Change Point Estimation of Bar Code Signals

Leming Qu* and Yi-Cheng Tu†

*Department of Mathematics, Boise State University, Boise ID 83725-1555, USA
qu@math.boisestate.edu.

†Department of Computer Sciences, Purdue University, West Lafayette, IN 47907-2066, USA
tuyc@cs.purdue.edu

Problem Definition

$f(t)$ representing bar code is a 0-1 step function for $t \in T = [0,1]$.

Want to recover $f(t)$ given the samples $y_i = y(t_i), i = 1, \ldots, M$, of the continuous-time observation

$$y(t) = \alpha \cdot G \star f(t) + \epsilon(t), \quad t \in T = [0,1]$$

where $\alpha > 0$ is the unknown amplitude, the $\epsilon(t)$ is the additive unobservable noise process

$$G \star f(t) = \int_T G(t-x) f(x) dx,$$

$$G(t) = \frac{1}{\sigma_0 \sqrt{2\pi}} exp(-\frac{t^2}{2\sigma_0^2}),$$

$\sigma_0 > 0$: the unknown standard deviation which increases as the scanner moves away from the bar code.

Since $\sigma_0 > 0$ unknown, this problem differs slightly from standard restoration problems of image processing in that the convolution kernel contains unknown quantities.

The problem closer to the blind deconvolution problems.

Previous work

- local approach: finding local minima and maxima in the derivative of

$$s(t) = \alpha \cdot G \star f(t),$$

- global approach: regularization methods for ill-posed inverse problems such as total variation based restoration.

Shortcomings of these approach:

- Local approach:
  (1) Locating local extrema of $s'(t)$ is sensitive to noise.
  (2) these local extrema are difficult to relate to the true change points of $f(t)$ due to 'convolution distortion'.

- Regularization:
  How to choose the regularization parameter?
  Computational extensive.

Our Approach

A hybrid of local and global approach.

Fully utilize the information about
$f(t)$: a 0-1 step function.

A nonlinear least squares solution to the change points of $f(t)$, α and σ_0 with the constraints of the ordered change points.

The local approach is used to provide the starting values for the global minimization problem.

Change Point Estimation

Assuming the total number of bars of $f(t)$ is the known integer K.

ξ_{2j-1} and ξ_{2j}: the beginning and ending location of the jth bar for $j = 1, \ldots, K$ of the bar code function $f(t)$.

$$f(t) = I(\xi_{2j-1} < t \leq \xi_{2j}),\ t \in T,\ j = 1, \ldots, K.$$

where

$$0 < \xi_1 < \xi_2 < \ldots < \xi_{2K-1} < \xi_{2K} < 1$$

are the ordered change points.

The goal of bar code reconstruction:
to recover the change points
$\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_{2K-1}, \xi_{2K})^T$
from the observed data $\boldsymbol{y} = (y_1, \ldots, y_M)^T$ at $\boldsymbol{t} = (t_1, \ldots, t_M)^T$, without any knowledge of α and σ_0.

$$s(t) = \alpha \int_T G(t-x) f(x) dx = \alpha \sum_{j=1}^{K} \int_{\xi_{2j-1}}^{\xi_{2j}} G(t-x) dx.$$

Let

$$r_i = s_i - y_i = \alpha \sum_{j=1}^{K} \int_{\xi_{2j-1}}^{\xi_{2j}} G(t-x) dx - y_i$$

be the ith residual, $\boldsymbol{r} = (r_1, \ldots, r_M)^T$ the residual vector and

$$h(\boldsymbol{\xi}, \alpha, \sigma_0) = \frac{1}{2} \sum_{i=1}^{M} r_i^2 = \frac{1}{2} r^T r$$

the residual sums of squares. We seek the least squares solution of $\boldsymbol{\xi}, \alpha$ and σ_0. That is to find $\widehat{\boldsymbol{\xi}}, \widehat{\alpha}$ and $\widehat{\sigma}_0$ which minimizes the merit function $h(\boldsymbol{\xi}, \alpha, \sigma_0)$ subject to the required conditions.

More explicitly, the constrained nonlinear least squares problem is

$$min_{\boldsymbol{\xi},\alpha,\sigma_0} h(\boldsymbol{\xi},\alpha,\sigma_0) \qquad (1)$$

such that

$$0 < \xi_1 < \xi_2 < \ldots < \xi_{2K-1} < \xi_{2K} < 1,$$

$$\sigma_0 \geq 0, \quad \alpha > 0.$$

These constraints are simply linear inequality constraints
$A[\boldsymbol{\xi}^T,\sigma_0,\alpha]^T < u$
with a sparse matrix A.

The recast of the bar code reconstruction into a constrained nonlinear least squares problem enables us to utilize the existing techniques for solving nonlinear least square problem subject to linear inequality constraints in the statistical and numerical analysis literature.

The Fletcher-Xu hybrid Gauss-Newton and BFGS method for nonlinear least squares problem are super linearly convergent. This method along with other five methods for constrained nonlinear least squares problems is implemented in the $clsSolve$ solver of the TOMLAB 4.7 optimization environment.

The Jacobian matrix of $\boldsymbol{r}$ is easily obtainable analytically.

Initial Estimation

The success of the $(2K+2)$ dimensional global minimization problem heavily depends on good starting values. Our numerical experiments indicated that simple starting values such as equally spaced grids on T for $\boldsymbol{\xi}$ did not give satisfactory solutions.

The local extremas of the $s'(t)$ are close to $\boldsymbol{\xi}$

The initial estimation of $\boldsymbol{\xi}$ is the following problem: given the noisy discrete observations of $s(t)$:

$$y_i = s(t_i) + \epsilon_i, \quad i = 1, \ldots, M,$$

finding the local extremas of $s'(t)$.

The above is the Nonparametric regression problem.

Estimating $s(t_i)$ first by $\widehat{s}(t_i)$, then using $\widehat{s}'(t)$ to estimate $s'(t)$.

In our simulation, we use wavelet thresholding method to estimate $s(t_i)$ first, then estimate $s'(t_i)$ based on $\widehat{s}(t_i)$ using a first derivative filter.

The initial σ_0 is estimated by propositions suggested by Joseph and Pavlidis.

The initial value of α is simply the ordinary least squares estimate given the initial values of $\boldsymbol{\xi}$ and σ_0 .

Numerical Experiments

Estimation of $s(t_i)$ is carried out by the soft Wavelet thresholding technique implemented in the Wavelet Toolbox in MATLAB.

The thresholds are chosen by a heuristic variant of the Stein's Unbiased Risk Estimate with multiplicative threshold rescaling using a single estimation of level noise based on the finest level wavelet coefficients.

The wavelet filter used is db6: the Daubechies wavelet with 6 vanishing moments.

The first derivative filter for estimating $s'(t_i)$ from $\widehat{s}(t_i)$ is used.

Conclusion

A nonlinear least squares estimation for change points of bar code signals is proposed. The local information contained in the derivative of the convolved signal is used to provide starting values for the global optimization solution. This hybrid approach uses all available information for parameter estimation to the full extent. If extra information such as the knowledge of the width of the thinnest or thickest black and white strips is available, they can be easily incorporated into the linear inequality constraints.

Currently, the value K of number of bars is assumed to be known in advance. A future research effort is to estimate the bar code without this assumption. Then model selection methods are needed for this situation.