

On Optimal Multiple Changepoint Algorithms for Large Data

Robert Maidstone^{1,†}, Toby Hocking², Guillem Rigai³ and Paul Fearnhead⁴

¹STOR-i Centre for Doctoral Training, Lancaster University

²McGill University and Genome Quebec Innovation Center

³Unité de Recherche en Génomique Végétale (URGV), INRA-CNRS-Université d'Evry

⁴Department of Mathematics and Statistics, Lancaster University

[†]Correspondence: r.maidstone@lancaster.ac.uk

September 11, 2014

Abstract

There is an increasing need for algorithms that can accurately detect changepoints in long time-series, or equivalent, data. Many common approaches to detecting changepoints, for example based on penalised likelihood or minimum description length, can be formulated in terms of minimising a cost over segmentations. Dynamic programming methods exist to solve this minimisation problem exactly, but these tend to scale at least quadratically in the length of the time-series. Algorithms, such as Binary Segmentation, exist that have a computational cost that is close to linear in the length of the time-series, but these are not guaranteed to find the optimal segmentation. Recently pruning ideas have been suggested that can speed up the dynamic programming algorithms, whilst still being guaranteed to find true minimum of the cost function. Here we extend these pruning methods, and introduce two new algorithms for segmenting data, FPOP and SNIP. Empirical results show that FPOP is substantially faster than existing dynamic programming methods, and unlike the existing methods its computational efficiency is robust to the number of changepoints in the data. We evaluate the method at detecting Copy Number Variations and observe that FPOP has a computational cost that is competitive with that of Binary Segmentation.

Keywords: Breakpoints, Dynamic Programming, FPOP, SNIP, Optimal Partitioning, pDPA, PELT, Segment Neighbourhood.

1 Introduction

Often time series data experiences multiple abrupt changes in structure which need to be taken into account if the data is to be modelled effectively. These changes known as changepoints (or equivalently breakpoints) cause the data to be split into segments which can then be modelled separately. Detecting changepoints, both accurately and efficiently, is required in a number of applications including financial data (Fryzlewicz, 2012), climate data (Killick et al., 2012; Reeves et al., 2007), EEG data (Lavielle, 2005) and the analysis of speech signals (Davis et al., 2006).

In Section 7 of this paper we look at detecting changes in DNA copy number in tumour microarray data. Regions in which this copy number is amplified or reduced from a baseline level can relate to tumorous cells and detection

of these areas is crucial for classifying tumour progression and type. Change point methods are widely used in this area (Zhang and Siegmund, 2007; Olshen et al., 2004; Picard et al., 2011) and moreover tumour microarray data has been used as a basis for benchmarking change point techniques both in terms of accuracy (Hocking et al., 2013) and speed (Hocking et al., 2014).

Many approaches to estimating the number and position of change points (e.g. Braun et al., 2000; Davis et al., 2006; Zhang and Siegmund, 2007) can be formulated in terms of defining a cost function for a segmentation. They then either minimise a penalised version of this cost, which we call the penalised minimisation problem; or minimise the cost under a constraint on the number of change points, which we call the constrained minimisation problem. If the cost function depends on the data through a sum of segment-specific costs then the minimisation can be done exactly using dynamic programming (Auger and Lawrence, 1989; Jackson et al., 2005). However these dynamic programming methods have a cost that increases at least quadratically with the amount of data. This large computational cost is an increasingly important issue as every larger data sets are needed to be analysed.

Alternatively, much faster algorithms exist that provide approximate solutions to the minimisation problem. The most widely used of these approximate techniques is Binary Segmentation (Scott and Knott, 1974). This takes a recursive approach, adding change points one at a time. With a new change point added in the position that would lead to the largest reduction in cost given the location of previous change points. Due to its simplicity, Binary Segmentation is computationally efficient, being roughly linear in the amount of data, however it only provides an approximate solution and can lead to poor estimation of the number and position of change points (Killick et al., 2012). Variations of Binary Segmentation, such as Circular Binary Segmentation (Olshen et al., 2004) and Wild Binary Segmentation (Fryzlewicz, 2012), can offer more accurate solutions for slight decreases in the computational efficiency.

A better solution, if possible, is to look at ways of speeding up the dynamic programming algorithms. Recent work has shown this is possible via the pruning of the solution space. Killick et al. (2012) present a technique for doing this which we shall refer to as *inequality based pruning*. This forms the basis of their method PELT which can be used to solve the penalised minimisation problem. Rigaiil (2010) develop a different pruning technique, *functional pruning*, and this is used in their pDPA method which can be used to solve the constrained minimisation problem. Both PELT and pDPA are optimal algorithms, in the sense that they find the true optimum of the minimisation problem they are trying to solve. However the pruning approaches they take are very different, and work well in different scenarios. PELT is most efficient in applications where the number of change points is large, and pDPA when there are few change points.

The focus of this paper is on these pruning techniques, with the aim of trying to combine ideas from PELT and pDPA. This leads to two new algorithms, FPOP and SNIP. The former uses functional pruning to solve the penalised minimisation problem, and the latter uses inequality based pruning to solve the constrained minimisation problem. We further show that FPOP always prunes more than PELT. Empirical results suggest that FPOP is efficient for large data sets regardless of the number of change points, and we observe that FPOP has a computational cost that is even competitive with Binary Segmentation.

The structure of the paper is as follows. We introduce the constrained and penalised optimisation problems for segmenting data in the next section. We then review the existing dynamic programming methods and pruning approaches for solving the penalised optimisation problem in Section 3 and for solving the constrained optimisation problem in Section 4. The new algorithms, FPOP and SNIP, are developed in Section 5, and compared empirically and theoretically with existing pruning methods in Section 6. We then evaluate FPOP empirically on both simulated and CNV data in Section 7. The paper ends with a discussion.

2 Model Definition

Assume we have data ordered by time, though the same ideas extend trivially to data ordered by any other attribute such as position along a chromosome. Denote the data by $\mathbf{y} = (y_1, \dots, y_n)$. We will use that notation that, for $s \geq t$, the set of observations from time t to time s is $\mathbf{y}_{t:s} = (y_t, \dots, y_s)$. If we assume that there are k changepoints in the data, this will correspond to the data being split into $k + 1$ distinct segments. We let the location of the j th changepoint be τ_j for $j = 1, \dots, k$, and set $\tau_0 = 0$ and $\tau_{k+1} = n$. The j th segment will consist of data points $y_{\tau_{j-1}+1}, \dots, y_{\tau_j}$. We let $\tau = (\tau_0, \dots, \tau_{k+1})$ be the set of changepoints.

The statistical problem we are considering is how to infer both the number of changepoints and their locations. The specific details of any approach will depend on the type of change, such as change in mean, variance or distribution, that we wish to detect. However a general framework that encompasses many changepoint detection methods is to introduce a cost function for each segment. The cost of a segmentation can then be defined in terms of the sum of the costs across the segments, and we can infer segmentations through minimising the segmentation cost.

Throughout we will let $\mathcal{C}(\mathbf{y}_{t+1:s})$, for $s \leq t$, denote the cost for a segment consisting of data points y_{t+1}, \dots, y_s . The cost of a segmentation, τ_1, \dots, τ_k is then

$$\sum_{j=0}^k \mathcal{C}(\mathbf{y}_{\tau_j+1:\tau_{j+1}}). \quad (1)$$

The form of this cost function will depend on the type of change we are wanting to detect. One generic approach to defining these segments is to introduce a model for the data within a segment, and then to let the cost be minus the maximum log-likelihood for the data in that segment. If our model is that the data is independent and identically distributed with segment-specific parameter μ then

$$\mathcal{C}(\mathbf{y}_{t+1:s}) = \min_{\mu} \sum_{i=t+1}^s -\log(p(y_i|\mu)). \quad (2)$$

In this formulation we are detecting changes in the value of the parameter, μ , across segments.

For example if μ is the mean in Normally distributed data, with known variance σ^2 , then the cost for a segment would simply be

$$\mathcal{C}(\mathbf{y}_{t+1:s}) = \frac{1}{2\sigma^2} \left[\min_{\mu} \sum_{i=t+1}^s \left(y_i - \frac{1}{s-t} \sum_{j=t+1}^s y_j \right)^2 \right], \quad (3)$$

which is just a quadratic error loss. We have removed a term that does not depend on the data and is linear in segment length, as this term does not affect the optimal segmentation. Note we will get the same optimal segmentation for any choice of $\sigma > 0$ using this cost function.

