Adv Comput Math (2013) 38:737–762
DOI 10.1007/s10444-011-9258-4

Edge detection from truncated Fourier data using spectral mollifiers

Doug Cochran · Anne Gelb · Yang Wang

Received: 17 September 2011 / Accepted: 6 November 2011 /
Published online: 17 December 2011

Abstract Edge detection from a finite number of Fourier coefficients is challenging as it requires extracting local information from global data. The problem is exacerbated when the input data is noisy since accurate high frequency information is critical for detecting edges. The noise furthermore increases oscillations in the Fourier reconstruction of piecewise smooth functions, especially near the discontinuities. The edge detection method in Gelb and Tadmor (Appl Comput Harmon Anal 7:101–135, 1999, SIAM J Numer Anal 38(4):1389–1408, 2000) introduced the idea of "concentration kernels" as a way of converging to the singular support of a piecewise smooth function. The kernels used there, however, and subsequent modifications to reduce the impact of noise, were generally *oscillatory*, and as a result oscillations were always prevalent in the neighborhoods of the jump discontinuities. This paper revisits concentration kernels, but insists on *uniform convergence* to the "sharp

Communicated by Yuesheng Xu.

A. Gelb and D. Cochran were supported in part by NSF-DMS-FRG award 0652833.
Y. Wang was supported in part by NSF-DMS awards 0813750 and 1043034.

D. Cochran
Department of Electrical Engineering,
Arizona State University, Tempe, AZ 85287, USA
e-mail: doug.cochran@asu.edu

A. Gelb
School of Mathematical and Statistical Sciences,
Arizona State University, Tempe, AZ 85287, USA
e-mail: Anne.Gelb@asu.edu

Y. Wang (✉)
Department of Mathematics, Michigan State University,
East Lansing, MI 48824, USA
e-mail: ywang@math.msu.edu

peaks" of the function, that is, the edge detection method converges to zero away from the jumps without introducing new oscillations near them. We show that this is achievable via an admissible class of spectral mollifiers. Our method furthermore suppresses the oscillations caused by added noise.

Keywords Edge detection · Truncated Fourier data · Gibbs phenomenon · Poisson summation formula · Spectral mollifier

Mathematics Subject Classifications (2010) 42A10 · 42A50 · 65T10

1 Introduction

Detecting edges in piecewise smooth functions and images represents a fundamental problem for a variety of applications, including signal processing and numerical partial differential equation simulation. Algorithms using given pixel data, such as those developed by Canny [2] and Sobel [12] are generally based on constructing the numerical gradient of the image. Point values corresponding to large gradients constitute the edges of the image. In a similar fashion, the celebrated ENO scheme measures the gradients of neighboring pixel values to avoid crossing over jump discontinuities, or shocks, in solving partial differential equations that admit shocks as part of their solutions [10].

The situation is somewhat different, however, when the input data is given as a finite number of Fourier coefficients. Specifically, there is increased difficulty in having to determine a local phenomenon (an edge) from Fourier data, which contains global information. It is possible, of course, to first approximate the piecewise smooth function from the given Fourier data and then use a gradient based method to locate edges. However, the non-physical Gibbs oscillations near the jump discontinuities may be seen as edges in the gradient based methods [8]. Moreover, any filtering introduced to mitigate the Gibbs oscillations will "smooth over" corresponding edges [9]. This issue is further aggravated when noise is present in the data [11]. Hence edge detection algorithms that work directly from given Fourier data are useful, especially when noise is considered in their design.

In [5, 6], Gelb and Tadmor introduced the notion *concentration kernels* to detect jump discontinuities directly from Fourier data. Specifically, suppose f is a piecewise smooth function on [0, 1] with finitely many jump discontinuities a_k for $k = 1, \cdots, K$ on [0, 1]. We define

$$[f](x) := \begin{cases} f(x+) - f(x-) & x \in (0, 1) \\ f(0+) - f(1-) & x = 0, 1 \end{cases} \tag{1.1}$$

as the jump function associated with f. Note that the defined value $[f](0) = [f](1) = f(0+) - f(1-)$ is natural because of the periodic extension of f to

$\mathbb{R}$. A concentration kernel $K_N(x)$ is defined to satisfy certain properties (see [6]), and when convolved with the data function f it has the property that $f * K_N(x) \longrightarrow [f](x)$ on $(0, 1)$. In a typical setting the function K_N is smooth and hence $f * K_N$ is continuous. Therefore the convergence of $f * K_N$ to $[f]$ cannot be uniform. The idea is to find kernels K_N such that $f * K_N(x) \longrightarrow 0$ in smooth regions, with $f * K_N$ concentrated mostly around discontinuities (and hence the name "concentration kernel"). While concentration kernels are defined in more general settings [6], input truncated Fourier data is ubiquitous in many applications and therefore will be our focus in this investigation. In particular we will revisit the generalized conjugate partial sum where the kernel $K_N(x)$ has the form

$$K_N(x) = \sum_{n=-N}^{N} \omega\left(\frac{n}{N}\right) e^{2\pi i n x}.$$

With this kernel $f * K_N$ becomes

$$f * K_N(x) = \sum_{n=-N}^{N} \omega\left(\frac{n}{N}\right) \widehat{f}(n) e^{2\pi i n x}. \tag{1.2}$$

Thus $f * K_N(x)$ becomes a conjugate partial sum with a spectral mollifier. It was shown in [5] that by taking $\omega(t) = 2\pi i t \tau(t)$ where $\tau(t)$ is even and satisfying $\int_{-1}^{1} \tau = 1$ one has

$$f * K_N = [f](x) + \mathcal{O}_x\left(\frac{\log N}{N}\right),$$

and thus the convergence $f * K_N(x) \longrightarrow [f](x)$ on $[0, 1)$. The notation $\mathcal{O}_x$ implies dependence on x, and thus the error term is itself a function of x, which is dependent on $f'(x)$ (defined in smooth regions) as well as the data lost in the truncation. In fact, in [3], the error was calibrated as

$$f * K_N(x) = [f](x) + \begin{cases} \mathcal{O}_x\left(\dfrac{\log N}{N}\right) & d(x) \leq \dfrac{\log N}{N} \\ \mathcal{O}\left(\dfrac{\log N}{(N d(x))^s}\right) & d(x) \gg \dfrac{1}{N} \end{cases} \tag{1.3}$$

where $d(x)$ denotes the distance between a point in the domain and the nearest discontinuity, and the convergence rate $s > 0$ depends upon the concentration factor chosen [3, 6]. However, this convergence rate *does not* address the problem of suppressing the undesirable Gibbs oscillations near the discontinuities. These oscillations can result in false edges, especially when noise is present. Hence a main focus of this paper is to study ways to control them.

Concentration kernels have been studied rather extensively, and a good review can be found in [15]. While methods using concentration kernels have steadily improved [3, 4, 11], and are used effectively in a variety of (noisy) applications, there are still issues remaining in terms of of convergence. It is well understood that a concentration kernel yielding fast convergence

in smooth regions away from discontinuities will necessarily induce Gibbs oscillations near them. On the other hand, too much emphasis placed on controlling the Gibbs oscillations near the jumps will both reduce the accuracy of the jump magnitude approximations as well as reduce the convergence rate away from them. As our analysis in this paper will demonstrate, there is a type of "uncertainty principle" trade-off in play. These problems are further exacerbated when there are multiple jumps in the domain, especially if they are in close proximity. Noise may also adversely affect the results. Several techniques have been introduced to address these problems, and even exploit the "uncertainty principle" pattern of oscillations observed near the jump discontinuities [4, 7, 11, 13].

