11-1-2005

Change Point Estimation of Bilevel Functions

Leming Qu
Boise State University, qu@math.boisestate.edu

Yi-Cheng Tu
Purdue University, tuyc@cs.purdue.edu

Change Point Estimation of Bilevel Functions

Leming Qu
Department of Mathematics
Boise State University

Yi-Cheng Tu
Department of Computer Science
Purdue University

Reconstruction of a bilevel function such as a bar code signal in a partially blind deconvolution problem is an important task in industrial processes. Existing methods are based on either the local approach or the regularization approach with a total variation penalty. This article reformulated the problem explicitly in terms of change points of the 0-1 step function. The bilevel function is then reconstructed by solving the nonlinear least squares problem subject to linear inequality constraints, with starting values provided by the local extremas of the derivative of the convolved signal from discrete noisy data. Simulation results show a considerable improvement of the quality of the bilevel function using the proposed hybrid approach over the local approach. The hybrid approach extends the workable range of the standard deviation of the Gaussian kernel significantly.

Key words: Bar code, 0-1 step function, nonlinear least squares, constrained optimization

Introduction

Reconstruction of bilevel functions has important practical applications in modern life. A particular example of bilevel functions is the bar code signal. The ubiquitous alternating black and white strips (the bar code) are now widely used in every day life and industrial processes.

The problem of recovering a bar code signal $f(t)$ from the noisy signal $y(t)$ detected by a bar code scanner (Esedoglu, 2004, Shellhammer, Goren, & Pavlidis, 1999) is to construct a one-dimensional 0-1 step function $f(t)$ given the samples $y_i = y(t_i)$, $i = 1, ..., M$, of the continuous-time observation

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

Leming Qu is an Assistant Professor of Statistics. His research interests include Wavelets in statistics, time series, Bayesian Analysis, statistical and computational inverse problems, nonparametric and semiparametric regression. Email him at qu@math.boisestate.edu. Yi-Cheng Tu is a doctoral student in Computer Science. Email at tuyc@cs.purdue.edu.

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

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

and $G(t)$ is a Gaussian kernel of the convolution:

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

where $\sigma > 0$ is the unknown standard deviation which increases as the scanner moves away from the bar code. According to Wittman, (2004), the $\sqrt{2}\sigma$ is assumed to be smaller than the width of the thinnest (WTB) bar. The approach proposed in this article can extend this range. The simulation result in section V shows that $f(t)$ can still be reconstructed even $\sqrt{2}\sigma$ is as large as 1.4 times the WTB at high signal to noise ratio level. Figure 1 illustrates the results of the convolution with additive noise for a UPC bar code (Wittman, 2004) encoding 0123456789.

Without using the special characteristics of a bar code signal, $f(t)$ in model (1) is treated as a bilevel function with levels 0 or 1. Thus, the goal is to reconstruct the bilevel function f(t)

from the noisy discrete data $\boldsymbol{y} = (y_1, \ldots, y_M)^T$. This problem differs slightly from standard restoration problems of image processing in that the convolution kernel contains unknown quantities. It is a partially blind deconvolution problem somewhat closer to blind deconvolution.

Previous Work and Current Approach

       Previous work on the bar code reconstruction problem (Wittman, 2004) is based on the following:

(a) local approach: finding local minima and maxima in the derivative of

$$s(t) = \alpha \cdot G * f(t). \qquad (2)$$

      Note the above is the noise-free version of model (1);

(b) global approach: regularization methods for ill-posed inverse problems such as total variation based restoration (Esedoglu, 2004).

      The approach (a) tries to relate the local minima and maxima in $s'(t)$ to the edges of bars which are the change points in $f(t)$. Locating these local extremas is sensitive to noise ε (t). Furthermore, these local extrema are difficult to relate to the true change points of $f(t)$ due to convolution distortion (Shellhammer, Goren, & Pavlidis, 1999). Such techniques use only local information and would have difficulty to detect thin bars closely adjacent to each other. For example, in Figure 1, the $s(t)$ is near flat around location 0.55 even though three adjacent thin bars stand there.