2.1 Finding the Optimal Segmentation

If we knew the number of changepoints in the data, k , then we could infer their location through minimising (1) over all segmentations with k changepoints. Normally however k is unknown, and also has to be estimated. A common approach is to define

$$C_{k,n} = \min_{\tau} \left[\sum_{j=0}^k \mathcal{C}(\mathbf{y}_{\tau_j+1:\tau_{j+1}}) \right], \quad (4)$$

the minimum cost of a segmenting data $y_{1:n}$ with k changepoints. As k increases we have more flexibility in our model for the data, so often $C_{k,n}$ will be monotonically decreasing in k and estimating the number of changepoints by minimising $C_{k,n}$ is not possible. One solution is to solve (4) for a fixed value of k which is either assumed to be known or chosen separately. We call this problem the *constrained case*.

If k is not known, then a common approach is to calculate $C_{k,n}$ and the corresponding optimal segmentations for a range of values, $k = 0, 1, \dots, K$, where K is some chosen maximum number. We can then estimate the number of changepoints by minimising $C_{k,n} + f(k, n)$ over k for some suitable penalty function $f(k, n)$. The most common choices of $f(k, n)$, for example SIC (Schwarz, 1978) and AIC (Akaike, 1974) are linear in k

If the penalty function is linear in k , with $f(k, n) = \beta k$ for some $\beta > 0$ (which may depend on n), then we can directly find the optimal number of changepoints and segmentation by noting that

$$\begin{aligned} \min_k [C_{k,n} + \beta k] &= \min_{k, \tau} \left[\sum_{j=0}^k \mathcal{C}(\mathbf{y}_{\tau_j+1:\tau_{j+1}}) \right] + \beta k \\ &= \min_{k, \tau} \left[\sum_{j=0}^k \mathcal{C}(\mathbf{y}_{\tau_j+1:\tau_{j+1}}) + \beta \right] - \beta. \end{aligned} \quad (5)$$

We call the minimisation problem in (5) the *penalised case*.

In both the constrained and penalised cases we need to solve a minimisation problem to find the optimal segmentation under our criteria. There are efficient dynamic programming algorithms for solving each of these minimisation problems. For the constrained case this is achieved using the Segment Neighbourhood Search algorithm (see Section 4.1), whilst for the penalised case this can be achieved using the Optimal Partitioning algorithm (see Section 3.1).

Solving the constrained case offers a way to get optimal segmentations for $k = 0, 1, \dots, K$ changepoints, and thus gives insight into how the segmentation varies with the number of segments. However, a big advantage of the penalised case is that it incorporates model selection into the problem itself, and therefore is often computationally more efficient when dealing with an unknown value of k .

2.2 Conditions for Pruning

The focus of this paper is on methods for speeding up these dynamic programming algorithms using pruning methods. The pruning methods can be applied under one of two conditions on the segment costs:

C1 The cost function satisfies

$$\mathcal{C}(\mathbf{y}_{t+1:s}) = \min_{\mu} \sum_{i=t+1}^s \gamma(y_i, \mu),$$

for some function $\gamma(\cdot, \cdot)$, with parameter μ .

C2 There exists a constant κ such that for all $t < s < T$,

$$\mathcal{C}(\mathbf{y}_{t+1:s}) + \mathcal{C}(\mathbf{y}_{s+1:T}) + \kappa \leq \mathcal{C}(\mathbf{y}_{t+1:T}).$$

Condition C1 will be used by functional pruning (which is discussed in Sections 4.2 and 5.1). Condition C2 will be used by the inequality based pruning (Section 3.2 and 5.2). Note that C1 is a stronger condition than C2. If C1 holds then C2 also holds with $\kappa = 0$.

For many practical cost functions these conditions hold; for example it is easily seen that for the negative log-likelihood (2) C1 holds with $\gamma(y_t, \mu) = -\log(p(y_t|\mu))$ and C2 holds with $\kappa = 0$.

3 Solving the Penalised Optimisation Problem

We first consider solving the penalised optimisation problem (5) using a dynamic programming approach. The initial algorithm, Optimal Partitioning (Jackson et al., 2005), will be discussed first before mentioning how pruning can be used to reduce the computational cost.

3.1 Optimal Partitioning

Consider segmenting the data $\mathbf{y}_{1:t}$. Denote $F(t)$ to be the minimum value of the penalised cost (5) for segmenting such data, with $F(0) = -\beta$. The idea of Optimal Partitioning is to split the minimisation over segmentations into the minimisation over the position of the last changepoint, and then the minimisation over the earlier changepoints. We can then use the fact that the minimisation over the earlier changepoints will give us the value $F(\tau^*)$ for some $\tau^* < t$

$$\begin{aligned} F(t) &= \min_{\tau, k} \sum_{j=0}^k [\mathcal{C}(\mathbf{y}_{\tau_j+1:\tau_{j+1}}) + \beta] - \beta, \\ &= \min_{\tau, k} \left\{ \sum_{j=0}^{k-1} [\mathcal{C}(\mathbf{y}_{\tau_j+1:\tau_{j+1}}) + \beta] + \mathcal{C}(\mathbf{y}_{\tau_k+1:t}) + \beta \right\} - \beta, \\ &= \min_{\tau^*} \left\{ \min_{\tau, k'} \sum_{j=0}^{k'} [\mathcal{C}(\mathbf{y}_{\tau_j+1:\tau_{j+1}}) + \beta] - \beta + \mathcal{C}(\mathbf{y}_{\tau^*+1:t}) + \beta \right\}, \\ &= \min_{\tau^*} \{F(\tau^*) + \mathcal{C}(\mathbf{y}_{\tau^*+1:t}) + \beta\}. \end{aligned}$$

Hence we obtain a simple recursion for the $F(t)$ values

$$F(t) = \min_{0 \leq \tau < t} [F(\tau) + \mathcal{C}(\mathbf{y}_{\tau+1:t}) + \beta]. \quad (6)$$

The segmentations themselves can be recovered by first taking the arguments which minimise (6)

$$\tau_t^* = \arg \min_{0 \leq \tau < t} [F(\tau) + \mathcal{C}(\mathbf{y}_{\tau+1:t}) + \beta], \quad (7)$$

which give the optimal location of the last changepoint in the segmentation of $\mathbf{y}_{1:t}$.

If we denote the vector of ordered changepoints in the optimal segmentation of $\mathbf{y}_{1:t}$ by $cp(t)$, with $cp(0) = \emptyset$, then the optimal changepoints up to a time t can be calculated recursively

$$cp(t) = (cp(\tau_t^*), \tau_t^*).$$

As equation (6) is calculated for time steps $t = 1, 2, \dots, n$ and each time step involves a minimisation over $\tau = 0, 1, \dots, t-1$ the computation actually takes $\mathcal{O}(n^2)$ time.

3.2 PELT

One way to increase the efficiency of Optimal Partitioning is discussed in Killick et al. (2012) where they introduce the PELT (Pruned Exact Linear Time) algorithm. PELT works by limiting the set of potential previous changepoints (i.e. the set over which τ is chosen from in the minimisation in equation 6). They show that if Condition C2 holds for some κ , and if

$$F(t) + \mathcal{C}(\mathbf{y}_{(t+1:s)}) + \kappa > F(s), \quad (8)$$

then at any future time $T > s$, t can never be the optimal location of the most recent changepoint prior to T .

This means that at every time step s the left hand side of equation (8) can be calculated for all potential values of the last changepoint. If the inequality holds for any individual t then that t can be discounted as a potential last changepoint for all future times. Thus the update rules (6) and (7) can be restricted to a reduced set of potential last changepoints, τ , to consider. This set, which we shall denote as R_t , can be updated simply by

$$R_{t+1} = \{\tau \in \{R_t \cup \{t\}\} : F(\tau) + \mathcal{C}(\mathbf{y}_{(\tau+1):t}) + \kappa \leq F(t)\}. \quad (9)$$

This pruning technique, which we shall refer to as *inequality based pruning*, forms the basis of the PELT method.

As at each time step in the PELT algorithm the minimisation is being run over fewer values it would be expected that this method would be more efficient than the basic Optimal Partitioning algorithm. In Killick et al. (2012) it is shown to be at least as efficient as Optimal Partitioning, with PELT's computational cost being bounded above by $\mathcal{O}(n^2)$. Under certain conditions the expected computational cost can be shown to be bounded by Ln for some constant $L < \infty$. These conditions are given fully in Killick et al. (2012), the most important of which is that the expected number of changepoints in the data increases linearly with the length of the data, n .

4 Solving the Constrained Optimisation Problem

We now consider solving the constrained optimisation problem (4) using dynamic programming. These methods assume a maximum number of changepoints that are to be considered, K , and then solve the constrained optimisation problem for all values of $k = 1, 2, \dots, K$. We first describe the initial algorithm, Segment Neighbourhood Search (Auger and Lawrence, 1989), and then an approach that uses pruning.