The fact that $[f]$ is discontinuous and is supported on a finite set makes the convergence of $K_N * f$ to $[f]$ tricky. Clearly there is no uniform convergence. As a result, the idea of convergence rate is not well defined. In this study we introduce the notion of *uniform convergence to sharp peaks*. Instead of the convergence directly to $[f]$, as defined in (1.1), we consider the approximation of $[f]$ by a sequence of smooth pulses, Q_N, which are increasingly concentrated at the jumps and which have few or no oscillations near or away from the jumps, see e.g. Fig. 1. To this end we introduce a class of spectral mollifiers, which can be viewed as concentration kernels but are modified from the conventional ones discussed in [6], so that given a finite number of Fourier coefficients, our new mollified partial Fourier sum approximates the functions of sharp peaks (smooth pulses) of Q_n *uniformly*. The Gibbs oscillations in the region of the jumps are controlled and the error in the smooth regions in (1.3) caused by the construction of $f'(x)$ and the impact of using truncated data is reduced. In other words, our method insists on a single convergence rate near the jump discontinuities and in smooth regions while at the same time maintaining sharp localization of detected edges. To best achieve this goal, we show that one must in general use a modified version of the generalized conjugate partial sum of the form[1]

$$T_N(x) = \sum_{n=-N}^{N} \omega\left(\frac{n}{\lambda_N}\right) \widehat{f}(n) e^{2\pi i n x}, \tag{1.4}$$

which differs from the generalized conjugate partial sum (1.2) in that λ_N typically must grow more slowly than N, i.e. λ_N/N tends to 0. Similar to the concentration kernels, the mollifier $\omega(t)$ will have the form $\omega(t) = 2\pi i t\widehat{\sigma}(t)$, where $\widehat{\sigma}$ can be drawn from a broad class of functions. For example, any function from the Schwartz class will qualify as long as $\sigma(0) = \int_{\mathbb{R}} \widehat{\sigma} = 1$. Our investigation determines an admissible sequence for λ_N. An optimal sequence ultimately depends upon the characteristics of the particular spectral mollifier $\omega(t)$ used as well as other factors such as noise.

[1]The fact that T_N and subsequent operators work on f is implicitly understood.

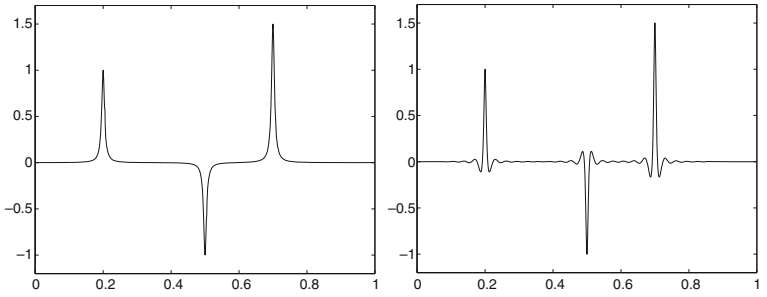

Fig. 1 $f_\lambda(x)$ for $\sigma(x) = \frac{1}{1+x^2}$ (*left*) and $\sigma(x) = \frac{\cos x}{1+x^2}$ (*right*), respectively and $\lambda = 200$

We point out that our insistence on a *single* convergence rate puts a theoretical limit on the rate of convergence, which is no better than $\mathcal{O}\left(\frac{\log N}{N}\right)$ (see Theorem 2.14). The idea of controlling both the Gibbs-like oscillations and the convergence away from jumps was first discussed in [6]. The focus in that investigation, however, was on a family of oscillatory concentration kernels, which allow for higher convergence rates away from jump discontinuities. We now revisit the spectral mollifiers introduced there as a way to obtain a uniform convergence bound throughout the domain, even in the presence of noise. A similar idea was also developed in [3], where the types of errors in the concentration kernel were separated in terms of the smoothness of the region. However, no attempt was made there to derive uniform convergence estimates.

The rest of the paper is organized as follows. In Section 2 we provide the mathematical analysis of edge detection from truncated Fourier data. A key ingredient is the classical Poisson Summation Formula. We introduce the notion of uniform convergence to sharp peaks and analyze the convergence of our method. In Section 3 we study the robustness of the method, especially when noise is present in the data. Our numerical results demonstrate the effectiveness of our method. We compare these results to those of the original concentration method [5], and in particular with those from [3, 4] when noise is present. Some concluding remarks are provided in Section 4.

2 Edge detection using spectral mollifiers

2.1 Edge detection from Fourier data

Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a piecewise smooth (say C^2) function with compact support. Assume that f has discontinuities at $\{a_k : 1 \le k \le K\}$, which are all jump discontinuities (edges). We define $E \subset [0, 1)$ to be the set of jumps of $f(x)$ in $[0, 1)$, i.e. $E = \{a_k : 1 \le k \le K\} \cap [0, 1)$. We shall also use E^+ to denote the set $E^+ = E \cup \{1\}$. This notation is needed to address the potential extra discontinuities of f introduced at boundary points from its periodic extension,

and will come handy in our proofs.[2] Throughout this paper we assume that f is supported on the unit interval [0, 1) and it is piecewise smooth with only jump discontinuities (edges).

To determine the edges of f, we will use the derivative Df of f, which is a distribution given by

$$Df(x) = \sum_{k=1}^{K} [f]_*(a_k)\delta(x - a_k) + g(x)$$

where $[f]_*(x) := f(x^+) - f(x^-)$ and $g(x) = f'(x)$ for $x \notin \{a_k\}$ and $g(a_k) = 0$. Taking the Fourier transform of Df yields

$$2\pi i\xi\, \widehat{f}(\xi) = \sum_{k=1}^{K} [f]_*(a_k)e^{-2\pi i a_k \xi} + \widehat{g}(\xi). \tag{2.1}$$

Note that g is piecewise C^1 with possible discontinuities also at $\{a_k\}$, and its Fourier transform $\widehat{g}(\xi)$ decays at the rate $\mathcal{O}(|\xi|^{-1})$ as $|\xi| \to \infty$. Thus (2.1) yields the property (c.f. [5])

$$2\pi i\xi\, \widehat{f}(\xi) = \sum_{k=1}^{K} [f]_*(a_k)e^{-2\pi i a_k \xi} + \mathcal{O}(|\xi|^{-1}). \tag{2.2}$$

Since differentiation of the function f yields Dirac pulses at the discontinuities, a quick and simple way to detect edges is to consider the conjugated sum of Df, whose Fourier transform is given by $2\pi i\xi\, \widehat{f}(\xi)$. The resulting edge detector, when properly normalized, turns out to be identical to the first order polynomial concentration kernel edge detector with $\omega(t) = \pi i t$ [5]. Normalized by the size of the data it yields

$$\tilde{T}_N(x) = \frac{2\pi i}{2N+1} \sum_{n=-N}^{N} n\, \widehat{f}(n)e^{2\pi i n x}. \tag{2.3}$$

Using (2.1) and the Dirichlet kernel, $D_N(x) = \sum_{n=-N}^{N} e^{2\pi i n x}$, we obtain

$$\tilde{T}_N(x) = \frac{1}{2N+1} \sum_{a \in E^+} [f]_*(a) D_N(x - a) + \frac{1}{2N+1} \sum_{n=-N}^{N} \widehat{g}(n)e^{2\pi i n x}. \tag{2.4}$$

For large N, $\frac{1}{2N+1} D_N(x - a)$ has a peak of height 1 at $x = a$ and it is easy to show that the second sum $\frac{1}{2N+1} \sum_{n=-N}^{N} \widehat{g}(n)e^{2\pi i n x} = o(1)$. As a result $\tilde{T}_N$ can be a simple and sometimes effective way to detect edges of f. Unfortunately, as is evident in Fig. 3 (top), due to the oscillations of D_N and lack of robustness in the presence of noise, $\tilde{T}_N$ with this choice of ω is not an ideal edge detector. Moreover, other concentration factors introduced in [5, 6] yield faster convergence away from the jump locations (peaks).

[2] This definition is necessary since we do not assume periodicity of f, as is usually done, see e.g. [5].

2.2 A mollified Fourier partial sum

Let $\sigma(x)$ be a function in a suitable class. We now consider the mollified partial Fourier sum

$$T_N[\sigma](x) = 2\pi i \sum_{n=-N}^{N} n\widehat{\sigma}(n)\,\widehat{f}(n)e^{2\pi inx}. \tag{2.5}$$

It follows from (2.1) that

$$\begin{aligned} T_N[\sigma](x) &= \sum_{n=-N}^{N} \widehat{\sigma}(n) \left(\sum_{a\in E^+} [f]_*(a)e^{-2\pi ian} + \widehat{g}(n) \right) e^{2\pi inx} \\ &= \sum_{a\in E^+} [f]_*(a) \sum_{n=-N}^{N} \widehat{\sigma}(n)e^{2\pi in(x-a)} + \sum_{n=-N}^{N} \widehat{\sigma}(n)\widehat{g}(n)e^{2\pi inx} \\ &= \sum_{a\in E} [f](a) \sum_{n=-N}^{N} \widehat{\sigma}(n)e^{2\pi in(x-a)} + \sum_{n=-N}^{N} \widehat{\sigma}(n)\widehat{g}(n)e^{2\pi inx} \\ &= \sum_{a\in E} [f](a)A_N[\sigma](x-a) + B_N[\sigma](x), \end{aligned}$$

where

$$A_N[\sigma](x) = \sum_{n=-N}^{N} \widehat{\sigma}(n)e^{2\pi inx} \quad \text{and} \quad B_N[\sigma](x) = \sum_{n=-N}^{N} \widehat{\sigma}(n)\widehat{g}(n)e^{2\pi inx}. \tag{2.6}$$

