ON THE ESTIMATION OF THE BANDWIDTH OF NONUNIFORMLY SAMPLED SIGNALS

Thomas Strohmer

Department of Mathematics, University of California, Davis
Davis, CA-95616, strohmer@math.ucdavis.edu

ABSTRACT

In many applications signals can only be sampled at nonuniformly spaced points. For a reliable reconstruction of the signal from its samples we require knowledge of the bandwidth of the signal, which however is often not known a priori. Therefore robust and efficient methods are needed that allow to estimate the bandwidth of a signal from nonuniform spaced, noisy samples. We present two procedures for bandwidth estimation. The first method is based on the discrete Bernstein inequality and Newton's divided differences and is computationally very efficient. The second method requires somewhat more computational effort, since it simultaneously estimates the bandwidth and provides a reconstruction of the signal. It is based on a multi-scale conjugate gradient algorithm for the solution of a nested sequence of Toeplitz systems and is particularly useful in case of noisy data. Examples from various applications demonstrate the performance of the proposed methods.

1. INTRODUCTION

In many practical applications, for instance in geophysics, spectroscopy, medical imaging, the signal cannot be sampled at regularly spaced points. Thus one is confronted with the problem of reconstructing an irregularly sampled signal. Often the signal can be assumed to be band-limited (or at least essentially band-limited), but the actual bandwidth is not known.

The problem of reconstructing a band-limited signal f from regularly and irregularly samples $\{f(t_j)\}_{j\in I}$ has attracted many mathematicians and engineers. Many theoretical results and efficient algorithms have been derived in the last years, see e.g. [1, 2, 3, 4] and the references cited therein. Most of these results are based on the assumption that the bandwidth of the signal to be recovered is known a priori.

However in many applications this assumption is not justified. Only a few theoretical results are known for bandwidth estimation from irregularly spaced data [5]. The importance of choosing a proper bandwidth in order to avoid underfitting or overfitting of noisy data is illustrated by means of a simulated numerical example in Figure 1.

In this note we present two procedures for bandwidth estimation. One method involves a discrete version of the Bernstein inequality, see Section 3. The other method, discussed in Section 2, simultaneously estimates the bandwidth and provides a reconstruction of the signal. It is based on a multi-scale conjugate gradient algorithm for the solution of a nested sequence of Toeplitz systems and is particularly useful in case of noisy data. Examples from various applications demonstrate the performance of the proposed methods.

Before we proceed we introduce some notation. Let $\mathbb{B}_N$ be the space of bandlimited functions

$$\mathbb{B}_N = \{f \in \boldsymbol{L}^2(\mathbb{R}) : \hat{f}(\omega) = 0 \,\text{for}\, |\omega| > N\}$$

where $\hat{f}$ denotes the Fourier transform of f. Let $\{t_j\}_{j\in\mathbb{Z}}$ be a set of sampling points and denote $y = \{f(t_j\}_{j\in\mathbb{Z}}$. We define the analysis operator T_N which maps signals in $\mathbb{B}_N$ to sequences in $\boldsymbol{\ell}^2(\mathbb{Z})$ via

$$T_N f = \{\langle f, \text{sinc}_N(\cdot - t_j)\rangle\}_{j\in\mathbb{Z}} = \{(f * \text{sinc}_N)(t_j)\}_{j\in\mathbb{Z}}$$

where $f * g$ denotes convolution. The adjoint operator is $T_N^* : \boldsymbol{\ell}^2(\mathbb{Z}) \mapsto \mathbb{B}_N$ given by

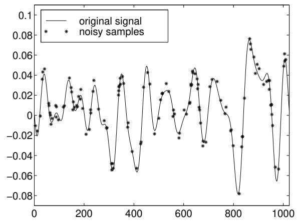
$$T_N^*\{c_j\}_{j\in\mathbb{Z}} = \sum_{j\in\mathbb{Z}} c_j \text{sinc}_N(\cdot - t_j)\,.$$