4.1 Segment Neighbourhood Search

Take the constrained case (4) which segments the data up to t , for $t \geq k + 1$, into $k + 1$ segments (using k changepoints), and denote the minimum value of the cost by $C_{k,t}$. The idea of Segment Neighbourhood Search is to derive a relationship between $C_{t,k}$ and $C_{s,k-1}$ for $s < t$:

$$\begin{aligned} C_{k,t} &= \min_{\tau} \sum_{j=0}^k \mathcal{C}(\mathbf{y}_{\tau_j+1:\tau_{j+1}}), \\ &= \min_{\tau_k} \left[\min_{\tau_{1:k-1}} \sum_{j=0}^{k-1} \mathcal{C}(\mathbf{y}_{\tau_j+1:\tau_{j+1}}) + \mathcal{C}(\mathbf{y}_{\tau_k+1:\tau_{k+1}}) \right], \\ &= \min_{\tau_k} [C_{k-1,\tau_k} + \mathcal{C}(\mathbf{y}_{\tau_k+1:\tau_{k+1}})]. \end{aligned}$$

Thus the following recursion is obtained:

$$C_{k,t} = \min_{\tau \in \{k, \dots, t-1\}} [C_{k-1,\tau} + \mathcal{C}(\mathbf{y}_{\tau+1:t})]. \quad (10)$$

If this is run for all values of t up to n and for $k = 2, \dots, K$, then the optimal segmentations with $1, \dots, K$ segments can be acquired.

To extract the optimal segmentation we first let $\tau_l^*(t)$ denote the optimal position of the last changepoint if we segment data $\mathbf{y}_{1:t}$ using l changepoints. This can be calculated as

$$\tau_l^*(t) = \arg \min_{\tau \in \{l, \dots, t-1\}} [C_{l-1,\tau} + \mathcal{C}(\mathbf{y}_{\tau+1:t})].$$

Then if we let $(\tau_1^k, \dots, \tau_k^k)$ be the set of changepoints in the optimal segmentation of $\mathbf{y}_{1:n}$ into $k + 1$ segments, we have $\tau_k^k = \tau_k^*(n)$. Furthermore we can calculate the other changepoint positions recursively for $l = k - 1, \dots, 1$ using

$$\tau_l^k(n) = \tau_l^*(\tau_{l+1}^k).$$

For a fixed value of k equation (10) is computed for $t \in 1, \dots, n$. Then for each t the minimisation is done for $\tau = 1, \dots, t - 1$. This means that $\mathcal{O}(n^2)$ calculations are needed. However, to also identify the optimal number of changepoints this then needs to be done for $k \in 1, \dots, K$ so the total computational cost in time can be seen to be $\mathcal{O}(Kn^2)$.

4.2 Pruned Segment Neighbourhood Search

Rigall (2010) has developed techniques to increase the efficiency of Segment Neighbourhood Search using functional pruning. These form the basis of a method called pruned Dynamic Programming Algorithm (pDPA). A more generic implementation of this method is presented in Cleynen et al. (2012). Here we describe how this algorithm can be used to calculate the $C_{k,t}$ values. Once these are calculated, the optimal segmentation can be extracted as in Segment Neighbourhood Search.

Assuming condition C1, the segment cost function can be split into the component parts $\gamma(y_i, \mu)$, which depend on the parameter μ . We can then define new cost functions, $Cost_{k,t}^\tau(\mu)$, as the minimal cost of segmenting data $y_{1:t}$ into k segments, with a most recent changepoint at τ , and where the segment after τ is conditioned to have parameter μ . Thus for $\tau \leq t - 1$,

$$Cost_{k,t}^\tau(\mu) = C_{k-1,\tau} + \sum_{i=\tau+1}^t \gamma(y_i, \mu), \quad (11)$$

and $Cost_{k,t}^t(\mu) = C_{k-1,t}$.

These functions, which are stored for each candidate changepoint, can then be updated at each new time step as for $\tau \leq t - 1$

$$Cost_{k,t}^\tau(\mu) = Cost_{k,t-1}^\tau(\mu) + \gamma(y_t, \mu). \quad (12)$$

By taking the minimum of $Cost_{k,t}^\tau(\mu)$ over μ , the individual terms of the right hand side of equation (10) can be recovered. Therefore, by further minimising over τ , the minimum cost $C_{k,t}$ can be returned

$$\begin{aligned} \min_{\tau} \min_{\mu} Cost_{k,t}^\tau(\mu) &= \min_{\tau} \min_{\mu} \left[C_{k-1,\tau} + \sum_{i=\tau+1}^t \gamma(y_i, \mu) \right], \\ &= \min_{\tau} \left[C_{k-1,\tau} + \min_{\mu} \sum_{i=\tau+1}^t \gamma(y_i, \mu) \right], \\ &= \min_{\tau} [C_{k-1,\tau} + \mathcal{C}(\mathbf{y}_{\tau+1:t})], \\ &= C_{k,t}. \end{aligned}$$

By interchanging the order of minimisation the values of the potential last changepoint, τ , can be pruned whilst allowing for changes in μ . First we define the function $Cost_{k,t}^*(\mu)$ as follows

$$Cost_{k,t}^*(\mu) = \min_{\tau} Cost_{k,t}^\tau(\mu).$$

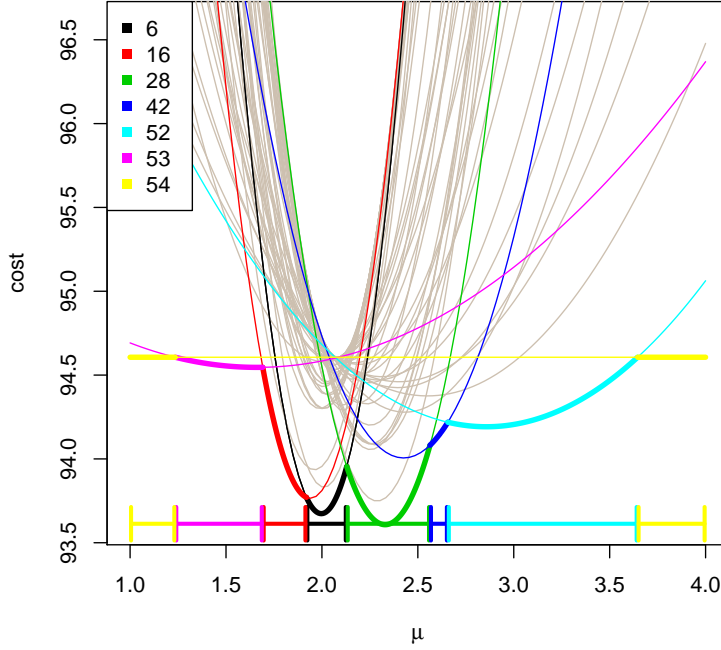


Figure 1: Cost functions, $Cost_{k,\tau}(\mu, t)$ for $\tau = 0, \dots, 46$ and $t = 46$ and the corresponding $C_k^*(\mu, t)$ (in bold) for a change in mean using a least square cost criteria. Coloured lines correspond to $Cost_{k,\tau}(\mu, t)$ that contribute to $C_k^*(\mu, t)$, with the coloured horizontal lines showing the intervals of μ for which each value of τ is such that $Cost_{k,\tau}(\mu, t) = C_k^*(\mu, t)$. Greyed out lines correspond to candidates which have previously been pruned, and do not contribute to $C_k^*(\mu, t)$.

We can now get a recursion for $Cost_{k,t}^*(\mu)$ by splitting the minimisation over the most recent changepoint τ into the two cases $\tau \leq t-1$ and $\tau = t$:

$$\begin{aligned} Cost_{k,t}^*(\mu) &= \min \left\{ \min_{\tau \leq t-1} Cost_{k,t}^\tau(\mu), Cost_{k,t}^t(\mu) \right\} \\ &= \min \left\{ \min_{\tau \leq t-1} Cost_{k,t-1}^\tau(\mu) + \gamma(y_t, \mu), C_{k-1,t} \right\}, \end{aligned}$$

which gives

$$Cost_{k,t}^*(\mu) = \min \{ Cost_{k,t-1}^*(\mu) + \gamma(y_t, \mu), C_{k-1,t} \}.$$

The idea of pDPA is the use this recursion for $Cost_{k,t}^*(\mu)$. We can then use the fact that $C_{k,t} = \min_{\mu} Cost_{k,t}^*(\mu)$ to calculate the $C_{k,t}$ values. In order to do this we need to be able to represent this function of μ in an efficient way. This can be done if μ is a scalar, because for any value of μ , $Cost_{k,t}^*(\mu)$ is equal to the value of $Cost_{k,t}^\tau(\mu)$ for some value of τ . Thus we can partition the possible values of μ into intervals, with each interval corresponding to a value for τ for which $Cost_{k,t}^*(\mu) = Cost_{k,t}^\tau(\mu)$.