In order to detect edges while suppressing the Gibbs phenomenon, we must first estimate $A_N[\sigma](x-a)$ and $B_N[\sigma](x)$. An important class of continuous functions for which $\sigma(x)$ belongs is

$$\mathcal{C} = \left\{\sigma : \sigma(0) = \max|\sigma| = 1 \text{ and } |\sigma(t)| + |\widehat{\sigma}(t)| \le C_0(1+|t|)^{-1-\delta} \text{ for some } C_0, \delta > 0\right\}. \tag{2.7}$$

Clearly any real function σ in the Schwartz class with $\sigma(0) = \max|\sigma| = 1$ is in $\mathcal{C}$.[3] For edge detection, it is desirable to restrict functions to a subspace $\mathcal{C}_0$ of bell-shaped functions given by

$$\mathcal{C}_0 = \left\{\sigma \in \mathcal{C} : \ \sigma'(x) \ge 0 \ \text{ on } (0,\infty) \text{ and } \ \sigma'(x) \le 0 \ \text{ on } (-\infty, 0).\right\}.$$

A key ingredient for edge detection will be the Poisson Summation Formula for functions in $\mathcal{C}$:

[3] We note that this is similar to the bump function described in [6]. The difference here is that we are starting with a derivative approximation (2.5).

Proposition 2.1 (Poisson summation formula) *Let $\sigma(x) \in \mathcal{C}$. Then*

$$\sum_{n\in\mathbb{Z}} \sigma(n) = \sum_{n\in\mathbb{Z}} \widehat{\sigma}(n).$$

The proof of Poisson Summation Formula can be found in most analysis textbooks, e.g. Stein and Weiss [14].

Lemma 2.2 *Let $\sigma \in \mathcal{C}$ and let $A_N[\sigma](x)$ be as in (2.6). Then*

$$A_N[\sigma](x) = \sum_{n\in\mathbb{Z}} \sigma(n+x) - \sum_{|n|>N} \widehat{\sigma}(n) e^{2\pi i n x}.$$

Proof Let $h(t) := \sigma(t+x)$. Then $\widehat{h}(\xi) = e^{2\pi i \xi x}\widehat{\sigma}(\xi)$. By the Poisson Summation Formula

$$\sum_{n\in\mathbb{Z}} h(n) = \sum_{n\in\mathbb{Z}} \sigma(n+x) = \sum_{n\in\mathbb{Z}} \widehat{h}(n) = \sum_{n\in\mathbb{Z}} \widehat{\sigma}(n) e^{2\pi i n x}.$$

It follows that

$$A_N[\sigma](x) = \sum_{n=-N}^{N} \widehat{\sigma}(n) e^{2\pi i n x} = \sum_{n\in\mathbb{Z}} \sigma(n+x) - \sum_{|n|>N} \widehat{\sigma}(n) e^{2\pi i n x}.$$

□

We shall now introduce two notations for the rest of the paper. For any function σ and nonzero $\lambda \in \mathbb{R}$ we shall use $Q[\sigma]$ and σ_λ to denote respectively

$$Q[\sigma](x) := \sum_{n\in\mathbb{Z}} \sigma(n+x), \qquad \sigma_\lambda(x) := \sigma(\lambda x), \tag{2.8}$$

which play an important role in our edge detection method. Note that $Q[\sigma](x)$ is continuous and $\mathbb{Z}$-periodic, and $\widehat{\sigma_\lambda}(\xi) = \lambda^{-1}\widehat{\sigma}(\lambda^{-1}\xi)$. By Lemma 2.2

$$T_N[\sigma](x) = \sum_{a\in E} [f](a)\, Q[\sigma](x-a) + C_N[\sigma](x) + B_N[\sigma](x), \tag{2.9}$$

where E is again the set of discontinuities of f in $[0, 1)$, B_N is given by (2.6), and

$$C_N[\sigma](x) = -\sum_{a\in E} \sum_{|n|>N} [f](a)\widehat{\sigma}(n) e^{2\pi i n(x-a)}. \tag{2.10}$$

The set of edges E will be determined by choosing σ so that $Q[\sigma](x)$ consists of sharply localized peaks at 0 and other integer points while both $B_N[\sigma]$, $C_N[\sigma]$ in (2.9) are suppressed as much as possible. This leads to the following key definition for the purpose of edge detection:

Definition 2.3 Let $F \subset [0,1]$ be a finite set. A sequence of functions $G_n(x)$ is said to *converge uniformly in* $[0,1]$ *to the sharp peaks at* F if there exist a $\sigma(x) \in \mathcal{C}$, a height function $h : F \to \mathbb{R}$, and a sequence $\lambda_N \to \infty$ such that

$$\lim_{N\to\infty} \left| G_N(x) - \sum_{a\in F} h(a)\sigma_{\lambda_N}(x-a) \right| = 0 \qquad \text{uniformly.}$$

Note that for any $\sigma \in \mathcal{C}$ and $\lambda \neq 0$ the term $h(a)\sigma_{\lambda_n}(x-a)$ has a peak at a with height $h(a)$. As λ_N becomes larger the width of the peak will shrink so the peak becomes sharper. The desirability of taking $\sigma \in \mathcal{C}_0$ is that there will be no oscillations near the peak, as such oscillations can be confused for real peaks. Figure 1 shows the plots of the function

$$f_\lambda(x) = \sigma_\lambda(x-0.2) - \sigma_\lambda(x-0.5) + 1.5\sigma_\lambda(x-0.7)$$

for $\sigma(x) = \frac{1}{1+x^2}$ (left) and $\sigma(x) = \frac{\cos x}{1+x^2}$ (right) respectively, with $\lambda = 200$. Oscillations near the peaks is not a problem in the first case but clearly is a problem in the second because σ is not in $\mathcal{C}_0$. For the purposes of edge detection, we require the height function h to be equal to the corresponding jump values $[f](x)$ at the discontinuities E.

Lemma 2.4 *Let* $E \subset [0,1)$ *be a the set of jump discontinuities of* f *and let* $E^+ = E \cup \{1\}$. *For any* $\sigma \in \mathcal{C}$ *and sequence* $\lambda_N \to \infty$ *we have*

$$\lim_{N\to\infty} \sup_{x\in[0,1]} \left| \sum_{a\in E^+} [f](a)\sigma_{\lambda_N}(x-a) - \sum_{a\in E} [f](a)\, Q[\sigma_{\lambda_N}](x-a) \right| = 0.$$

Proof Substitute in $Q[\sigma_{\lambda_N}](x) := \sum_{n\in\mathbb{Z}} \sigma_{\lambda_N}(n+x)$ we have

$$\begin{aligned}\sum_{a\in E} [f](a)\, Q[\sigma_{\lambda_N}](x-a) &= \sum_{n\in\mathbb{Z}} \sum_{a\in E} [f](a)\sigma_{\lambda_N}(x-a+n) \\ &= \sum_{n\in\mathbb{Z}} \sum_{a\in E\cap(0,1)} [f](a)\sigma_{\lambda_N}(x-a+n) \\ &\quad + \sum_{n\in\mathbb{Z}} [f](0)\sigma_{\lambda_N}(x+n).\end{aligned}$$

Set $h(x) := \sum_{a\in E^+} [f](a)\sigma_{\lambda_N}(x-a) - \sum_{a\in E} [f](a)\, Q[\sigma_{\lambda_N}](x-a)$. It follows that

$$h(x) = \sum_{n\neq 0} \sum_{a\in E\cap(0,1)} [f](a)\sigma_{\lambda_N}(x-a+n) + \sum_{n\neq 0,-1} [f](0)\sigma_{\lambda_N}(x+n).$$

Now there exists some $\delta > 0$ such that $|\sigma(x)| \leq C|x|^{-1-\delta}$ for all $x \neq 0$. Hence for any $x, a \in (0,1)$ we have

$$\sum_{n\neq 0} |\sigma_{\lambda_N}(x-a+n)| \leq \frac{2C}{|\lambda_N|^{1+\delta}} \sum_{n=1}^{\infty} \frac{1}{(n+x-a)^{1+\delta}} \leq \frac{C(a)}{|\lambda_N|^{1+\delta}},$$

where $C(a)$ is a constant depending only on a (more precisely, only on the distance between a and the boundary points 0 and 1). A similar estimates also holds for other terms in $h(x)$. Thus there exists a constant C^* depending only on E such that

$$\sup_{x\in[0,1]} |h(x)| \le C^* \sum_{a\in E} |[f](a)|\lambda_N^{-1-\delta}. \tag{2.11}$$

The lemma now follows. □

A consequence of the above lemma is that by proving $T_N[\sigma_{\lambda_N}](x) - \sum_{a\in E}[f](a)\,Q[\sigma_{\lambda_N}](x-a)$ tends to 0 uniformly, which we establish later, we have equivalently shown that $T_N[\sigma_{\lambda_N}](x)$ converges uniformly to sharp peaks at the discontinuities of f with height function $[f]$.

