIX B. PROOF OF LEMMA 1

Proof. Let $\mathcal{F}_{i,j}^n$ denote the set of piecewise constant functions with change point s at i and j, $1 \leq i < j \leq n$. Then for

fixed i and j ($>i$), the problem of finding the least squares fit is

$$\min_{f_{i,j}^n \in \mathcal{F}_{i,j}^n} \sum_{t=1}^{n} (Y_t - f_{i,j}^n)^2 = \sum_{t=i+1}^{j} (Y_t - \bar{Y}_1)^2 + \sum_{t \le i \text{ ort} > j} (Y_t - \bar{Y}_2)^2,$$

where $\bar{Y}_1 = \sum_{t=i+1}^{j} Y_t/(j-i)$ and $\bar{Y}_1 = \sum_{t \le i \text{ ort} > j} Y_t/(n-j+i)$ are the least squares fit in segments $(i+1, \ldots, j)$ and $(j+1, \ldots, n, 1, \ldots, i)$. Then

$$\begin{aligned}
&\min_{i,j} \min_{f_{i,j}^n \in \mathcal{F}_{i,j}^n} \sum_{t=1}^{n} (Y_t - f_{i,j}^n)^2 \\
\Leftrightarrow\ &\min_{i,j} \sum_{t=i+1}^{j} (Y_t - \bar{Y}_1)^2 + \sum_{t \le i \text{ ort} > j} (Y_t - \bar{Y}_2)^2 \\
\Leftrightarrow\ &\min_{i,j} \sum_{t=i+1}^{j} Y_t^2 - (j-i)\bar{Y}_1^2 + \sum_{t \le i \text{ ort} > j} Y_t^2 - (n-j+i)\bar{Y}_2^2 \\
\Leftrightarrow\ &\min_{i,j} \sum_{t=1}^{n} Y_t^2 - (j-i)\bar{Y}_1^2 - (n-j+i)\bar{Y}_2^2 \\
\Leftrightarrow\ &\max_{i,j} (j-i)\bar{Y}_1^2 + (n-j+i)\bar{Y}_2^2 \\
\Leftrightarrow\ &\max_{i,j} (j-i)\bar{Y}_1^2 + (n-j+i)\bar{Y}_2^2 - \left(\sum_{t=1}^{n} Y_t\right)^2 /n \\
\Leftrightarrow\ &\max_{i,j} (j-i)\bar{Y}_1^2 + (n-j+i)\bar{Y}_2^2 \\
&\qquad - \left((j-i)\bar{Y}_1 + (n-j+i)\bar{Y}_2\right)^2 /n \\
\Leftrightarrow\ &\max_{i,j} \frac{(j-i)(n-j+i)}{n} (\bar{Y}_1 - \bar{Y}_2)^2 \\
\Leftrightarrow\ &\max_{i,j} |Z_{i,j}|^2,
\end{aligned}$$

which completes the proof. □

Received 23 December 2022

REFERENCES

Behrendt, S. and Schweikert, K. (2021). A note on adaptive group lasso for structural break time series. *Econometrics and Statistics* **17** 156–172. MR4214852

Brouwers, N., Van Cauwenberghe, C., Engelborghs, S., Lambert, J.-C., Bettens, K., Le Bastard, N., Pasquier, F., Gil Montoya, A., Peeters, K., Mattheijssens, M., Vandenberghe, Rik, De Deyn, P. P., Cruts, M., Amouyel, P., Sleegers, K. and Van Broeckhoven, C. (2012). Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing c3b/c4b binding sites. *Molecular Psychiatry* **17**(2) 223–233.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Science & Business Media. MR2807761

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of American Statistical Association* **96** 1348–1360. MR1946581

Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Altshuler, D. M., Aburatani, H., Jones, K. W., Tyler-Smith, C., Hurles, M. E., Carter, N. P., Scherer, S. W., and Lee, C. (2006). Copy number variation: new insights in genome diversity. *Genome Research* **16**(8) 949–961.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *Annals of Statistics* **42**(6) 2243–2281. MR3269979

Huang, J., Jiao, Y., Liu, Y. and Lu, X. (2018). A constructive approach to l0 penalized regression. *Journal of Machine Learning Research* **19** 1–37. MR3862417

Ingason, A., Rujescu, D., Cichon, S., Sigurdsson, E., Sigmundsson, T., Pietilainen, O. P. H., Buizer-Voskamp, J. E., Strengman, E., Francks, C., Muglia, P., Gylfason, A., Gustafsson, O., Olason, P. I., Steinberg, S., Hansen, T., Jakobsen, K. D., Rasmussen, H. B., Giegling, I., Moller, H. J., Hartmann, A., Crombie, C., Fraser, G., Walker, N., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Bramon, E., Kiemeney, L. A. L. M., Franke, B., Murray, R., Vassos, E., Toulopoulou, T., Muhleisen, T. W., Tosato, S., Ruggeri, M., Djurovic, S., Andreassen, O. A., Zhang, Z., Werge, T., Ophoff, R. A., Rietschel, M., Nothen, M. M., Petursson, H., Stefansson, H., Peltonen, L., Collier, D., Stefansson, K. and St Clair, D. M. (2011). Copy number variations of chromosome 16p13.1 region associated with schizophrenia. *Molecular Psychiatry* **16**(1) 17–25. MR2956262