To make the idea concrete, an example of $Cost_{k,t}^*(\mu)$ is given in Figure 1 for a change in mean using a least square cost criteria. Each $Cost_{k,t}^\tau(\mu)$ is a quadratic function in this example. In this example there are 6 intervals of μ corresponding to 5 different values of τ for which $Cost_{k,t}^*(\mu) = Cost_{k,t}^\tau(\mu)$. The pDPA algorithm needs to just store the 5 different $Cost_{k,t}^\tau(\mu)$ functions, and the corresponding sets.

Formally speaking we define the set of intervals for which $Cost_{k,t}^*(\mu) = Cost_{k,t}^\tau(\mu)$ as $Set_{k,t}^\tau$. The recursion for $Cost_{k,t}^*(\mu)$ can be used to induce a recursion for these sets. First define:

$$I_{k,t}^\tau = \{ \mu : Cost_{k,t}^\tau(\mu) \leq C_{k-1,t} \}. \quad (13)$$

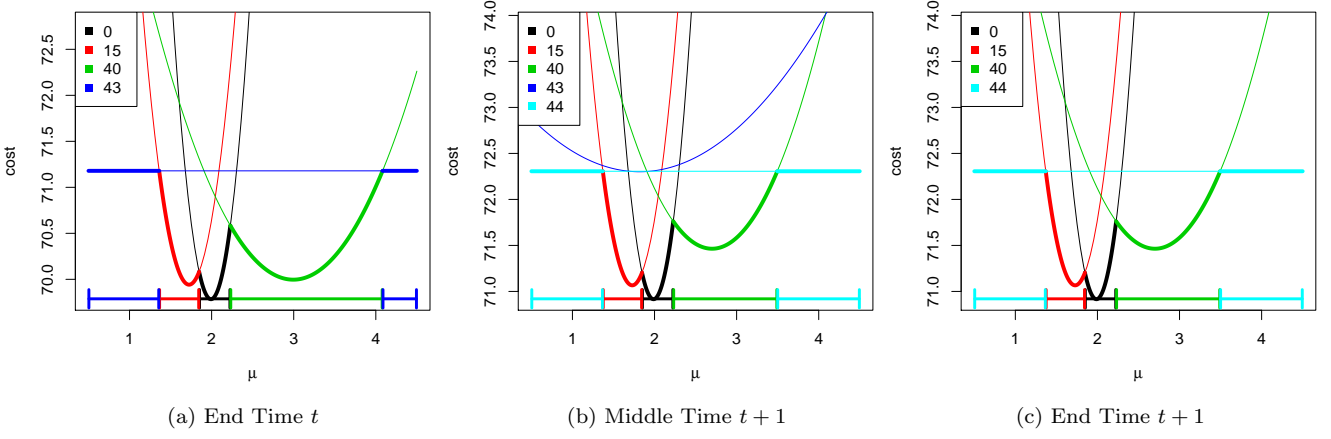


Figure 2: Example of pDPA algorithm over two time-steps. On each plot we show individual $Cost_{k,t}^{\tau}(\mu)$ functions that are stored, together with the intervals (along the bottom) for which each candidate last changepoint is optimal. In bold is the value of $Cost_{k,t}^{*}(\mu)$. For this example $t = 43$ and we are detecting a change in mean (see Section 2). (a) 4 candidates are optimal for some interval of μ , however at $t = 44$ (b), when the candidate functions are updated and the new candidate is added, then the candidate $\tau = 43$ is no longer optimal for any μ and hence can be pruned (c).

Then, for $\tau \leq t - 1$ we have

$$\begin{aligned} Set_{k,t}^{\tau} &= \{\mu : Cost_{k,t}^{\tau}(\mu) = Cost_{k,t}^{*}(\mu)\} \\ &= \{\mu : Cost_{k,t-1}^{\tau}(\mu) + \gamma(y_t, \mu) = \min\{Cost_{k,t-1}^{*}(\mu) + \gamma(y_t, \mu), C_{k-1,t}\}\}. \end{aligned}$$

Remembering that $Cost_{k,t-1}^{\tau}(\mu) + \gamma(y_t, \mu) \geq Cost_{k,t-1}^{*}(\mu) + \gamma(y_t, \mu)$, we have that for μ to be in $Set_{k,t}^{\tau}$ we need that $Cost_{k,t-1}^{\tau}(\mu) = Cost_{k,t-1}^{*}(\mu)$, and that $Cost_{k,t-1}^{\tau}(\mu) + \gamma(y_t, \mu) \leq C_{k-1,t}$. The former condition corresponds to μ being in $Set_{k,t-1}^{\tau}$ and the second that μ is in $I_{k,t}^{\tau}$. So for $\tau \leq t - 1$

$$Set_{k,t}^{\tau} = Set_{k,t-1}^{\tau} \cap I_{k,t}^{\tau}.$$

If this $Set_{k,t}^{\tau} = \emptyset$ then the value τ can be pruned, as $Set_{k,T}^{\tau} = \emptyset$ for all $T > t$.

If we denote the range of values μ can take to be D , then we further have that

$$Set_{k,t}^t = D \setminus \left[\bigcup_{\tau} I_{k,t}^{\tau} \right],$$

where t can be pruned straight away if $Set_{k,t}^t = \emptyset$.

An example of the pDPA recursion is given in Figure 2 for a change in mean using a least square cost criteria. The left-hand plot shows $Cost_{k,t}^{*}(\mu)$. In this example there are 5 intervals of μ corresponding to 4 different values of τ for which $Cost_{k,t}^{*}(\mu) = Cost_{k,t}^{\tau}(\mu)$. When we analyse the next data point, we update each of these four $Cost_{k,t}^{\tau}(\mu)$ functions, using $Cost_{k,t+1}^{\tau}(\mu) = Cost_{k,t}^{\tau}(\mu) + \gamma(y_{t+1}, \mu)$, and introduce a new curve corresponding to a change-point at time $t + 1$, $Cost_{k,t+1}^{t+1}(\mu) = C_{k-1,t+1}$ (see middle plot). We can then prune the functions which are no longer optimal for any μ values, and in this case we remove one such function (see right-hand plot).

pDPA can be shown to be bounded in time by $\mathcal{O}(Kn^2)$. Rigail (2010) further analyses the time complexity of pDPA and shows it empirically to be $\mathcal{O}(Kn \log n)$, further indications towards this will be presented in Section 7. However pDPA has a computational overhead relative to Segment Neighbourhood Search, as it requires calculating

and storing the $Cost_{k,t}^\tau(\mu)$ functions and the corresponding sets $Set_{k,t}^\tau$. Currently implementations of pDPA have only been possible for models with scalar segment parameters μ , due to the difficulty of calculating the sets in higher dimensions. Being able to efficiently store and update the $Cost_{k,t}^\tau(\mu)$ have also restricted application primarily to models where $\gamma(y, \mu)$ corresponds to the log-likelihood of an exponential family. However this still includes a wide-range of changepoint applications, including that of detecting CNVs that we consider in Section 7. The cost of updating the sets depends heavily on whether the updates (13) can be calculated analytically, or whether they require the use of numerical methods.

5 New Changepoint Algorithms

Two natural ways of extending the two methods introduced above will be examined in this section. These are, respectively, to apply functional pruning (Section 4.2) to Optimal Partitioning, and to apply inequality based pruning (Section 3.2) to Segment Neighbourhood Search. These lead to two new algorithms, which we call Functional Pruning Optimal Partitioning (FPOP) and Segment Neighbourhood with Inequality Pruning (SNIP).

5.1 Functional Pruning Optimal Partitioning

Functional Pruning Optimal Partitioning (FPOP) provides a version of Optimal Partitioning (Jackson et al., 2005) which utilises functional pruning to increase the efficiency. As will be discussed in Section 6 and shown in Section 7 FPOP provides an alternative to PELT which is more efficient in certain scenarios. The approach used by FPOP is similar to the approach for pDPA in Section 4.2, however the theory is slightly simpler here as there is no longer the need to condition on the number of changepoints.

We assume condition C1 holds, that the cost function, $\mathcal{C}(\mathbf{y}_{\tau+1:t})$, can be split into component parts $\gamma(y_i, \mu)$ which depend on the parameter μ . Cost functions $Cost_t^\tau$ can then be defined as the minimal cost of the data up to time t , conditional on the last changepoint being at τ and the last segment having parameter μ . Thus for $\tau \leq t - 1$

$$Cost_t^\tau(\mu) = F(\tau) + \beta + \sum_{i=\tau+1}^t \gamma(y_i, \mu), \quad (14)$$

and $Cost_t^t(\mu) = F(t) + \beta$.