An *admissible* sequence for σ is constructed by bounding the error terms in (2.9). We therefore have

Definition 2.5 Let $\sigma \in \mathcal{C}$ and $\{\lambda_N\}_{N\in\mathbb{N}}$ be a sequence of positive numbers. We call $\{\lambda_N\}$ an *admissible* sequence for σ if $\lambda_N \to \infty$ and

$$\lim_{N\to\infty} \sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| = 0 \quad \text{and} \quad \lim_{N\to\infty} \sum_{n=-N}^{N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1} = 0.$$

Note that $\widehat{\sigma_{\lambda_N}}(\xi) := \lambda_N^{-1}\widehat{\sigma}(\lambda_N^{-1}x)$.

Theorem 2.6 *Let $\sigma \in \mathcal{C}$ and $\{\lambda_N\}_{N\in\mathbb{N}}$ be an admissible sequence for σ. Then $T_N[\sigma_{\lambda_N}](x)$ converges uniformly on $[0,1)$ to sharp peaks with height function $[f]$ at E, the set of discontinuities of f in $[0,1)$. More precisely,*

$$\lim_{N\to\infty} \left(T_N[\sigma_{\lambda_N}](x) - \sum_{a\in E}[f](a)\,Q[\sigma_{\lambda_N}](x-a) \right) = 0 \tag{2.12}$$

uniformly in $[0,1)$.

Proof By (2.9) we have

$$T_N[\sigma_{\lambda_N}](x) = \sum_{a\in E}[f](a)\,Q[\sigma_{\lambda_N}](x-a) + C_N[\sigma_{\lambda_N}](x) + B_N[\sigma_{\lambda_N}](x).$$

Observe that

$$\begin{aligned} \left|C_N[\sigma_{\lambda_N}](x)\right| &\le \sum_{a\in E}\left|[f](a)\right| \sum_{|n|>N} \left|\widehat{\sigma_{\lambda_N}}(n)e^{2\pi i n(x-a)}\right| \\ &= \left(\sum_{a\in E}\left|[f](a)\right|\right) \sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)|. \end{aligned}$$

Because $\{\lambda_N\}$ is admissible for σ, it follows that $\left\|C_N[\sigma_{\lambda_N}]\right\|_\infty = o(1)$. Also by (2.6) $B_N[\sigma_{\lambda_N}](x) = \sum_{n=-N}^{N} \widehat{\sigma_{\lambda_N}}(n)\widehat{g}(n)e^{2\pi i n x}$. Since g is piecewise smooth with finitely many jump singularities we have $\widehat{g}(n) = \mathcal{O}(n^{-1})$. Thus there exists a constant $C_0 > 0$ such that

$$\left\|B_N[\sigma_{\lambda_N}]\right\|_\infty \le C_0 \sum_{n=-N}^{N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1} = o(1).$$

These estimates yield

$$\lim_{N\to\infty} \left(C_N[\sigma_{\lambda_N}](x) + B_N[\sigma_{\lambda_N}](x)\right) = 0 \tag{2.13}$$

uniformly. It follows that

$$\lim_{N\to\infty} \left(T_N[\sigma_{\lambda_N}](x) - \sum_{a\in E} [f](a)\, Q[\sigma_{\lambda_N}](x-a) \right) = 0$$

uniformly, proving the theorem. □

Corollary 2.7 *Let $\sigma \in \mathcal{C}$ and $\{\lambda_N\}_{N\in\mathbb{N}}$ be an admissible sequence for σ. Then for $x \in [0, 1)$,*

$$\lim_{N\to\infty} T_N[\sigma_{\lambda_N}](x) = [f](x).$$

Proof The corollary follows immediately from the observation that $\lim_{N\to\infty} Q[\sigma_{\lambda_N}](x-a) = 0$ if $x - a \notin \mathbb{Z}$ and $\lim_{N\to\infty} Q[\sigma_{\lambda_N}](x-a) = 1$ if $x - a \in \mathbb{Z}$. □

Remark 2.8 Although $T_N[\sigma_{\lambda_N}]$ converges uniformly to sharp peaks at the edges of f, with a fixed finite data size N it is important that we choose the function $\sigma(x) \in \mathcal{C}$ so that the peaks of $T_N[\sigma_{\lambda_N}](x)$ at the edges are as sharp as possible for detection, and that no artificial oscillations are introduced in the vicinity. That said, if $\sigma(x)$ is oscillatory, then $T_N[\sigma_{\lambda_N}](x)$ will produce multiple peaks bundled together at the edges, instead of just one peak. This is in fact what happens with the oscillatory kernels introduced in [5, 6]. To eliminate this artifact its is desirable that we choose a bell-shaped $\sigma \in \mathcal{C}_0$, see Fig. 1. Also, it is preferable to choose $\sigma(x)$ to be an even function so that $\widehat{\sigma}$ is real. One such σ we often use in practice is the Gaussian $\sigma(x) = e^{-\pi^2 x^2}$, which has $\widehat{\sigma}(t) = \frac{1}{\sqrt{\pi}} e^{-t^2}$.

The following theorem shows that admissible sequences for any function $\sigma \in \mathcal{C}$ are easy to find.

Theorem 2.9 *Let $\{\lambda_N\}$ be a sequence with $\lim_N \lambda_N = \infty$.*

(1) *Assume that $\lim_N \lambda_N/N = 0$. Then $\{\lambda_N\}$ is an admissible sequence for any $\sigma \in \mathcal{C}$.*
(2) *Assume that $\limsup_N \lambda_N/N = a > 0$. Then $\{\lambda_N\}$ is an admissible sequence for $\sigma \in \mathcal{C}$ if and only if $\operatorname{supp}(\widehat{\sigma}) \subseteq [-a^{-1}, a^{-1}]$.*

(3) *Assume that* $\limsup_N \lambda_N/N = \infty$. *Then* $\{\lambda_N\}$ *is not an admissible sequence for any* $\sigma \in \mathcal{C}$.

Proof

Part (1) We only need to show that for $\widehat{\sigma_{\lambda_N}}(t) = \lambda_N^{-1}\widehat{\sigma}(\lambda_N^{-1}t)$ we have

$$\lim_{N\to\infty} \sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| = 0 \quad \text{and} \quad \lim_{N\to\infty} \sum_{n=-N}^{N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1} = 0.$$

Now $\sigma \in \mathcal{C}$ so that there exist constants $C_0, \delta > 0$ and such that $|\widehat{\sigma_{\lambda_N}}(t)| = \lambda_N^{-1}|\widehat{\sigma}(\lambda_N^{-1}t)| \le C_0\lambda_N^{-1}(|\lambda_N^{-1}t|+1)^{-1-\delta}$. It follows that

$$\begin{aligned}\sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| &\le C_0\lambda_N^{-1} \sum_{|n|>N}^{\infty} \frac{1}{|\lambda_N^{-1}n|^{1+\delta}} \\ &\le C_0\lambda_N^{\delta} \sum_{|n|>N}^{\infty} \frac{1}{|n|^{1+\delta}} \\ &= \mathcal{O}\left(\frac{\lambda_N^{\delta}}{N^{\delta}}\right).\end{aligned}$$

Therefore,

$$\lim_{N\to\infty} \sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| = 0.$$

To establish the other estimate, we break down the sum into two parts,

$$\sum_{n=-N}^{N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1} = \sum_{|n|\le\lambda_N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1} + \sum_{\lambda_N<|n|\le N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1}.$$

The first sum has a bound

$$\sum_{|n|\le\lambda_N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1} \le \frac{\|\widehat{\sigma}\|_{\infty}}{\lambda_N} \sum_{|n|\le\lambda_N} \frac{1}{|n|+1} = \mathcal{O}\left(\frac{\log\lambda_N}{\lambda_N}\right).$$

The second sum has a bound

$$\begin{aligned}\sum_{|n|\lambda_N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1} &\le C_0\,\lambda_N^{-1} \sum_{|n|>\lambda_N} \frac{1}{(|\lambda_N^{-1}n|+1)^{1+\delta}(n+1)} \\ &\le C_0\lambda_N^{\delta} \sum_{|n|>\lambda_N} \frac{1}{|n|^{2+\delta}} \\ &= \mathcal{O}(\lambda_N^{-1}).\end{aligned}$$

Thus both sums tend to 0 as $N\to\infty$, proving the theorem.