      To overcome these shortcomings, approach (b) in Esedoglu (2004) tried to recover $f(t)$ by regularization using the total variation penalty, a technique commonly used in image restoration literature. It models systematically the interaction of neighboring bars in $f(t)$ under convolution with $G(t)$, as well as the estimation of α and σ from global information contained in the $y(t)$. It is proved in Esedoglu (2004) that under certain regularity conditions, the infimum of the total-variation energy is attained.

Numerical results show that bar codes from highly degraded signals can be recovered.

      The regularization approach in inverse problems must deal with the choice of the regularization parameter, a difficult problem itself. In Esedoglu (2004), there are two regularization parameters which need to be chosen. In the numerical results of Esedoglu (2004), the regularization parameters are preselected and kept fixed.

      It is felt that all the existing works did not fully utilize the information about $f(t)$: a bilevel function with levels 0 or 1. To recover $f(t)$ is to recover the change points of $f(t)$ for $t \in$ T. The number of change points in $f(t)$ is twice the number of bars in the bar code. Recovering the $f(t)$, $t \in T$ is usually an ill-posed problem, although recovering the change points is a well-posed problem if the number of observations exceed the number of unknown parameters. The well-posed problem is essentially to estimate the nonlinear parameters in finite dimension with linear inequality constraints. The formulation of the bilevel function deconvolution was not found in terms of the change points explicitly in the existing literature. Therefore, a nonlinear least squares solution is proposed to the change points of $f(t)$, σ and α with the constraints of the ordered change points. The local approach is used to provide the starting values for the global minimization problem. This method is a hybrid of local and global approaches in spirit.

Change Point Estimation

      Assume the total number of bars of $f(t)$ is the known integer K. For example, $K = 22$ for the UPC bar code in the test problem. Denote ξ_{2j-1} and ξ_{2j} as the beginning and ending location of the jth bar for $j = 1, \ldots, K$ of the bilevel function $f(t)$.

      Then $f(t)$ can be defined explicitly as:

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

where $I()$ is the usual indicator function and

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

are the ordered change points.

The goal of the bilevel function reconstruction is to recover the change points ξ = $(\xi_1, \xi_2, \dots, \xi_{2K-1}, \xi_{2K})^T$ from the observed data y = $(y_1, \dots, y_M)^T$ at $t = (t_1, \dots, t_M)^T$, without any knowledge of σ and α.

With the special structure of $f(t)$, the convolution $G * f(t)$ can be explicitly expressed as a function of ξ and σ. Thus,

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

Let

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

be the ith residual. Denote $r = (r_1, \dots, r_M)^T$ the residual vector and

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

the residual sums of squares. The least squares solution of ξ, σ and α is sought. That is to find $\hat{\xi}$, $\hat{\sigma}$, and $\hat{\alpha}$ which minimizes the merit function $h(\xi, \sigma, \alpha)$ subject to the required conditions.

More explicitly, the constrained nonlinear least squares problem is

$$min_{\xi, \sigma, \alpha} \, h(\xi, \sigma, \alpha) \qquad (3)$$

such that

$$0 < \xi_1 < \xi_2 < \cdots < \xi_{2K-1} < \xi_{2K} < 1, \sigma > 0, \alpha > 0.$$

These constraints are simply linear inequality constraints $A[\xi^T, \sigma, \alpha]^T < u$ with a sparse matrix A whose elements are

$$
\begin{aligned}
A[1,1] &= -1, \\
A[i, i-1] &= 1, \\
A[i,i] &= 1 \quad \textbf{for } i = 2,\dots,2K, \\
A[2K+1, 2K] &= 1, \\
A[2K+2, 2K+1] &= -1, \\
A(2K+3, 2K+2] &= -1;
\end{aligned}
$$

and $u = (0, \dots, 0, 1, 0, 0)^T$ is a $(2K+3)$ column vector.

The recast of the bilevel function 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 (Fletcher & Xu, 1987) for nonlinear least squares problem is super linearly convergent. This method along with other five methods for constrained nonlinear least squares problems is implemented in the *cls Solve* solver of the TOMLAB 4.7 optimization environment [9].