These functions which only need to be stored for each candidate changepoint can then be recursively updated at each time step, $\tau \leq t - 1$

$$Cost_t^\tau(\mu) = Cost_{t-1}^\tau(\mu) + \gamma(y_t, \mu). \quad (15)$$

Given the cost functions $Cost_t^\tau(\mu)$ the minimal cost $F(t)$ can be returned by minimising over both τ and μ :

$$\begin{aligned} \min_{\tau} \min_{\mu} Cost_t^\tau(\mu) &= \min_{\tau} \min_{\mu} \left[F(\tau) + \beta + \sum_{i=\tau+1}^t \gamma(y_i, \mu) \right], \\ &= \min_{\tau} \left[F(\tau) + \beta + \min_{\mu} \sum_{i=\tau+1}^t \gamma(y_i, \mu) \right], \\ &= \min_{\tau} [F(\tau) + \beta + \mathcal{C}(\mathbf{y}_{\tau+1:t})], \\ &= F(t). \end{aligned}$$

As before, by interchanging the order of minimisation, the values of the potential last changepoint, τ , can be pruned whilst allowing for a varying μ . Firstly we will define the function $Cost_t^*(\mu)$, the minimal cost of segmenting data

$y_{1:t}$ conditional on the last segment having parameter μ :

$$Cost_t^*(\mu) = \min_{\tau} Cost_t^{\tau}(\mu).$$

We will update these functions recursively over time, and use $F(t) = \min_{\mu} Cost_t^*(\mu)$ to then obtain the solution of the penalised minimisation problem. The recursions for $Cost_t^*(\mu)$ are obtained by splitting the minimisation over τ into $\tau \leq t-1$ and $\tau = t$

$$\begin{aligned} Cost_t^*(\mu) &= \min \left\{ \min_{\tau \leq t-1} Cost_t^{\tau}(\mu), Cost_t^t(\mu) \right\}, \\ &= \min \left\{ \min_{\tau \leq t-1} Cost_{t-1}^{\tau}(\mu) + \gamma(y_t, \mu), Cost_t^t(\mu) \right\}, \end{aligned}$$

which then gives:

$$Cost_t^*(\mu) = \min\{Cost_{t-1}^*(\mu) + \gamma(y_t, \mu), F(t) + \beta\}.$$

To implement this recursion we need to be able to efficiently store and update $Cost_t^*(\mu)$. As before we do this by partitioning the space of possible μ values, D , into sets where each set corresponds to a value τ for which $Cost_t^*(\mu) = Cost_t^{\tau}(\mu)$. We then need to be able to update these sets, and store $Cost_t^{\tau}(\mu)$ just for each τ for which the corresponding set is non-empty.

This can be achieved by first defining

$$I_t^{\tau} = \{\mu : Cost_t^{\tau}(\mu) \leq F(t) + \beta\}. \quad (16)$$

Then, for $\tau \leq t-1$, we define

$$\begin{aligned} Set_t^{\tau} &= \{\mu : Cost_t^{\tau}(\mu) = Cost_t^*(\mu)\} \\ &= \{\mu : Cost_{t-1}^{\tau}(\mu) + \gamma(y_t, \mu) = \min\{Cost_{t-1}^*(\mu) + \gamma(y_t, \mu), F(t) + \beta\}\} \end{aligned}$$

Remembering that $Cost_{t-1}^{\tau}(\mu) + \gamma(y_t, \mu) \geq Cost_{t-1}^*(\mu) + \gamma(y_t, \mu)$, we have that for μ to be in Set_t^{τ} we need that $Cost_{t-1}^{\tau}(\mu) = Cost_{t-1}^*(\mu)$, and that $Cost_{t-1}^{\tau}(\mu) + \gamma(y_t, \mu) \leq F(t) + \beta$. The former condition corresponds to μ being in Set_{t-1}^{τ} and the second that μ is in I_t^{τ} , so for $\tau \leq t-1$

$$Set_t^{\tau} = Set_{t-1}^{\tau} \cap I_t^{\tau}.$$

If this $Set_t^{\tau} = \emptyset$ then the value τ can be pruned, as then $Set_T^{\tau} = \emptyset$ for all $T > t$.

If we denote the range of values μ can take to be D , then we further have that

$$Set_t^t = D \setminus \left[\bigcup_{\tau} I_t^{\tau} \right],$$

where t can be pruned straight away if $Set_t^t = \emptyset$.

This updating of the candidate functions and sets is illustrated in Figure 3 where the $Cost$ functions and Set intervals are displayed across two time steps. In this example a change in mean has been considered, using a least squares cost. The bold line on the left-hand graph corresponds to the function $Cost_t^*(\mu)$ and is made up of 7 pieces which relate to 6 candidate last changepoints. As the next time point is analysed the six $Cost_t^{\tau}(\mu)$ functions are updated using the formula $Cost_{t+1}^{\tau}(\mu) = Cost_t^{\tau}(\mu) + \gamma(y_{t+1}, \mu)$ and a new function, $Cost_{t+1}^{t+1}(\mu) = F(t+1) + \beta$, is introduced corresponding to placing a changepoint at time $t+1$ (see middle plot). The functions which are no longer optimal for any values of μ (i.e. do not form any part of $Cost_{t+1}^*(\mu)$) can then be pruned, and one such function is removed in the right-hand plot.

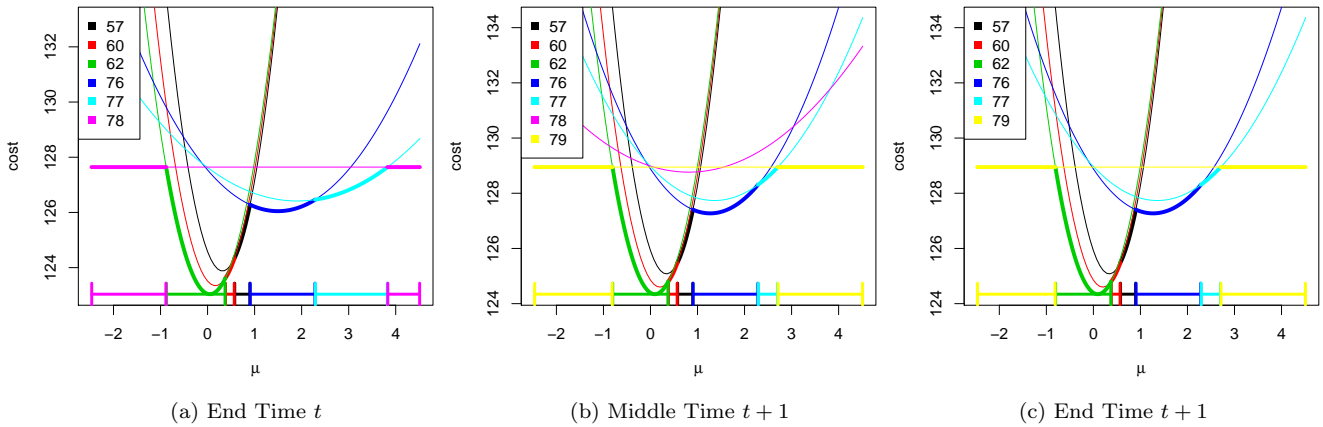


Figure 3: Candidate functions over two time steps, the intervals shown along the bottom correspond to the intervals of μ for which each candidate last changepoint is optimal. When $t = 78$ (a) 6 candidates are optimal for some interval of μ , however at $t = 79$ (b), when the candidate functions are updated and the new candidate is added, then candidate $\tau = 78$ is no longer optimal for any μ and hence can be pruned (c).

Once again we denote the set of potential last changes to consider as R_t and then restrict the update rules (6) and (7) to $\tau \in R_t$. This set can then be recursively updated at each time step

$$R_{t+1} = \{\tau \in \{R_t \cup \{t\}\} : \text{Set}_t^\tau \neq \emptyset\}. \quad (17)$$

These steps can then be applied directly to an Optimal Partitioning algorithm to form the FPOP method and the full pseudocode for this is presented in Algorithm 1.

5.2 Segment Neighbourhood with Inequality Pruning

In a similar vein to Section 5.1, Segment Neighbourhood Search can also benefit from using pruning methods. In Section 4.2 the method pDPA was discussed as a fast pruned version of Segment Neighbourhood Search. In this section a new method, Segment Neighbourhood with Inequality Pruning (SNIP), will be introduced. This takes the Segment Neighbourhood Search algorithm and uses inequality based pruning to increase the speed.

Under condition (C2) the following result can be proved for Segment Neighbourhood Search and this will enable points to be pruned from the candidate changepoint set.