Part (2) Suppose $\operatorname{supp}(\widehat{\sigma}) \subseteq [-a^{-1}, a^{-1}]$. Observe that we have already shown in the proof of Part (1) that $\lim_N \sum_{|n|\le\lambda_N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1} = 0$ as long as $\lambda_N \to \infty$. So we only need to establish $\lim_{N\to\infty} \sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| = 0$. For any $\varepsilon > 0$ we have $N/\lambda_N > a^{-1} - \varepsilon$ for sufficiently large N by the hypothesis. Hence

$$\sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| \le \sum_{|n|>(a^{-1}-\varepsilon)\lambda_N} |\widehat{\sigma_{\lambda_N}}(n)| = \lambda_N^{-1} \sum_{|n|>(a^{-1}-\varepsilon)\lambda_N} \widehat{\sigma}\left(\frac{n}{\lambda_N}\right).$$

Now $\operatorname{supp}(\widehat{\sigma}) \subseteq [-a^{-1}, a^{-1}]$. Thus $\widehat{\sigma}(\frac{n}{\lambda_N}) = 0$ whenever $|n| > a^{-1}\lambda_N$. Considering the above as a Riemann sum it follows that

$$\lim_{N\to\infty} \sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| \le \int_{a^{-1}-\varepsilon\le|t|\le a^{-1}-\varepsilon} |\widehat{\sigma}(t)|\, dt \le 2\|\widehat{\sigma}\|_\infty \varepsilon.$$

This proves the if direction of Part (2). For the only if direction, take a subsequence λ_{N_k} such that $\limsup_k \lambda_{N_k}/N_k = a > 0$. Again, the Riemann sum estimate of $\sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)|$ yields easily

$$\lim_{k\to\infty} \sum_{|n|>N_k} |\widehat{\sigma}_{N_k}(n)| = \int_{|t|\ge a^{-1}} |\widehat{\sigma}(t)|\, dt.$$

The right hand integral is 0 only if $\operatorname{supp}(\widehat{\sigma}) \subseteq [-a^{-1}, a^{-1}]$.

Part (3) This follows from Part (2) immediately. □

Remark 2.10 The classic concentration kernel introduced in [5, 6] is equivalent to the function $2\pi i\widehat{\sigma}(t)$ with $\lambda_N = N$ up to a normalization constant. In our method we normalize σ so that $\sigma(0) = \int_{\mathbb{R}} \widehat{\sigma} = 1$. With classic concentration kernels, by taking $\lambda_N = N$, the values $\frac{k}{N}$ stays within $[-1, 1]$. This can be viewed in our construction as taking $\widehat{\sigma}\chi_{[-1,1]}$ as the mollifier, with normalization $\int_{\mathbb{R}} \widehat{\sigma}\chi_{[-1,1]} = \int_{-1}^{1} \widehat{\sigma} = 1$. In virtually all examples in the concentration kernel literature, either $\widehat{\sigma}$ does not have compact support, in which case $\lambda_N = N$ is not admissible, or $\widehat{\sigma}$ has compact support on $[-1, 1]$ but σ does not belong to $\mathcal{C}_0$ so it is oscillatory. The kernels described there were often chosen to enable fast convergence away from the jump discontinuities, yielding Gibbs oscillations in the physical domain near the jump discontinuities as a necessary byproduct. These effects are seen in the numerical examples in Section 3. Another side effect of these oscillations is that when there are edges that are

in close proximity (or not adequately resolved by the given number of Fourier coefficients), the values of the edges are often not captured correctly. Although this problem can be addressed by choosing $\sigma \in \mathcal{C}$ so that $\widehat{\sigma}$ is compactly supported, it is generally difficult to construct such a $\sigma \in \mathcal{C}_0$. Instead, nonlinear postprocessing techniques, e.g. [7], are often used to mitigate the unwanted oscillations.

Remark 2.11 The rate of the uniform convergence of (2.12) is the rate of uniform convergence of $C_N[\sigma_{\lambda_N}](x) + B_N[\sigma_{\lambda_N}](x)$ in (2.13) with the decay rate of $\|C_N[\sigma_{\lambda_N}]\|_\infty$ governed by $\sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)|$, and the decay rate of $\|B_N[\sigma_{\lambda_N}]\|_\infty$ governed by $\sum_{n=-N}^{N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1}$. For effective edge detection, $\sigma(x)$ and the sequence $\{\lambda_n\}$ should be chosen so that both

$$\sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| \to 0 \quad \text{and} \quad \sum_{n=-N}^{N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1} \to 0$$

as quickly as possible. The following corollary shows what convergence rates are attainable for special classes of mollifiers. In Theorem 2.14 we prove that the convergence rate will be *no better than* $\mathcal{O}(\log N/N)$, regardless of the choices for σ and λ_N.

Corollary 2.12 *Under the assumptions of Theorem* 2.6 *we have*

(A) *Assume that* $|\widehat{\sigma}(t)| \le C_0(1+|t|)^{-1-\delta}$ *for some* $C_0, \delta > 0$. *Then for* $\lambda_N = c\, N^q \log^{1-q} N$ *where* $c > 0$ *and* $q = \frac{\delta}{1+\delta}$ *we have*

$$T_N[\sigma_{\lambda_N}](x) - \sum_{a\in E} h_f(a)\, Q[\sigma_{\lambda_N}](x-a) = \mathcal{O}\left(\frac{\log^q N}{N^q}\right). \tag{2.14}$$

In particular if σ *is in the Schwartz class then* q *can be arbitrarily close to* 1.

(B) *Assume that* $|\widehat{\sigma}(t)| \le C_0 e^{-\alpha|t|^\delta}$ *for some* $C_0, \alpha > 0$ *and* $\delta \ge 1$. *Then for* $\lambda_N = c\, N/\log^{q-1} N$ *where* $q = 1 + \delta^{-1}$ *and* $0 < c \le \alpha^{q-1}$ *we have*

$$T_N[\sigma_{\lambda_N}](x) - \sum_{a\in E} h_f(a)\, Q[\sigma_{\lambda_N}](x-a) = \mathcal{O}\left(\frac{\log^q N}{N}\right). \tag{2.15}$$

Proof We need only establish the desired bounds for both $\sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)|$ and $\sum_{n=-N}^{N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1}$, as described in Remark 2.11. Suppose that $|\widehat{\sigma}(t)| \le C_0(1+|t|)^{-1-\delta}$. Then, by the proof of Theorem 2.9, we have

$$\sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| = \mathcal{O}\left(\frac{\lambda_N^\delta}{N^\delta}\right) \quad \text{and} \quad \sum_{n=-N}^{N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1} = \mathcal{O}\left(\frac{\log \lambda_N}{\lambda_N}\right).$$

It is straightforward to check that for $\lambda_N = c\,N^q \log^{1-q} N$ we have both $\mathcal{O}(\frac{\lambda_N^\delta}{N^\delta}) = \mathcal{O}(\frac{\log^q N}{N^q})$ and $\mathcal{O}(\frac{\log \lambda_N}{\lambda_N}) = \mathcal{O}(\frac{\log^q N}{N^q})$, proving part (A). Note that if σ is in the Schwartz class then δ can be arbitrarily large, and hence q can be arbitrarily close to 1.