For $f \in \mathbb{B}_N$ we have $T_N f = \{f(t_j)\}_{j\in\mathbb{Z}}$ since in this case T corresponds to a reproducing kernel on $\mathbb{B}_N$. It is known that if the set $\{\text{sinc}_N(\cdot - t_j)\}$ is a frame for $\mathbb{B}_N$ then the operator equation

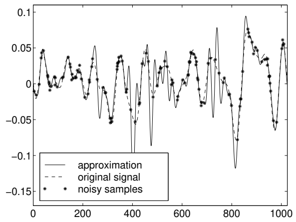
$$T_N f = y \tag{1}$$

has a unique solution, in other words f can be reconstructed from its samples $f(t_j)$.

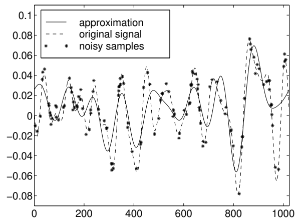
One efficient method for the solution of $T_N f = y$ is the conjugate gradient method applied to the normal equations $T_N^* T_N f = T_N^* y$. This method is particularly useful for the irregular sampling problem [4]. However to form the system $T_N f = y$ we already assume knowledge of the bandwidth of f, which may not be known in practice. Thus we have to find a way to estimate the bandwidth of f from its (noisy) samples $f(t_j)$ in order to reconstruct f.



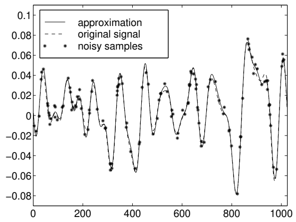
(a) Bandlimited signal and noisy samples



(b) Reconstruction using a too large bandwidth (overfit)



(c) Reconstruction using a too small bandwidth (underfit)



(d) Reconstruction using proposed algorithm

Figure 1: A good estimate of the bandwidth of a nonuniformly sampled signal is essential for signal reconstruction from noisy data in order to avoid overfitting and underfitting of the data. The proposed algorithm automatically adapts to solution of optimal bandwidth.

2. A MULTI-LEVEL CG METHOD FOR BANDWIDTH ESTIMATION

2.1. Theoretical considerations

Assume that we are given the data $y = \{y_j\}$ with $y_j = f(t_j) + \delta_j$ and $\sum_j |y_j - f(t_j)|^2 \leq \delta^2$ where $f \in \mathbb{B}_M$ for some unknown M. We propose following multi-level approach for combined bandwidth estimation and signal reconstruction.

Start at bandwidth or *level* $N = 0$ and apply the CG method to $T_0 f^{(0)} = y$ where the superscript in $f^{(0)}$ indicates the dependence of the solution on the chosen bandwidth $N = 0$. We run the CG method until a certain level-dependent *inner stopping criterion* is fulfilled, say, after k iterations, providing an approximation $f_k^{(0)}$. In other words we check if $\|T_0 f^{(0)} - y\|^2$ is smaller than a certain bound (which will be specified below). If this level-dependent *outer stopping criterion* is satisfied, we accept $f_k^{(0)}$ as approximate solution, otherwise we switch to the next level $N = 1$, where we use $f_k^{(0)}$ as initial guess $f_0^{(1)}$ for the CG method to solve $T_1 f^{(1)} = y$. We proceed for increasing levels N until the method terminates at a level $N \leq M$. The mathematical difficulty of this approach is to find well-defined inner and outer stopping criteria to guarantee that:

- the multi-level CG method converges to f for $\delta \to 0$.
- we do not iterate too long at a certain level N, because the actual solution may belong to a "higher-level" space $\mathbb{B}_M$ with $M > N$.
- we do not stop too early at a certain level N because the solution may actually belong to $\mathbb{B}_N$.

Hanke [6] showed by a counterexample that for noisy data the error of the iterates of the CG method does not monotonically decrease when the iteration is terminated with the usual stopping rules. Therefore an important issue is to find a stopping criterion, which guarantees monotonicity of the iterates at each level.