Theorem 5.1 *Assume that there exists a constant, κ , such that condition C2 holds. If, for any $k \geq 1$ and $t < s$*

$$C_{k-1,t} + \mathcal{C}(\mathbf{y}_{t+1:s}) + \kappa > C_{k-1,s} \quad (18)$$

then at any future time $T > s$, t cannot be the position of the last changepoint in the optimal segmentation of $y_{1:T}$ with k changepoints.

Proof. The idea of the proof is to show that a segmentation of $y_{1:T}$ into k segments with the last changepoint at s will be better than one with the last changepoint at t for all $T > s$.

Algorithm 1: Functional Pruning Optimal Partitioning (FPOP)

Input : Set of data of the form $\mathbf{y}_{1:n} = (y_1, \dots, y_n)$,

A measure of fit $\gamma(\cdot, \cdot)$ dependent on the data and the mean,

A penalty β which does not depend on the number or location of the changepoints.

Let $n = \text{length of data}$, and set $F(0) = -\beta$, $cp(0) = 0$;

then let $R_1 = \{0\}$;

and set $D = \text{the range of } \mu$;

$Set_0^0 = D$;

$Cost_0^0(\mu) = F(0) + \beta = 0$;

for $t = 1, \dots, n$ **do**

for $\tau \in R_t$ **do**

$Cost_t^\tau(\mu) = Cost_{t-1}^\tau(\mu) + \gamma(y_t, \mu)$;

 Calculate $F(t) = \min_{\tau \in R_t} (\min_{\mu \in Set_t^\tau} [Cost_t^\tau(\mu)])$;

 Let $\tau_t = \arg \min_{\tau \in R_t} (\min_{\mu \in Set_t^\tau} [Cost_t^\tau(\mu)])$;

 Set $cp(t) = (cp(\tau_t), \tau_t)$;

$Cost_t^t(\mu) = F(t) + \beta$;

$Set_t^t = D$;

for $\tau \in R_t$ **do**

$I_t^\tau = \{\mu : Cost_t^\tau(\mu) \leq F(t) + \beta\}$;

$Set_t^\tau = Set_{t-1}^\tau \cap I_t^\tau$;

$Set_t^t = Set_t^t \setminus I_t^\tau$;

$R_{t+1} = \{\tau \in \{R_t \cup \{t\}\} : Set_t^\tau \neq \emptyset\}$;

Output: The changepoints recorded in $cp(n)$.

Assume that (18) is true. Now for any $s < T \leq n$

$$\begin{aligned} C_{k-1,t} + \mathcal{C}(\mathbf{y}_{t+1:s}) + \kappa &> C_{k-1,s}, \\ C_{k-1,t} + \mathcal{C}(\mathbf{y}_{t+1:s}) + \kappa + \mathcal{C}(\mathbf{y}_{s+1:T}) &> C_{k-1,s} + \mathcal{C}(\mathbf{y}_{s+1:T}), \\ C_{k-1,t} + \mathcal{C}(\mathbf{y}_{t+1,T}) &> C_{k-1,s} + \mathcal{C}(\mathbf{y}_{s+1,T}), \end{aligned} \quad (\text{by C2}).$$

Therefore for any $T > s$ the cost $C_{k-1,t} + \mathcal{C}(\mathbf{y}_{t+1,T}) > C_{k,T}$ and hence t cannot be the optimal location of the last changepoint when segmenting $\mathbf{y}_{1:T}$ with k changepoints. \blacksquare

Theorem 5.1 implies that the update rule (10) can be restricted to a reduced set over τ of potential last changes to consider. This set, which we shall denote as $R_{k,t}$, can be updated simply by

$$R_{k,t+1} = \{v \in \{R_{k,t} \cup \{t\}\} : C_{k-1,v} + \mathcal{C}(\mathbf{y}_{v+1,t}) + \kappa < C_{k-1,t}\}. \quad (19)$$

This new algorithm, SNIP, is described fully in Algorithm 2.

6 Comparisons Between Pruning Methods

Functional and inequality based pruning both offer increases in the efficiency in solving both the penalised and constrained problems, however their use depends on the assumptions which can be made on the cost function.

Algorithm 2: Segment Neighbourhood with Inequality Pruning (SNIP)

Input : Set of data of the form $\mathbf{y}_{1:n} = (y_1, \dots, y_n)$,

A measure of fit $\mathcal{C}(\cdot)$ dependent on the data (needs to be minimised),

An integer, K , specifying the maximum number of changepoints to find,

A constant κ that satisfies: $\mathcal{C}(\mathbf{y}_{t+1:s}) + \mathcal{C}(\mathbf{y}_{s+1:T}) + \kappa \leq \mathcal{C}(\mathbf{y}_{t+1:T})$.

Let n =length of data;

Set $C_{0,t} = \mathcal{C}(\mathbf{y}_{1:t})$, for all $t \in \{1, \dots, n\}$;

for $k = 1, \dots, K$ **do**

Set $R_{k,k+1} = \{k\}$. **for** $t = k + 1, \dots, n$ **do**

 Calculate $C_{k,t} = \min_{v \in R_{k,t}} (C_{k-1,v} + \mathcal{C}(\mathbf{y}_{v+1:t}))$;

 Set $R_{k,t+1} = \{v \in \{R_{k,t} \cup \{t\}\} : C_{k-1,v} + \mathcal{C}(\mathbf{y}_{v+1,t}) + \kappa < C_{k-1,t}\}$;

Set $\tau_{k,1} = \arg \min_{v \in R_{k,n}} (C_{k-1,v} + \mathcal{C}(\mathbf{y}_{v+1,n}))$;

for $i = 2, \dots, k$ **do**

 Let $\tau_{k,i} = \arg \min_{v \in R_{k-i, \tau_{k,i-1}}} (C_{k-1,v} + \mathcal{C}(\mathbf{y}_{v+1, \tau_{k,i-1}}))$;

Output: For $k = 0, \dots, K$: the total measure of fit, $C_{k,n}$, for k changepoints and the location of the changepoints for that fit, $\boldsymbol{\tau}_{k,(1:k)}$.

Inequality based pruning is dependent on the assumption C2, while functional pruning requires the slightly stronger condition C1. Furthermore, inequality based pruning has a large computational overhead, and is currently only feasible for detecting a change in a univariate parameter.

If we consider models for which both pruning methods can be implemented, we can compare the extent to which the methods prune. This will give some insight into when the different pruning methods would be expected to work well. To explore this in Figures 4 and 5 we look at the amount of candidates stored by functional and inequality based pruning in each of the two optimisation problems.

As Figure 4 illustrates, PELT prunes very rarely; only when evidence of a change is particularly high. In contrast, FPOP prunes more frequently keeping the candidate set small throughout. Figure 5 shows similar results for the constrained problem. While pDPA constantly prunes, SNIP only prunes sporadically. In addition SNIP fails to prune much at all for low values of k .

Figures 4 and 5 give strong empirical evidence that functional pruning prunes more points than the inequality based method. In fact it can be shown that any point pruned by inequality based pruning will also be pruned at the same time step by functional pruning. This result holds for both the penalised and constrained case and is stated formally in Theorem 6.1.

Theorem 6.1 *Let $\mathcal{C}(\cdot)$ be a cost function that satisfies condition C1, and consider solving either the constrained or penalised optimisation problem using dynamic programming and either inequality or functional pruning.*

Any point pruned by inequality based pruning at time t will also have been pruned by functional pruning at the same time.

Proof. We prove this for pruning of optimal partitioning, with the ideas extending directly to the pruning of the Segment Neighbourhood algorithm.