Proof of (B) requires a better estimate for $\sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)|$. By the hypothesis,

$$\sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| \le \frac{C_0}{\lambda_N} \sum_{|n|>N} e^{-\alpha|\lambda_N^{-1} n|^\delta}.$$

A standard integral estimate provides a bound for the right hand side by $2C_0 \int_{A_N}^\infty e^{-\alpha t^\delta}\,dt$, where $A_N = N/\lambda_N$. Since $\delta \ge 1$, for $t \ge A_N$ we have

$$e^{-\alpha t^\delta} \le e^{-\alpha A_N^\delta} e^{-\alpha(t^\delta - A_N^\delta)} \le e^{-\alpha A_N^\delta} e^{-\alpha(t-A_N)A_N^{\delta-1}}.$$

It follows that

$$\int_{A_N}^\infty e^{-\alpha t^\delta}\,dt \le e^{-\alpha A_N^\delta} \int_{A_N}^\infty e^{-\alpha(t-A_N)A_N^{\delta-1}}\,dt \le \frac{1}{\alpha A_N^{\delta-1}} e^{-\alpha A_N^\delta}. \tag{2.16}$$

Now $A_N = N/\lambda_N = c^{-1}\log^{q-1} N$ where $q - 1 = 1/\delta$. Substituting it in (2.16) immediately yields $\sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| = o(\frac{\log^q N}{N^{\frac{\alpha}{c^\delta}}})$, which is $o(\frac{\log^q N}{N})$ since $\alpha/c^\delta \ge 1$ by assumption. Furthermore one can check that $\mathcal{O}(\frac{\log \lambda_N}{\lambda_N}) = \mathcal{O}(\frac{\log^q N}{N})$ in this case, proving (B). □

Remark 2.13 As we have shown, for edge detection we want to choose $\sigma(x)$ and the sequence $\{\lambda_N\}$ such that

$$P_N := \sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)| + \sum_{n=-N}^{N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1} \to 0 \tag{2.17}$$

as quickly as possible. To achieve this, we may be tempted to choose σ so that $\widehat{\sigma}$ decays very quickly. At the same time, $T_N[\sigma_{\lambda_N}](x)$ should resolve the sharp peaks of $[f]$ at the discontinuities E of f for fixed N. The sharpness of the peaks at the edges depends on how concentrated $\sigma_{\lambda_N}(x) = \sigma(\lambda_N x)$ is at 0. Clearly, the larger λ_N is, the more concentrated $\sigma_{\lambda_N}(x)$ is. However, these two objectives are conflicting. By making λ_N larger, the decay of $\sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)|$ will be slower. In addition, the Heisenberg Uncertainty Principle, which states that

$$\left(\int_{\mathbb{R}} x^2 |\sigma(x)|^2\,dx\right)\left(\int_{\mathbb{R}} t^2 |\widehat{\sigma}(t)|^2\,dt\right) \ge \frac{\|f\|_2^4}{16\pi^2},$$

also implies that by making $\widehat{\sigma}$ more concentrated at 0, σ itself will be less concentrated around 0. Our next theorem provides a more quantitative analysis of this dilemma by showing that the rate of decay of P_N is in general no better than $\mathcal{O}(\log N/N)$.

Theorem 2.14 *Let $\sigma \in \mathcal{C}$ and $\{\lambda_N\}_{N\in\mathbb{N}}$ be a sequence such that $\lim_N \lambda_N = \infty$. Assume that $\int_{\mathbb{R}} \sigma \neq 0$. Then*

$$\liminf \frac{P_N N}{\log N} > 0,$$

where P_N is given in (2.17). In other words, P_N decays no faster than $\mathcal{O}\left(\frac{\log N}{N}\right)$.

Proof Assume otherwise. Then there exists a sequence λ_{N_k} such that $\lim_k \frac{P_{N_k} N_k}{\log N_k} = 0$. By restricting ourselves to this subsequence, we may without loss of generality assume that

$$\lim \frac{P_N N}{\log N} = 0.$$

Now set $R_N := \sum_{|n|>N} |\widehat{\sigma_{\lambda_N}}(n)|$ and $S_N := \sum_{n=-N}^{N} \frac{|\widehat{\sigma_{\lambda_N}}(n)|}{|n|+1}$. The results from Theorem 2.9 clearly imply that $\lim_N R_N = 0$. We first show that if $\lim_N R_N = 0$ if and only if $\limsup \lambda_N/N < \infty$. Let us assume otherwise, that $\limsup \lambda_N/N = \infty$. Then there is a subsequence $N_k = a_k \lambda_{N_k}$ such that $\lim_k a_k = 0$. Observe that

$$\lim_k \frac{1}{\lambda_{N_k}} \sum_{n\in\mathbb{Z}} \left|\widehat{\sigma}\left(\frac{n}{\lambda_{N_k}}\right)\right| = \int_{\mathbb{R}} |\widehat{\sigma}| = \|\widehat{\sigma}\|_1.$$

Therefore we have

$$\begin{aligned} R_{N_k} &= \frac{1}{\lambda_{N_k}} \sum_{|n|>N_k} \left|\widehat{\sigma}\left(\frac{n}{\lambda_{N_k}}\right)\right| \\ &= \frac{1}{\lambda_{N_k}} \sum_{n\in\mathbb{Z}} \left|\widehat{\sigma}\left(\frac{n}{\lambda_{N_k}}\right)\right| - \frac{1}{\lambda_{N_k}} \sum_{|n|\leq N_k} \left|\widehat{\sigma}\left(\frac{n}{\lambda_{N_k}}\right)\right| \\ &= \|\widehat{\sigma}\|_1 - \frac{1}{\lambda_{N_k}} \sum_{|n|\leq a_k\lambda_{N_k}} \left|\widehat{\sigma}\left(\frac{n}{\lambda_{N_k}}\right)\right| \\ &\geq \|\widehat{\sigma}\|_1 - a_k\|\widehat{\sigma}\|_\infty. \end{aligned}$$

It follows that $\liminf_k R_{N_k} \geq \|\widehat{\sigma}\|_1 > 0$, a contradiction. Thus $\limsup \lambda_N/N < \infty$, which means λ_N cannot grow faster than $\mathcal{O}(N)$.

Next we show that the decay rate for S_N is at best $\mathcal{O}(\log \lambda_N/\lambda_N)$. By assumption, $\widehat{\sigma}(0) = \int_{\mathbb{R}} \sigma \neq 0$. So there is a $\delta_0 > 0$ such that $|\widehat{\sigma}(t)| \geq c_0$ for some $c_0 > 0$. Now without loss of generality we may assume that $\delta_0\lambda_N \leq N$. Then

$$S_N := \frac{1}{\lambda_N} \sum_{n=-N}^{N} \frac{|\widehat{\sigma}(n\lambda_N^{-1})|}{|n|+1} \geq \frac{c_0}{\lambda_N} \sum_{|n|\leq\delta_0\lambda_N} \frac{1}{|n|+1} \geq \frac{c_1 \log \lambda_N}{\lambda_N}$$

for sufficiently large N, where $c_1 > 0$ is some constant.

Finally, since $\lambda_N = \mathcal{O}(N)$ it satisfies $\frac{\log \lambda_N}{\lambda_N} \geq c_2 \frac{\log N}{N}$ for some $c_2 > 0$. It follows that

$$\liminf_N \frac{S_N N}{\log N} \geq c_1 c_2 > 0.$$

Hence S_N must be at least $\mathcal{O}\left(\frac{\log \lambda_n}{\lambda_n}\right)$, and the theorem is proved. □

Remark 2.15 We can easily see from the proof that if $\widehat{\sigma}$ is not compactly supported then the rate of decay will be strictly slower than $\mathcal{O}(\log N/N)$. On the other hand, the $\mathcal{O}(\log N/N)$ can be achieved by choosing a suitable σ such that $\widehat{\sigma}$ is compactly supported. In this case we can in fact set $\lambda_N = b\,N$ for some appropriate $b > 0$. However, it is a challenge to construct a $\sigma \in \mathcal{C}_0$ such that $\widehat{\sigma}$ is compactly supported. There are other tradeoffs. By the Uncertainty Principle, if $\widehat{\sigma}$ is compactly supported then σ in general is less concentrated. Also, it is possible that optimal σ may not be in a closed form. This will be considered in future investigations.

3 Robustness and numerical results

In this section we present several numerical results from our edge detection method, and also analyze the effects of noise. Consider the following test function:

$$f(x) = a(x)\sin(\pi x), \quad \text{where} \quad a(x) = \begin{cases} 1 & x \in \left[0, \frac{1}{4}\right) \cup \left[\frac{1}{2}, \frac{3}{4}\right) \\ 0 & x \in \left[\frac{1}{4}, \frac{1}{2}\right) \\ -1 & x \in \left[\frac{3}{4}, 1\right]. \end{cases} \tag{3.1}$$

Figure 2 plots $f(x)$ and the real and imaginary parts of $\widehat{f}(n)$, $-200 \leq n \leq 200$.