The Gauss-Newton method needs the gradient of the merit function $h(\xi, \sigma, \alpha)$, which is the product of the Jacobian matrix of r and r. The Jacobian matrix of r is easily obtained by:

$$
\begin{aligned}
\frac{\partial r_i}{\partial \xi_{2j-1}} &= -\alpha \cdot G(t_i - \xi_{2j-1}), \\
\frac{\partial r_i}{\partial \xi_{2j}} &= \alpha \cdot G(t_i - \xi_{2j}), \\
\frac{\partial r_i}{\partial \sigma} &= \frac{\alpha}{\sigma} \sum_{j=1}^{K} \int_{\xi_{2j-1}}^{\xi_{2j}} \left[-1 + \frac{(t_i - x)^2}{\sigma^2} \right] G(t_i - x) dx, \\
\frac{\partial r_i}{\partial \alpha} &= \sum_{j=1}^{K} \int_{\xi_{2j-1}}^{\xi_{2j}} G(t_i - x) dx.
\end{aligned}
$$

The success of the $(2K+2)$ dimensional global minimization problem (3) heavily depends on good starting values. Our numerical experiments indicated that simple starting values such as equally spaced grids on T for ξ did not give satisfactory solutions. Next, the initial

parameter estimation based on the local approach is discussed.

Initial Estimation

The local extremas of the derivative of $s(t)$ are close to ξ, the edges of the bars. Then the initial estimation of ξ in terms of the local extremas of the derivative signal is the following problem: given the noisy discrete observations of $s(t)$:

$$y_i = s(t_i) + \varepsilon_i, i = 1,...,M,$$

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

There are approaches of estimating $s'(t)$ based on different smoothing or denoising methods. Many of these try first to find $\hat{s}(t)$, the estimate of $s(t)$, using the chosen smoothing or denoising method; then estimate $s'(t)$ based on $\hat{s}'(t)$. See, for example, the spline regression based method in Zhou and Wolfe (2000) or the wavelet denoising based method in Cai (2002). For equally spaced $\{t_i\}$ and when M is a power of 2, there exists a fast algorithm with complexity $O(M)$ to carry out discrete wavelet transform (DWT). In this simulation, the wavelet thresholding method is used to estimate $s(t_i)$ first, then estimate $s'(t_i)$ based on $\hat{s}(t_i)$ using a first derivative filter.

After obtaining the initial estimate ξ_0 of ξ by the K pairs of local maxima and minima of $\hat{s}'(t_i)$, σ_0 is estimated, the initial σ, by techniques suggested in Joseph and Pavlidis (1993), Joseph and Pavlidis (1994), and Esedoglu (2004). Proposition 1 of Joseph and Pavlidis (1993) suggested approximating σ by the distance from the last local maxima of $\hat{s}(t_i)$ to the last local minima of $\hat{s}'(t_i)$. Proposition 2 of Joseph and Pavlidis (1994) suggested approximating σ by $\left|\hat{s}'(t^*)/\hat{s}^{(3)}(t^*)\right|$ where t^* is a point such that $\left|\hat{s}'(t^*)\right| \geq \left|\hat{s}'(t_i)\right|$ for $i = 1, ..., M$. The smaller vale of $\hat{\sigma}$ based on the two propositions is used first. If it is outside the reasonable range [0.001, 0.02], then the value 0.0079 is used as suggested in Esedoglu (2004) for the true σ ranging from 0.011 to 0.013. In contrast, the thinnest bar has width 0.0132 in

this simulation study. The initial value of α, α_0, is simply the ordinary least squares estimate given the ξ_0 and σ_0.

Simulation Results

In the experiment, a clean bar code $f(t)$ encoding 0123456789 was blurred by convolving it with $G(t)$ of known σ, amplified by the amplitude $\alpha = 1$, sampled at $t_i = i/M$, for $i = 1, ..., M$, followed by the addition of white Gaussian noise $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. The amount of the added noise makes the signal-to-noise ratio $SNR = 20log_{10}(std(s)/\sigma_\varepsilon)$ at the specified level.