Based on results obtained in a more general framework of moment problems [7] we terminate the CG method at level N and switch to level $N+1$ if following (level-dependent) inner stopping criterion is satisfied for the first time

$$\sum_j |f_k^{N)}(t_j) - y_j|^2 \leq 2\eta(\delta + \|\mathrm{sinc}_N * f - f\|)\|y\| \quad (2)$$

where $\eta > 1$. The multi-level algorithm is terminated (i.e., we stop the outer iterations) if following (level-independent) outer stopping criterion holds

$$\sum_j |f_k^{N)}(t_j) - y_j|^2 \leq 2\eta\delta\|y\| \,. \quad (3)$$

Clearly we cannot compute $\|\mathrm{sinc}_N * f - f\|$, since f is not known. In [7] the reader can find a procedure that describes how $\|\mathrm{sinc}_N * f - f\|$ can be estimated recursively.

2.2. Numerical implementation

For the numerical implementation of this method we use the finite-dimensional model outlined in [4]. We assume that we are given the samples $s(t_j), j = 0, \dots r-1$, where s is a periodic signal of period L, i.e., $s(j) = s(j + kL)$. The space of bandlimited functions is now given by

$$\mathbb{B}_N = \{s \in \boldsymbol{\ell}^2(\mathbb{Z}_L) : \hat{s}(k) = 0 \,\text{for}\, |k| > N\}$$

where $\hat{s}$ denotes the discrete Fourier transform of s. Using this model one can show that solving the normal equations $T_N^* T_N f^{(N)} = T_N^*\{f(t_j)\}$ is equivalent to solving a certain Toeplitz system $A_N x^{(N)} = b^{(N)}$, where $x^{(N)}$ contains the non-zero Fourier coefficients of $f^{(N)}$, cf. [4].

Let A_N be the Toeplitz matrix at level N, let A_{N+1} denote the Toeplitz matrix at level $N+1$ and let $\{a_k\}_{k=0}^{2(N+1)}$ be the first column of A_{N+1}. An important observation from a numerical point of view is that A_N is embedded in A_{N+1} and as follows

$$A_{N+1} = \begin{bmatrix} a_0 & \cdots & \overline{a_{2(N+1)}} \\ \vdots & \boxed{A_N} & \vdots \\ a_{2(N+1)} & \cdots & a_0 \end{bmatrix}.$$

An analogous result holds for the right hand side $b^{(N)}$, see also [8]. This observation reduces the computational effort of the multi-level CG method considerably, since only two new entries have to be calculated to establish the system matrix at the next-higher level.

Remark: The proposed method can be easily generalized to higher dimensions.

3. BANDWIDTH ESTIMATION BASED ON A DISCRETE BERNSTEIN INEQUALITY

The estimation method described in this section is due to K. Gröchenig, G. Zimmermann and the author. The key observation is the fact that for $f \in \mathbb{B}_N$ it follows from a discrete version of *Bernstein's inequality* [9]. Let Δs denote the difference sequence $\Delta s(n) = s(n+1) - s(n)$ (and by periodicity $\Delta s(L-1) = s(0) - s(L-1)$. Then for all $s \in \mathbb{B}_N$

$$\|\Delta s\| \leq 2 \sin \frac{\pi N}{L} \|s\| \tag{4}$$

hence we obtain following lower bound for regularly sampled signals

$$N \geq \frac{L}{\pi \|s\|} \arcsin \frac{\|\Delta s\|}{2}. \tag{5}$$

We approximate the difference operator using (higher-order) divided differences applied to the given nonuniformly spaced samples. A more comprehensive discussion of this method will be given elsewhere. This method is useful to obtain an initial guess for the bandwidth for the multi-level CG method.

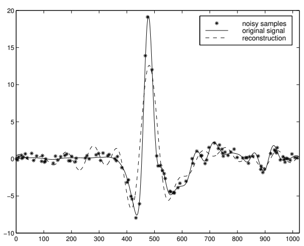
4. NUMERICAL EXPERIMENTS

We present examples from spectroscopy and medical imaging to demonstrate the performance of the method presented in Section 2.