Jun Shin, S., Wu, Y. and Hao, N. (2020). A backward procedure for change-point detection with applications to copy number variation detection. *Canadian Journal of Statistics* **48**(3) 366–385. doi: https://doi.org/10.1002/cjs.11535. MR4148604

Niu, Y. S. and Zhang, H. (2012). The screening and ranking algorithm to detect dna copy number variations. *The Annals of Applied Statistics* **6**(3) 1306. MR3012531

Olshen, A. B., Venkatraman, E. S., Lucito, R. and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based dna copy number data. *Biostatistics* **5**(4) 557–572.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y. H., Hicks, J., Spence, S. J., Lee, A. T., Puura, K., Lehtimäki, T., Ledbetter, D., Gregersen, P. K., Bregman, J., Sutcliffe, J. S., Jobanputra, V., Chung, W., Warburton, D., King, M. C., Skuse, D., Geschwind, D. H., Gilliam, T. C., Ye, K. and Wigler, M. (2007). Strong association of de novo copy number mutations with autism. *Science* **316**(5823) 445–449.

Shlien, A. and Malkin, D. (2009). Copy number variations and cancer. *Genome Medicine* **1**(6) 62–62.

Tibshirani, R. and Wang, P. (2008). Spatial smoothing and hot spot detection for cgh data using the fused lasso. *Biostatistics* **9**(1) 18–29.

Tibshirani, R., Saunders, M. A., Rosset, S., Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of The Royal Statistical Society Series B-statistical Methodology* **67**(1) 91–108. MR2136641

Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H. H., Peoples, A., Makarov, V., Yoon, S., Bhandari, A., Corominas, R., Iakoucheva, L. M., Krastoshevsky, O., Krause, V., Larach-Walters, V., Welsh, D. K., Craig, D., Kelsoe, J. R., Gershon, E. S., Leal, S. M., Dell Aquila, M., Morris, D. W., Gill, M., Corvin, A., Insel, P. A., McClellan, J., King, M. C., Karayiorgou, M., Levy, D. L., Delisi L. E. and Sebat, J. (2011). Duplications of the neuropeptide receptor gene vipr2 confer significant risk for schizophrenia. *Nature* **471**(7339) 499–503.

Venkatraman, E. S. and Olshen, A. B. (2007). A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**(6) 657–663.

Vostrikova, L. I. (1981). Detection of the disorder in multidimensional random-processes. *Doklady Akademii Nauk SSSR* **259**(2) 270–274. MR0625215

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F. A., Hakonarson, H. and Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data.

Genome Research **17**(11) 1665–1674.

Wen, C., Zhang, A., Quan, S. and Wang, X. (2017). BeSS: An R package for best subset selection in linear, logistic and CoxPH models. arXiv:1709.06254.

Wen, C., Wang, X. and Zhang, A. (2023). ℓ_0 trend filtering. *INFORMS Journal on Computing*. doi: https://doi.org/10.1287/ijoc.2021.0313.

Xi, R., Hadjipanayis, A. G., Luquette, L. J., Kim, T.-M., Lee, E., Zhang, J., Johnson, M. D., Muzny, D. M., Wheeler, D. A., Gibbs, R. A., Kucherlapati, R. and Park, P. J. (2011). Copy number variation detection in whole-genome sequencing data using the bayesian information criterion. *Proceedings of the National Academy of Sciences of the United States of America* **108**(46) 18583–18584.

Xiao, F., Min, X. and Zhang, H. (2014). Modified screening and ranking algorithm for copy number variation detection. *Bioinformatics* **31**(9) 1341–1348.

Xiu, X., Liu, W., Li, L. and Kong, L. (2019). Alternating direction method of multipliers for nonconvex fused regression problems. *Computational Statistics & Data Analysis* **136** 59–71. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2019.01.002. MR3944672

Yao, Y. (1988). Estimating the number of change-points via Schwarz' criterion. *Statistics and Probability Letters* **6**(3) 181–189. MR0919373

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* **38**(2) 894–942. doi: https://doi.org/10.1214/09-AOS729. MR2604701

Zhao, P. and Yu, B. (2006). On model selection consistency of lasso. *The Journal of Machine Learning Research* **7** 2541–2563. MR2274449

Xiaopu Wang
Department of Statistics and Finance, School of Management
University of Science and Technology of China
Hefei
China
E-mail address: wxp98@mail.ustc.edu.cn

Xueqin Wang
Department of Statistics and Finance, School of Management
University of Science and Technology of China
Hefei
China
E-mail address: wangxq20@ustc.edu.cn

Aijun Zhang
Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam Road
Hong Kong
E-mail address: ajzhang@umich.edu

Canhong Wen
Department of Statistics and Finance, School of Management
University of Science and Technology of China
Hefei
China
E-mail address: wench@ustc.edu.cn