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 Stein's Unbiased Risk Estimate with multiplicative threshold rescaling using a single estimation of level noise based on the finest level wavelet coefficients (Donoho & Johnstone, 1995). The wavelet filter used is db6: the Daubechies wavelet with 6 vanishing moments.

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

$$d1 = [-0.015964, -0.121482, -0.193357,$$
$$0.00, 0.193357, 0.121482, 0.015964];$$

as recently suggested in Farid and Simoncelli (2004). This filter is significantly more accurate than those commonly used in the image and signal processing literature.

Table 1 shows the Monte-Carlo approximations to $MSE = E\left(\left\|\hat{\xi} - \xi\right\|_2^2 /(2K)\right)$ of this method based on 100 independent simulations as the σ is varied at 0.011, 0.012 and 0.0125, the sample size M is varied dynamically from $M = 256$ through 1024, and SNR is varied from high to moderate levels. The range of the σ is in contrast with the WTB equal to 0.0132.

Table 2 shows the Monte-Carlo approximations to $MSE = E(\|\xi_0 - \xi\|_2^2/(2K))$. The results show a considerable reduction of MSE for $\hat{\xi}$ over ξ_0 in most cases. The most significant improvement occurs for the three tested values of σ with high SNR.

Table 1. Monte Carlo Approximations To $MSE = E\left(\left\|\hat{\xi} - \xi\right\|_2^2 / (2K)\right)$

		Means for the following levels of SNR		
σ	M	$SNR = 38$	$SNR = 28$	$SNR = 21$
0.0110	256	3.402e-09	3.358e-08	1.006e-04
0.0110	512	1.565e-09	1.628e-08	5.159e-05
0.0110	1024	7.769e-10	7.703e-09	1.471e-05
0.0120	256	3.685e-09	9.822e-05	2.992e-04
0.0120	512	1.939e-09	7.846e-06	3.084e-04
0.0120	1024	9.744e-10	1.220e-05	2.450e-04
0.0125	256	6.153e-05	2.020e-04	4.457e-04
0.0125	512	2.063e-09	5.569e-05	4.502e-04
0.0125	1024	1.055e-09	6.317e-05	3.847e-04

Table 2. Monte Carlo Approximations To $MSE = E(\|\xi_0 - \xi\|_2^2 / (2K))$

		Means for the following levels of SNR		
σ	M	$SNR = 38$	$SNR = 28$	$SNR = 21$
0.0110	256	1.512e-05	1.657e-05	1.399e-04
0.0110	512	1.052e-05	1.128e-05	6.394e-05
0.0110	1024	9.366e-06	1.002e-05	3.060e-05
0.0120	256	2.050e-05	1.324e-04	3.486e-04
0.0120	512	1.277e-05	2.471e-05	3.496e-04
0.0120	1024	1.129e-05	2.670e-05	2.918e-04
0.0125	256	8.834e-05	2.485e-04	5.345e-04
0.0125	512	1.478e-05	8.177e-05	5.068e-04
0.0125	1024	1.264e-05	8.423e-05	4.555e-04