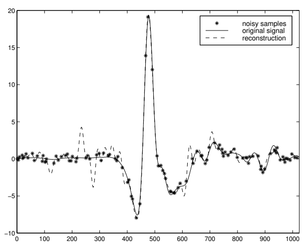
Experiment 1: Signal reconstruction in spectroscopy

In the first example we consider a signal from spectroscopy. The original signal of bandwidth $M = 30$ consists of 1024 regularly spaced and nearly noise-free samples. We have added zero-mean white noise with a signal-to-noise ratio of 12% and have sampled this noisy signal at 107 nonuniformly spaced sampling points. The setup of this experiment simulates a typical situation in spectroscopy. Since the samples are contaminated by noise, we cannot expect to recover the signal completely.

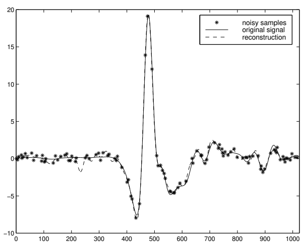
Figure 2 illustrates the resulting approximations by applying the ACT algorithm proposed in [4] using (a) a too small bandwidth (b) a too large bandwidth (c) the original bandwidth (d) and applying the proposed multi-level algorithm which in this example terminates at bandwidth 23. Although the bandwidth of the approximation is smaller than that of the original signal, the SNR of this approximation is better than the SNR using the correct bandwidth 30. The reason is that the stopping criteria also serve as *regularization parameter* that force the algorithm to terminate before the noisy in the data severely affects the reconstruction procedure.



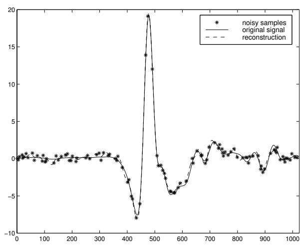
(a) Using a too small bandwidth provides a too smooth reconstruction, SNR = 6.479.



(b) Using a too large bandwidth provides a highly oscillating solution, SNR = 11.913.



(c) Even using the actual bandwidth may result in overfitting the data due to the noise, SNR = 18.804.



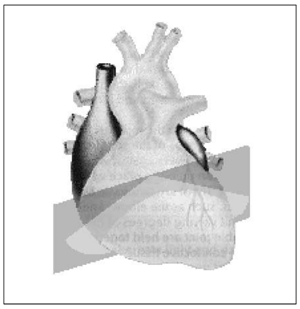
(d) Regularized solution using proposed method, SNR = 21.019.

Figure 2: Optimal control of the bandwidth of the solution is essential for signal reconstruction from noisy nonuniformly spaced samples in order to avoid overfitting and underfitting of the data. The proposed method automatically adapts to the solution of optimal bandwidth.

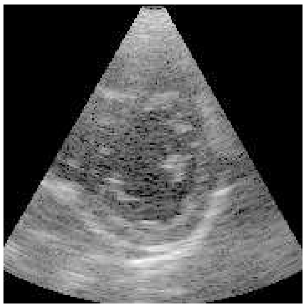
Experiment 2: Contour recovery in Echocardiography

In clinical cardiac studies the evaluation of cardiac function using parameters of left ventricular contractibility is an important constituent of an echocardiographic examination. These parameters are derived using boundary tracing of endocardial borders of the Left Ventricle (LV). The extraction of the boundary of the LV comprises two steps, once the ultrasound image of a cross section of the LV is given, see Figure 3(a)–(d). First an edge detection is applied to the ultrasound image to detect the boundary of the LV, cf. Figure 3(c).