For a cost function which can be decomposed into pointwise costs, it's clear that condition C2 holds when $\kappa = 0$ and

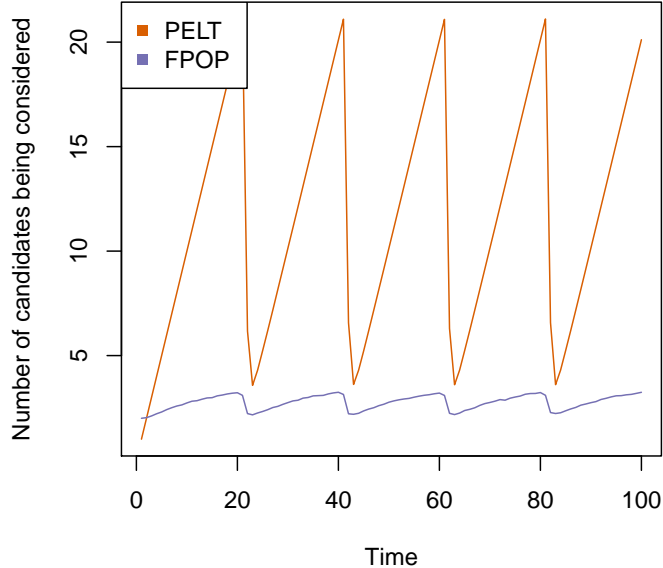


Figure 4: Comparison of the number of candidate changepoints stored over time by FPOP (purple) and PELT (orange). Averaged over 1000 data sets with changepoints at $t = 20, 40, 60$ and 80 .

hence inequality based pruning can be used. Recall that the point τ (where $\tau < t$, the current time point) is pruned by inequality based pruning in the penalised case if

$$F(\tau) + \mathcal{C}(\mathbf{y}_{\tau+1:t}) \geq F(t),$$

Then, by letting $\hat{\mu}_\tau$ be the value of μ such that $Cost_t^\tau(\mu)$ is minimised, this is equivalent to

$$Cost_t^\tau(\hat{\mu}_\tau) - \beta \geq F(t),$$

Which can be generalised for all μ to

$$Cost_t^\tau(\mu) \geq F(t) + \beta.$$

Therefore equation (16) holds for no value of μ and hence $I_t^\tau = \emptyset$ and furthermore $Set_{t+1}^\tau = Set_t^\tau \cap I_t^\tau = \emptyset$ meaning that τ is pruned under functional pruning. ■

7 Empirical evaluation of FPOP

As explained in Section 6 functional pruning leads to a better pruning in the following sense: any point pruned by inequality based pruning will also be pruned by functional pruning. However, functional pruning is computationally more demanding than inequality based pruning. We thus decided to empirically compare the performance of FPOP to PELT, pDPA and Binary Segmentation (Binseg).

To do so, we implement FPOP for the quadratic loss in C++. More precisely, we consider the quadratic cost (3). We assess the runtimes of FPOP on both real microarray data as well as synthetic data. All algorithms were implemented in C++. For pDPA and Binary Segmentation we had to input a maximum number of changepoints to search for, which we denote K as before.

Note that the Binary Segmentation heuristic is computationally extremely fast. That is because its complexity is on average $\mathcal{O}(n \log K)$. Furthermore, it relies on a few fairly simple operations and hence the constant in front of this

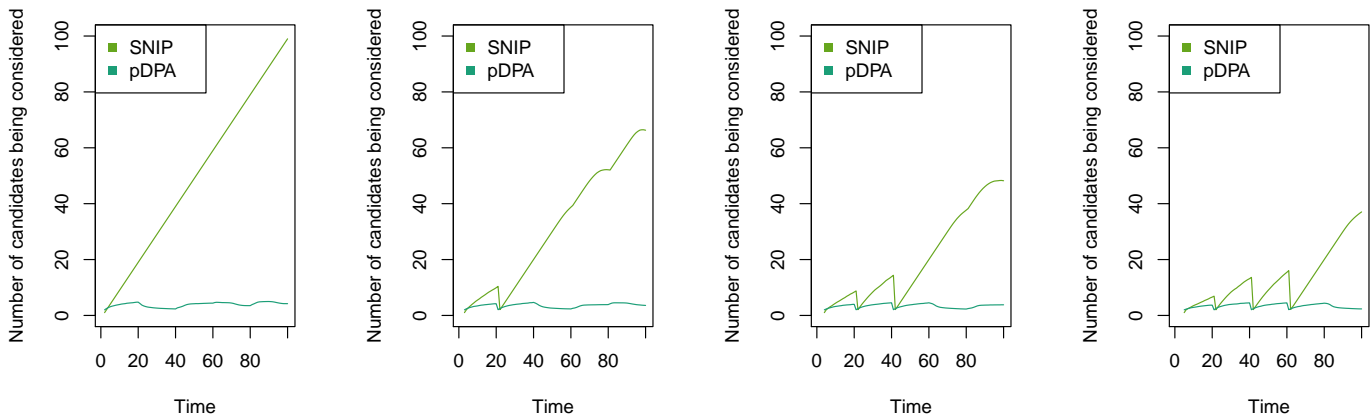


Figure 5: Comparison of the number of candidate changepoints stored over time by pDPA (teal) and SNIP (green) at multiple values of k in the algorithms (going from left to right $k = 2, 3, 4, 5$). Averaged over 1000 data sets with changepoints at $t = 20, 40, 60$ and 80 .

big \mathcal{O} notation is expected to be quite low. For this reason we believe that Binary Segmentation is good reference point in terms of speed and we do not think it is possible to be much faster. However, unlike the other algorithms, Binary Segmentation is not guaranteed to find the optimal segmentation. For a fuller investigation of the loss of accuracy that can occur when using Binary Segmentation see Killick et al. (2012).

7.1 Speed benchmark: 4467 chromosomes from tumour microarrays

Hocking et al. (2014) proposed to benchmark the speed of segmentation algorithms on a database of 4467 problems of size varying from 25 to 153662 data points. These data come from different microarrays data sets (Affymetrix, Nimblegen, BAC/PAC) and different tumour types (leukaemia, lymphoma, neuroblastoma, medulloblastoma).

We compared FPOP to several other segmentation algorithms: pDPA (Rigail, 2010), PELT (Killick et al., 2012), and Binary Segmentation (Binseg). For such data, we expect a small number of changes and for all profiles we ran pDPA and Binseg with a maximum number of changes $K = 52$.

We used the R `system.time` function to measure the execution time of all 4 algorithms on all 4467 segmentation problems. The R source code for these timings is in `benchmark/systemtime.arrays.R` in the opfp project repository on R-Forge: <https://r-forge.r-project.org/projects/opfp/>

It can be seen in Figure 6 that in general FPOP is faster than PELT and pDPA, but about two times slower than Binseg. Note that `system.time` is inaccurate for small times due to rounding, as pointed out in Figure 6 middle and right. For these small profiles we also assessed the performances of FPOP, PELT and Binseg using the `microbenchmark` package and confirmed that FPOP is faster than PELT. For these small problems we were surprised to observe that FPOP exhibited about the same speed as Binseg (microbenchmark results figure not shown).

7.2 Speed benchmark: simulated data with different number of changes

The speed of PELT, Binseg and pDPA depends on the underlying number of changes. For pDPA and Binseg the relationship is clear as to cope with a larger number of changes one need to increase the maximum number of changes to look for K . For a fixed size signal the runtime dependency is expected to be in $\mathcal{O}(\log K)$ for Binseg and in $\mathcal{O}(K)$ for pDPA.

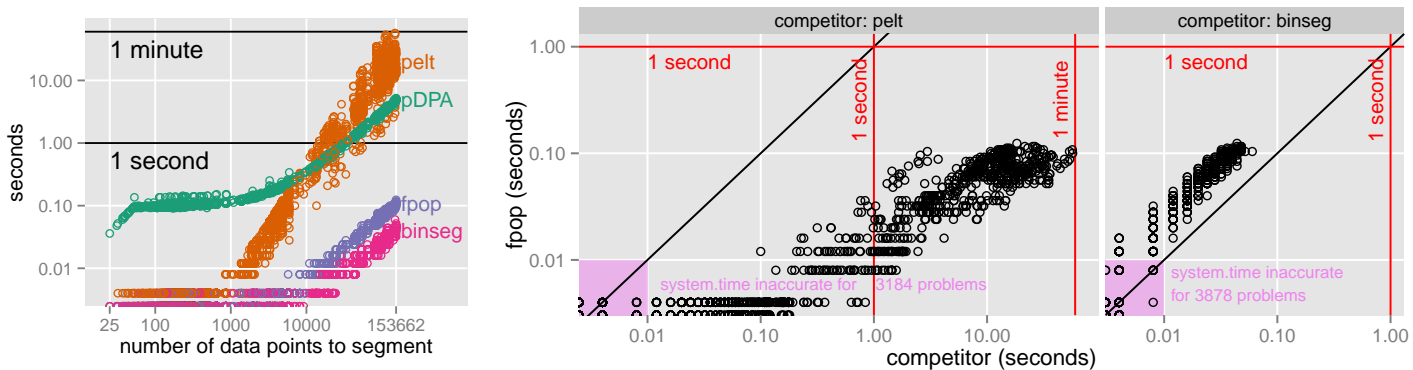


Figure 6: (Left) Runtimes of FPOP, PELT, pDPA and Binseg as a function of the length of the profile using on the tumour micro array benchmark. (Middle) Runtimes of PELT and FPOP for the same profiles. (Right) Runtimes of Binseg and FPOP for the same profiles.