Fgures 1 and 2 present results from two of these experiments. Figure 1 is an example of completely successful reconstruction while the Figure 2 an example that the estimation fails when the noise level or the blur factor σ gets too high. It seems that when $\left\|\hat{\xi} - \xi\right\|_2^2 / 44$ is less than the scale of $1.0e$-5, a successful reconstruction is obtained.
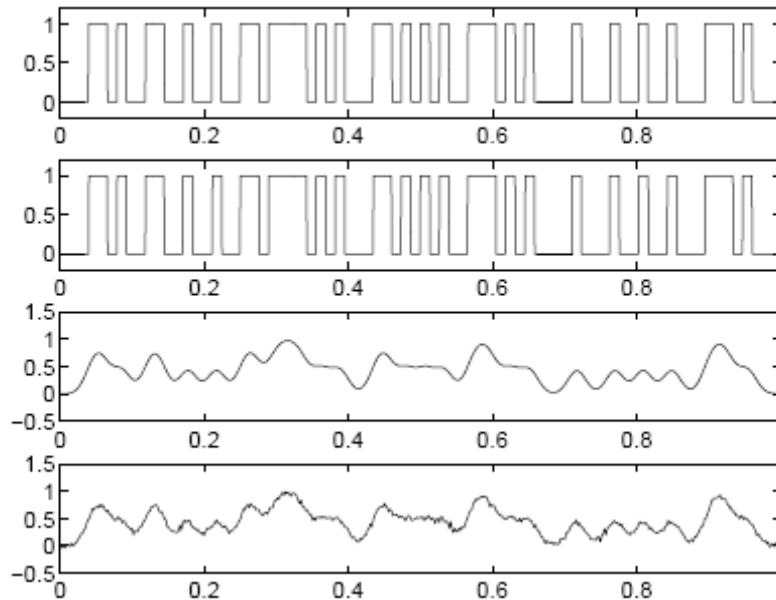
Figure 1. Top to bottom: the bar code, the reconstructed bar code, the corresponding clean convolved signal, the noisy convolved signal. The true parameter used to generate the corrupted signal: $M = 512$, $\sigma = 0.0118$, $\alpha = 1$, $SNR = 18$. The estimated parameter: $\hat{\sigma} = 0.0119$, $\hat{\alpha} = 1.0036$. The square error $\left\| \hat{\xi} - \xi \right\|_2^2 / 44 = 9.6e\text{-}8$. CPU time: 51 seconds.
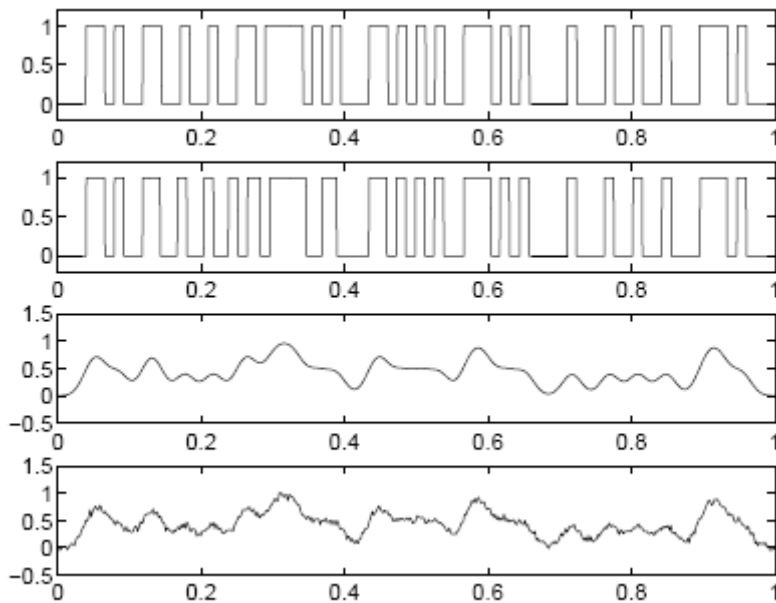

Figure 2. Top to bottom: the bar code, the failed reconstructed bar code, the corresponding clean convolved signal, the noisy convolved signal. The true parameter used to generate the corrupted signal: $M = 512$, $\sigma = 0.0129$, $\alpha = 1$, $SNR = 16$. The estimated parameter: $\hat{\sigma} = 0.0136$, $\hat{\alpha} = 1.0287$. The square error $\left\| \hat{\xi} - \xi \right\|_2^2 / 44 = 2.1e\text{-}4$. CPU time: 66 seconds.

Setting the cut-off $\left\|\hat{\xi} - \xi\right\|_2^2 / 44$ at $1.0e\text{-}5$ for a successful reconstruction, Table 3 displays the percentage of successful reconstructions of ξ. Table 4 displays the percentage of successful reconstructions of ξ_0. The successful rate of our hybrid approach is 100% in most cases with high *SNR*. The local approach alone will not work.

The simulation result also indicates that $\hat{\sigma}$ gives much better solution than the initial estimate σ_0 in terms of the reduction of *MSE*. The same is true for $\hat{\alpha}$.