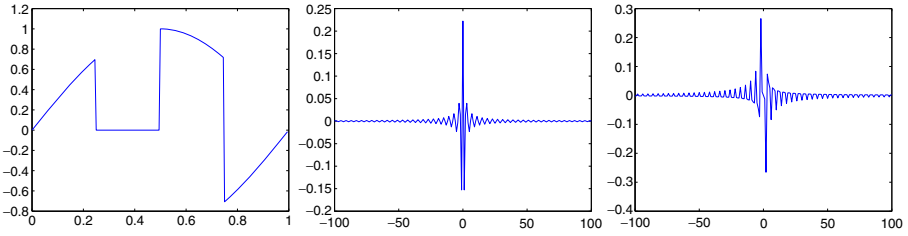

Fig. 2 (*left*) $f(x)$ (*middle*) real and (*right*) imaginary part of $\widehat{f}(n)$

We compare our edge detector, $T_N[\sigma_{\lambda_N}]$ in (2.5), to the un-mollified edge detector function, $\tilde{T}_N$ in (2.3), as well as to the concentration kernel edge detector in [6]

$$K_N[\sigma](x) = 2\pi i C_\sigma \sum_{n=-N}^{N} n\widehat{\sigma}\left(\frac{n}{N}\right)\widehat{f}(n)e^{2\pi i n x}, \tag{3.2}$$

where C_σ is the normalization constant so that $C_\sigma \int_{-1}^{1} \widehat{\sigma} = 1$ (see Remark 2.10 in Section 2). For our numerical tests we have considered several different mollifiers σ:

(1) (Gaussian) $\sigma(x) = e^{-\pi^2 x^2}$, which yields

$$\widehat{\sigma}(t) = \frac{1}{\sqrt{\pi}} e^{-t^2} \quad \text{and} \quad \widehat{\sigma_{\lambda_N}}(x) = \lambda_N^{-1}\widehat{\sigma}(\lambda_N^{-1}x) = \frac{1}{\sqrt{\pi}\lambda_N} e^{-\lambda_N^{-2}x^2}.$$

(2) $\sigma(x) = (1+4\pi^2 x^2)^{-1} \in \mathcal{C}_0$, with $\widehat{\sigma}_2(t) = \frac{1}{2}e^{-|t|}$, $\sigma(x) = (1+4\pi^2 x^2)^{-1} \in \mathcal{C}_0$.

(3) $\sigma(x)$ such that $\widehat{\sigma}(t) = c_\alpha(1+t^2)^{-\alpha}$ for $\alpha \geq 1$, where $c_\alpha = 1/\int_0^\pi \cos^{2\alpha-2}(x)dx$ is the normalization constant so that $\sigma(0) = 1$.

For the concentration kernel method we have chosen an exponential type function with compact support on $[-1, 1]$,

$$\widehat{\sigma}(t) = C_\alpha e^{\frac{1}{\alpha|t|(|t|-1)}}, \qquad t \in (-1, 1) \tag{3.3}$$

where $\alpha > 0$ is a parameter. C_α is the normalization constant so that $\int_{-1}^{1} \widehat{\sigma} = 1$. This concentration kernel has been used often because it yields fast decay away from discontinuities [6]. We set $\alpha = 4$ for our tests. Our edge detector, $T_N[\sigma_{\lambda_N}]$, with the Gaussian σ yields sharp peaks at the correct locations with the correct heights for the test function in (3.1). The unmollified edge detector, $\tilde{T}_N(x)$, shows Gibbs oscillations, and when the data size N is small the oscillations can obscure edges or yield false edges. The concentration kernel detector $K_N[\sigma]$ show marked improvement, but the Gibbs oscillations persist. In contrast, the new mollified edge detector $T_N[\sigma_{\lambda_N}](x)$ shows no Gibbs phenomenon. Figure 3 shows the plots of $\tilde{T}_N$, K_N, and $T_N[\sigma_{\lambda_N}]$ for $N = 40, 80, 160$, respectively. For $T_N[\sigma_{\lambda_N}]$ we have chosen $\lambda_N = N/\sqrt{\log N}$. We also tested our method using the other two types of mollifiers σ, which yielded comparable results.

The effect of λ_N can be seen from the plots in Fig. 4. In this example we choose $\sigma(x)$ given by its Fourier transform $\widehat{\sigma}(t) = \frac{8}{3\pi}(1+t^2)^{-3}$. For the plots we fix $N = 160$ and let $\lambda_N = 10, 30, 90$ respectively. One can see that a smaller λ_N leads to a smoother (less oscillatory) $T_N[\sigma_{\lambda_N}]$ while a larger λ_N leads to sharper peaks at the edges. Nevertheless in all three cases the locations and the heights of the peaks accurately represent the actual locations and heights of the jumps in $f(x)$. Our numerical experiments indicate that this mollifier works as well as the Gaussian mollifier. Generally $\sigma(x)$ given by $\widehat{\sigma}(t) = c_\alpha(1+t^2)^{-\alpha}$ for $\alpha \geq 1$ works well as a mollifier, where $c_\alpha = 1/\int_0^\pi \cos^{2\alpha-2}(x)dx$ is the normalization constant so that $\sigma(0) = 1$.

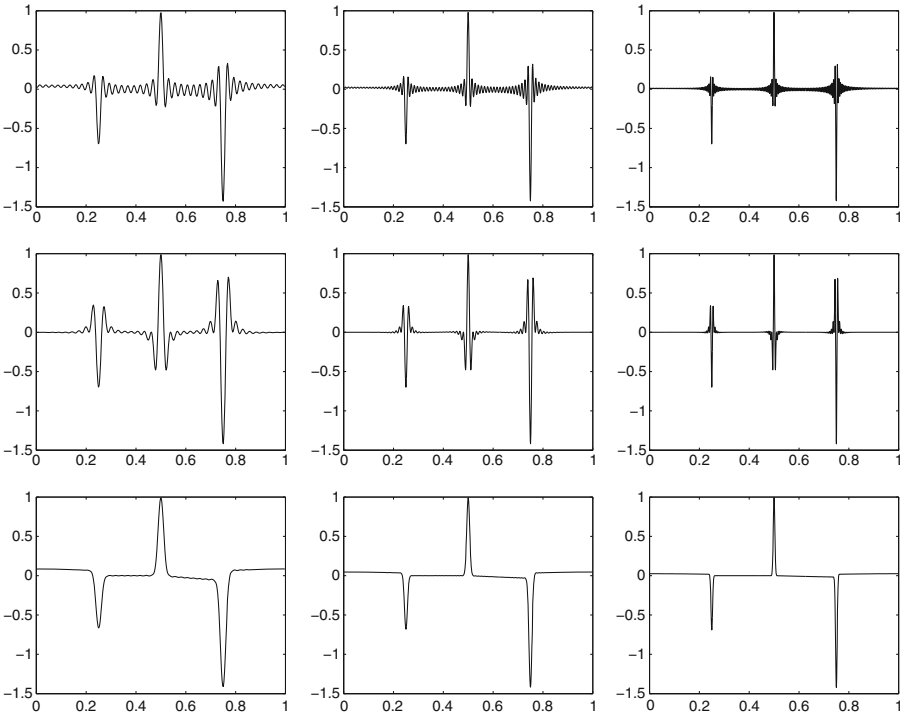

Fig. 3 *Top row*: $\tilde{T}_N(x)$ for $N = 40$, 80, and 160, respectively. *Middle row*: $K_N[\sigma](x)$ for the same N's. *Bottom row*: $T_N[\sigma_{\lambda_N}](x)$ for the same N's, with $\lambda_N = N/\sqrt{\log N}$

We now examine the effect of random noise on our method. Prior investigations [1, 3, 4, 11], demonstrated some of the challenges noisy Fourier coefficients presented for detecting edges. Several noise reduction techniques have been suggested, including designing concentration factors that specifically reduce the variance [1], or that attempt to separate the noise out of the high end frequency coefficients [3]. Another technique uses a matched filter approach [4]. Most recently, a method was designed employing statistical

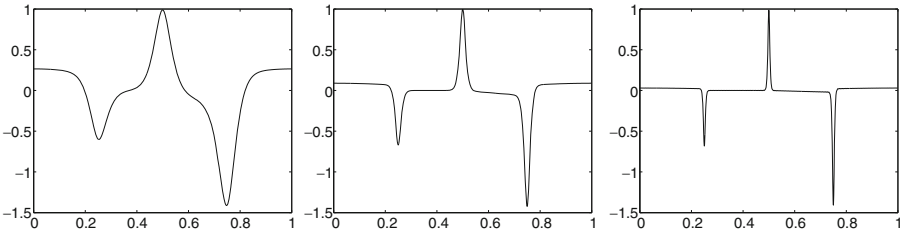

Fig. 4 $N = 160$ and $\lambda_N = 10$, 30, and 90, respectively

hypothesis testing [11]. Our mollifier approach evidently provides a more natural antidote to noisy Fourier data.