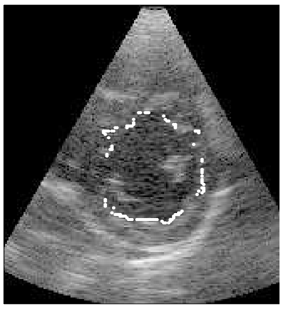
However this procedure may be hampered by the presence of interfering biological structures (such as papillar muscles), the unevenness of boundary contrast, and various kinds of noise [10]. Thus edge detection often provides only a set of nonuniformly spaced, perturbed boundary points rather than a connected boundary. Therefore a second step is required, to recover the original boundary from the detected edge points, cf. Figure 3(d). Since the shape of the Left Ventricle is definitely smooth, bandlimited functions are particularly well suited to model its boundary. For more details how to transform the problem of recovering the contour from detected noisy boundary points into a 1-D nonuniform sampling problem we refer the reader to [8].



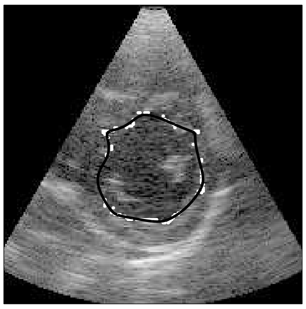
(a) 2-D echocardiography



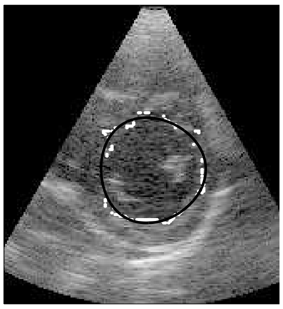
(b) Cross section of Left Ventricle



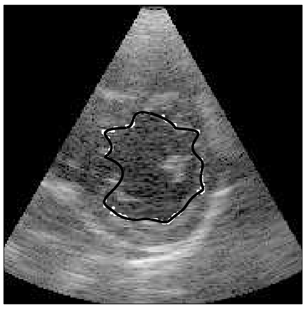
(c) Detected boundary points



(d) Recovered boundary of LV using proposed method



(e) Underfitted solution



(f) Overfitted solution

Figure 3: The recovery of the boundary of the Left Ventricle from 2-D ultrasound images is a basic step in echocardiography to extract relevant parameters of cardiac function. The contour in (d) has been computed by the multi-level method, it provides the optimal balance between fitting the data and preserving smoothness of the solution.

5. REFERENCES

[1] P. L. Butzer and G. Hinsen, "Two-dimensional nonuniform sampling expansions - an iterative approach," *I, II. Appl. Anal. 32*, pp. 53–68 and 69–85, 1989.

[2] J. Benedetto, "Irregular sampling and frames," in *Wavelets: A Tutorial in Theory and Applications* (C. K. Chui, ed.), pp. 445–507, Academic Press, 1992.

[3] H. G. Feichtinger and K. Gröchenig, "Error analysis in regular and irregular sampling theory," *Applicable Analysis*, vol. 50, pp. 167–189, 1993.

[4] H. G. Feichtinger, K. Gröchenig, and T. Strohmer, "Efficient numerical methods in non-uniform sampling theory," *Numerische Mathematik*, vol. 69, pp. 423–440, 1995.

[5] T. Pogany, "Statistical estimation of the bandwidth from irregularly spaced data," *Signal Processing*, vol. 54, no. 1, pp. 142–148, 1996.

[6] M. Hanke, "Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems," *Numer. Funct. Anal. Optim.*, vol. 18, no. 9-10, pp. 971–993, 1997.

[7] O. Scherzer and T. Strohmer, "A multi–level algorithm for the solution of moment problems," *Num.Funct.Anal.Opt.*, vol. 19, no. 3–4, pp. 353–375, 1998.

[8] T. Strohmer, "A Levinson-Galerkin algorithm for regularized trigonometric approximation." submitted, 1998.

[9] K. Gröchenig, "Irregular sampling of wavelet and short time Fourier transform," *Constr. Approx.*, vol. 9, pp. 283–297, 1993.

[10] M. Süssner, M. Budil, T. Strohmer, M. Greher, G. Porenta, and T. Binder, "Contour detection using artificial neuronal network presegmention," in *Proc. Computers in Cardiology*, (Vienna), 1995.