The computational time for finding the solution is relatively fast. For example, for the most time consuming scenario $\sigma = 0.125$, $M = 1024$, *SNR* = 21, the average CPU time is 86 seconds, contrary to the reported 6 minutes using the regularization approach in Wittman (2004).

Most current bar code decoding works only in the range when σ is up to 0.7 times the WTB. The result in Table 3 indicates that this approach can attain 100% successful rate even when σ is up to 0.95 times the WTB at the *SNR* = 38 level. In contrast, the local approach alone can attain a successful rate of 94% when σ is up to 0.83 times the WTB at the *SNR* = 38 level as indicated in Table 4. This approach seems to break down at $\sigma = 0.0120$ and *SNR* = 21.

Table 3. Percentage Of Successful Reconstruction Of ξ

		Percentages for the following levels of SNR		
σ	M	$SNR = 38$	$SNR = 28$	$SNR = 21$
0.011000	256	100	100	85
0.011000	512	100	100	90
0.011000	1024	100	100	98
0.012000	256	100	81	54
0.012000	512	100	97	52
0.012000	1024	100	98	51
0.012500	256	88	63	40
0.012500	512	100	89	37
0.012500	1024	100	87	33

Table 4. Percentage Of Successful Reconstruction Of ξ_0

σ	M	Percentages for the following levels of SNR		
		$SNR = 38$	$SNR = 28$	$SNR = 21$
0.011000	256	0	0	0
0.011000	512	19	13	1
0.011000	1024	94	54	8
0.012000	256	0	0	0
0.012000	512	0	0	0
0.012000	1024	0	0	0
0.012500	256	0	0	0
0.012500	512	0	0	0
0.012500	1024	0	0	0

Conclusion

A nonlinear least squares estimation for change points of a bilevel function 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. Monte Carlo simulation results show the good performance of the hybrid approach over the local approach. It also indicates that the hybrid approach extends the workable range of the standard deviation of the Gaussian kernel significantly.

If extra information such as the WTB or the width of thickest bar or space is available, they can be easily incorporated into the linear inequality constraints. Actually, in UPC barcode system, the WTB is always known. A barcode is decoded based on the bar widths relative to the WTB. Each should be 1, 2, 3, or 4 times the WTB. This knowledge could lead to an interesting combinatorial problem for bar code deconvolution and an additional constraint: choosing the change points to reflect the fact that the widths are 1 through 4 times the WTB.

Currently, the value K of the number of bars is assumed to be known in advance. In real bar code, the range of possible K values will be fairly small. A future research effort is to estimate the bilevel function without this assumption. Then model selection methods are needed for such problem.

References

Cai, T. (2002). On adaptive wavelet estimation of a derivative and other related linear inverse problems. *Journal of Statistical Planning and Inference, 108*, 329-349

Donoho, D., & Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of American Statistical Association, 90*, 1200-1224

Esedoglu, S. (2004). Blind deconvolution of bar code signals. *Inverse Problems, 20*, 121-135

Farid, H. & Simoncelli, E. P. (2004). Differentiation of discrete multidimensional signals. *IEEE Transactions on Image Processing, 13*(4), 496-508.

Fletcher, R. & Xu, C. (1987). Hybrid methods for nonlinear least squares. *IMA Journal of Numerical Analysis, 7*, 371-389.

Joseph, E. & Pavlidis, T. (1993). Deblurring of bilevel waveforms. *IEEE Transactions on Image Processing, 2*, 223-235.

Joseph, E. & Pavlidis, T. (1994). Bar code waveform recognition using peak locations. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 16*, 630-640.

Shellhammer, S. J., Goren, D. P., & Pavlidis, T. (March, 1999). Novel signal-processing techniques in barcode scanning. *IEEE Robotics & Automation Magazine*, 57-65

Tomlab Optimization Inc., (2005). User's Guide For TOMLAB 4.7. http://tomlab.biz.

Wittman, T. (September, 2004). Lost in the supermarket: Decoding blurry barcodes. *SIAM News*.

Zhou, S. & Wolfe, D. A., (2000). On derivative estimation in spline regression. *Statistica Sinica, 10*, 93-108.