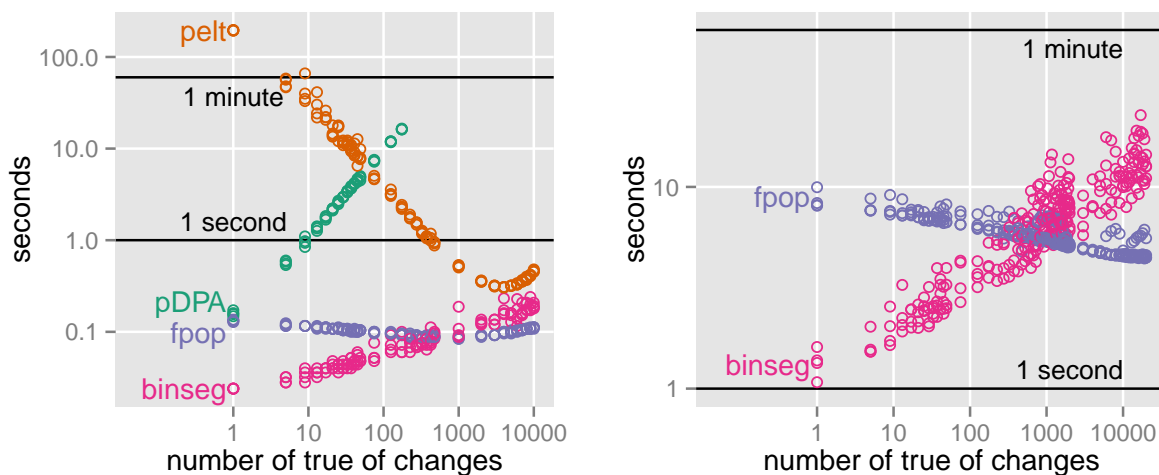


Figure 7: Runtimes as a function of the true number of changepoints. (Left) For FPOP, PELT, pDPA and Binseg and $n = 2 \times 10^5$ (Right) For Binseg and FPOP and $n = 10^7$

For PELT the relationship is less clear, however we expect pruning to be more efficient if there is a large number of changepoints. Hence for a fixed size signal we expect the runtime of PELT to improve with the underlying number of changes.

Based on Section 6 we expect FPOP to be more efficient than PELT and pDPA. Thus it seems reasonable to expect FPOP to be efficient for the whole range of K . This is what we empirically check in this section.

To do that we simulated Gaussian signal with 200000 data points and varied the number of changes and timed the algorithms for each of these signals. We then repeat the same experience for signals with 10^7 and timed FPOP and Binseg only. The R source code for these timings is in `benchmark/systemtime.simulation.R` in the `opfp` project repository on R-Forge: <https://r-forge.r-project.org/projects/opfp/>.

It can be seen in Figure 7 that FPOP is always faster than pDPA and PELT. Interestingly for both $n = 2 \times 10^5$ and 10^7 it is faster than Binseg for a true number of changepoints larger than 500.

7.3 Accuracy benchmark: the neuroblastoma data set

Hocking et al. (2013) proposed to benchmark the changepoint detection accuracy of segmentation models by using annotated regions defined by experts when they visually inspected scatterplots of the data. The neuroblastoma data are a set of 575 copy number microarrays of neuroblastoma tumours, and each chromosome is a separate segmentation problem. The benchmark is to use a training set of n annotated chromosomes to learn a segmentation model, and then quantify the annotation error on a test set. Let d_1, \dots, d_n be the number of data points to segment on each chromosome in the training data set, and let $\mathbf{y}_1 \in \mathbb{R}^{d_1}, \dots, \mathbf{y}_n \in \mathbb{R}^{d_n}$ be the vectors of noisy data for each chromosome in the training set.

Both PELT and pDPA have been applied to this benchmark by first defining $\beta = \lambda d_i$ for all $i \in \{1, \dots, n\}$, and then choosing the constant λ that maximises agreement with the annotated regions in the training set.

Since FPOP computes the same segmentation as PELT, we obtain the same error rate for FPOP on the neuroblastoma benchmark. As shown on <http://cbio.enscm.fr/~thocking/neuroblastoma/accuracy.html>, FPOP (fpop) achieves 2.2% test error. This is the smallest error across all algorithms that have been tested.

8 Discussion

We have introduced two new algorithms for detecting changepoints, FPOP and SNIP. A natural question is which of these, and the existing algorithms, pDPA and PELT, should be used in which applications. There are two stages to answering this question. The first is whether to detect changepoints through solving the constrained or the penalised optimisation problem, and the second is whether to use functional or inequality based pruning.

The advantage of solving the constrained optimisation problem is that this gives optimal segmentations for a range of numbers of changepoints. The disadvantage is that solving it is slower than solving the penalised optimisation problem, particularly if there are many changepoints. In interactive situations where you wish to explore segmentations of the data, then solving the constrained problem is to be preferred (Hocking et al., 2014). However in non-interactive scenarios when the penalty parameter is known in advance, it will be faster to solve the penalised problem to recover the single segmentation of interest.

The decision as to which pruning method to use is purely one of computational efficiency. We have shown that functional pruning always prunes more than inequality based pruning, and empirically have seen that this difference can be large, particularly if there are few changepoints. However functional pruning can be applied less widely. Not only does it require a stronger condition on the cost functions, but currently its implementation has been restricted to detecting changes in a uni-variate parameter from a model in the exponential family. Even for situations where functional pruning can be applied, its computational overhead per non-pruned candidate is higher.

Our experience suggests that you should prefer functional pruning in the situations where it can be applied. For example FPOP was always faster than PELT for detecting a change in mean in the empirical studies we conducted, the difference in speed is particularly large in situations where there are few changepoints. Furthermore we observed FPOP's computational speed was robust to changes in the number of changepoints to be detected, and was even competitive with, and sometimes faster than, Binary Segmentation.

Acknowledgements: We thank Adam Letchford for helpful comments and discussions. This research was supported by EPSRC grant EP/K014463/1. Maidstone gratefully acknowledges funding from EPSRC via the STOR-i Centre for Doctoral Training. The authors would like to thank the Isaac Newton Institute for Mathematical Sciences,

Cambridge, for support and hospitality during the programme *Inference for Change-Point and Related Processes* where work on this paper was undertaken.

References

- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19:716–723.
- Auger, I. E. and Lawrence, C. E. (1989). Algorithms for the Optimal Identification of Segment Neighborhoods. *Bulletin of Mathematical Biology*, 51:39–54.
- Braun, J. V., Braun, R. K., and Muller, H. G. (2000). Multiple Change-point Fitting via Quasilikelihood, With Application to DNA Sequence Segmentation. *Biometrika*, 87:301–314.
- Cleynen, A., Koskas, M., and Rigaiil, G. (2012). A Generic Implementation of the Pruned Dynamic Programming Algorithm. *ArXiv e-prints*.
- Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006). Structural Break Estimation for Nonstationary Time Series Models. *Journal of the American Statistical Association*, 101:223–239.
- Fryzlewicz, P. (2012). Wild Binary Segmentation for Multiple Change-Point Detection. *Annals of Statistics*, to appear.
- Hocking, T. D., Boeva, V., Rigaiil, G., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., Richer, W., Bourdeaut, F., Suguro, M., Seto, M., Bach, F., and Vert, J.-P. (2014). SegAnnDB: Interactive Web-Based Genomic Segmentation. *Bioinformatics*, 30:1539–46.
- Hocking, T. D., Schleiermacher, G., Janoueix-lerosey, I., Boeva, V., Cappo, J., Delattre, O., Bach, F., and Vert, J.-P. (2013). Learning Smoothing Models of Copy Number Profiles Using Breakpoint Annotations. *BNC Bioinformatics*, 14.
- Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumoussis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., and Tsai, T. T. (2005). An Algorithm for Optimal Partitioning of Data on an Interval. *IEE Signal Processing Letters*, 12:105–108.
- Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal Detection of Change-points With a Linear Computational Cost. *Journal of the American Statistical Association*, 107:1590–1598.
- Lavielle, M. (2005). Using Penalized Contrasts for the Change-Point Problem. *Signal Processing*, 85:1501–1510.
- 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*, 5:557–572.
- Picard, F., Lebarbier, E., Hoebeke, M., Rigaiil, G., Thiam, B., and Robin, S. (2011). Joint Segmentation, Calling, and Normalization of Multiple CGH Profiles. *Biostatistics*, 12:413–428.
- Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. Q. (2007). A Review and Comparison of Change-point Detection Techniques for Climate Data. *Journal of Applied Meteorology and Climatology*, 46:900–915.
- Rigaiil, G. (2010). Pruned Dynamic Programming for Optimal Multiple Change-Point Detection. *ArXiv e-prints*.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6:461–464.

Scott, A. J. and Knott, M. (1974). A Cluster Analysis Method for Grouping Means in the Analysis of Variance. *Biometrics*, 30:507–512.

Zhang, N. R. and Siegmund, D. O. (2007). A Modified Bayes Information Criterion With Applications to the Analysis of Comparative Genomic Hybridization Data. *Biometrics*, 63:22–32.