Suppose that the given Fourier data $\{\widehat{f}(n)\}_{n=-N}^{N}$ is corrupted by an additive noise ε_n, i.e. we are given $z_n = \widehat{f}(n) + \varepsilon_n$ instead of $\widehat{f}(n)$ for $-N \le n \le N$. We shall assume that the noise $\{\varepsilon_n\}$ are complex i.i.d. random variables with mean 0 and standard deviation δ.

Since $\widehat{f}(n)$ decays at the rate $\mathcal{O}(\frac{1}{n})$, for large n the noise ε_n can exceed and even overwhelm $\widehat{f}(n)$, rendering the data z_n useless. As a result there is an effective limit on the size of the Fourier data that can be meaningfully used. By (2.2), $\widehat{f}(n)$ is approximately bounded by $\frac{M}{2\pi n}$ for large n, where M is the size of the largest jump of f. Thus for n much larger than $\frac{M}{2\pi\delta}$, the noise ε_n will dominate the signal $\widehat{f}(n)$. In this case it may be better to simply discard those $\widehat{f}(n)$.

Our edge detection method, (2.5), applied to noisy data $\{z_n\}$ yields

$$T_N[\sigma_{\lambda_N}](x) = 2\pi i \sum_{n=-N}^{N} n\widehat{\sigma_{\lambda_N}}(n)\,\widehat{f}(n)e^{2\pi inx} + 2\pi i \sum_{n=-N}^{N} n\varepsilon_n\widehat{\sigma_{\lambda_N}}(n)e^{2\pi inx}.$$

By comparison, the un-mollified edge detector function, $\tilde{T}_N$ in (2.3), applied to noisy data gives

$$\tilde{T}_N(x) = \frac{2\pi i}{2N+1}\left(\sum_{n=-N}^{N} n\,\widehat{f}(n)e^{2\pi inx} + \sum_{n=-N}^{N} n\varepsilon_n e^{2\pi inx}\right).$$

The corresponding random noise contribution to the error terms are

$$E_N[\sigma_{\lambda_N}](x) := 2\pi i \sum_{n=-N}^{N} n\varepsilon_n\widehat{\sigma_{\lambda_N}}(n)e^{2\pi inx} \quad \text{and} \quad \tilde{E}_N(x) := \frac{2\pi i}{2N+1}\sum_{n=-N}^{N} n\varepsilon_n e^{2\pi inx}. \tag{3.4}$$

To understand the impact of noise $\{\varepsilon_n\}$ on edge detection, we must analyze $E_N[\sigma_{\lambda_N}](x)$, $\tilde{E}_N(x)$. Note that for the concentration kernel method the error is simply $E_N[\sigma_{\lambda_N}](x)$ with $\lambda_N = N$. Clearly both errors have 0 mean. The variances of the methods are quite different, however, with

$$\begin{aligned}\text{Var}\left(\tilde{E}_N(x)\right) &= \frac{4\pi^2}{(2N+1)^2}\sum_{n=-N}^{N}\text{Var}\left(n\varepsilon_n e^{2\pi inx}\right)\\ &= \frac{4\pi^2\delta^2}{(2N+1)^2}\sum_{n=-N}^{N} n^2\\ &\ge \frac{2\pi^2\delta^2 N}{3}.\end{aligned}$$

and, with $\widehat{\sigma_{\lambda_N}}(x) = \lambda_N^{-1}\widehat{\sigma}(\lambda_N^{-1}x)$,

$$\begin{aligned}\text{Var}\left(E_N[\sigma_{\lambda_N}](x)\right) &= 4\pi^2 \sum_{n=-N}^{N} \text{Var}\left(n\varepsilon_n \widehat{\sigma_{\lambda_N}}(n) e^{2\pi i n x}\right) \\ &= \frac{4\pi^2\delta^2}{\lambda_N^2} \sum_{n=-N}^{N} n^2 |\widehat{\sigma}(\lambda_N^{-1}n)|^2.\end{aligned}$$

Note that when λ_N is large the sum $\lambda_N^{-3}\sum_{n=-N}^{N} n^2|\widehat{\sigma}(\lambda_N^{-1}n)|^2$ can be approximated by $\int_{-b_N}^{b_N} t^2|\widehat{\sigma}(t)|^2\,dt$ where $b_N = N/\lambda_N$. In the case of concentration kernels where $\lambda_N = N$ we have

$$\text{Var}\left(E_N[\sigma_N](x)\right) \approx \frac{4\pi^2\delta^2}{N}\int_{-1}^{1} t^2|\widehat{\sigma}(t)|^2\,dt.$$

So the error is in the same order as for the unmollified edge detector. In our method we have flexibility in the choice of λ_N, and in general $\lambda_N = o(N)$. Thus it typically performs better under additive noise.

Lemma 3.1 *Assume that $g(t) \geq 0$ is increasing on $[0, a]$ and decreasing on (a, ∞). Then*

$$\sum_{n=1}^{\infty} g\left(\frac{n}{\lambda}\right) \leq \lambda \int_{1/\lambda}^{\infty} g(t)dt + g(a).$$

Proof Assume that $K/\lambda \leq a < (K+1)/\lambda$. By standard integral estimation technique we have

$$\sum_{n=1}^{K} g\left(\frac{n}{\lambda}\right) \leq \lambda \int_{\frac{1}{\lambda}}^{\frac{K+1}{\lambda}} g(t)dt,$$

and

$$\sum_{n=K+1}^{\infty} g\left(\frac{n}{\lambda}\right) \leq \lambda \int_{\frac{K}{\lambda}}^{\infty} g(t)dt.$$

It follows that

$$\sum_{n=1}^{\infty} g\left(\frac{n}{\lambda}\right) \leq \lambda \int_{\frac{1}{\lambda}}^{\infty} g(t)dt + \lambda \int_{\frac{K}{\lambda}}^{\frac{K+1}{\lambda}} g(t)dt \leq \lambda \int_{\frac{1}{\lambda}}^{\infty} g(t)dt + g(a).$$

□

Corollary 3.2 *Assume that $\sigma \in \mathcal{C}$ and $t^2|\widehat{\sigma}(t)|^2$ has exactly two local maxima on $\mathbb{R}$. Then for $\widehat{\sigma_{\lambda_N}}(x) = \lambda_N^{-1}\widehat{\sigma}(\lambda_N^{-1}x)$ we have*

$$\text{Var}\left(E_N[\sigma_{\lambda_N}](x)\right) \leq 4\pi^2\delta^2\left(\lambda_N \int_{\mathbb{R}} t^2|\widehat{\sigma}(t)|^2\,dt + 2\sup_t t^2|\widehat{\sigma}(t)|^2\right). \tag{3.5}$$

Corollary 3.2 is a straightforward application of Lemma 3.1, and we omit the proof here.

Using (3.5) we can now obtain bounds on the variance of $T_N[\sigma_{\lambda_N}]$ for the mollifiers used in this paper:

(A) For the Gaussian mollifier given by $\sigma(x) = e^{-\pi^2 x^2}$ and $\widehat{\sigma}(t) = \frac{1}{\sqrt{\pi}} e^{-t^2}$ the bound for Var $\left(E_N[\sigma_{\lambda_N}](x)\right)$ is

$$\text{Var}\left(E_N[\sigma_{\lambda_N}](x)\right) \leq \pi^{\frac{3}{2}} \delta^2 \lambda_N + \frac{4\pi}{e} \delta^2. \tag{3.6}$$

(B) For the mollifier σ given by $\widehat{\sigma}(t) = \frac{8}{3\pi}(1+t^2)^{-3}$ the bound for Var $\left(E_N[\sigma_{\lambda_N}](x)\right)$ is

$$\text{Var}\left(E_N[\sigma_{\lambda_N}](x)\right) \leq \frac{7}{9} \delta^2 \lambda_N + 2\delta^2. \tag{3.7}$$

The advantage of using $T_N[\sigma_{\lambda_N}]$ as an edge detector is that we can control the impact of the random noise to a certain degree by controlling λ_N and $\int_{\mathbb{R}} t^2 |\widehat{\sigma}(t)|^2 \, dt$. The trade-off is that by taking a smaller λ_N and/or $\int_{\mathbb{R}} t^2 |\widehat{\sigma}(t)|^2 \, dt$ we are choosing a smoother edge detector. But as we have seen from previous examples, even a smoother edge detector $T_N[\sigma_{\lambda_N}]$ yields accurate locations and heights of the edges. Figure 5 plots the edge detectors K_N and $T_N[\sigma_{\lambda_N}]$ under the effect of additive noise, where in both cases we take σ to be Gaussian (Again, the other mollifiers yield very similar results). The data size is $N = 300$ and we fix $\lambda_N = 40$. We add complex white Gaussian noise of standard

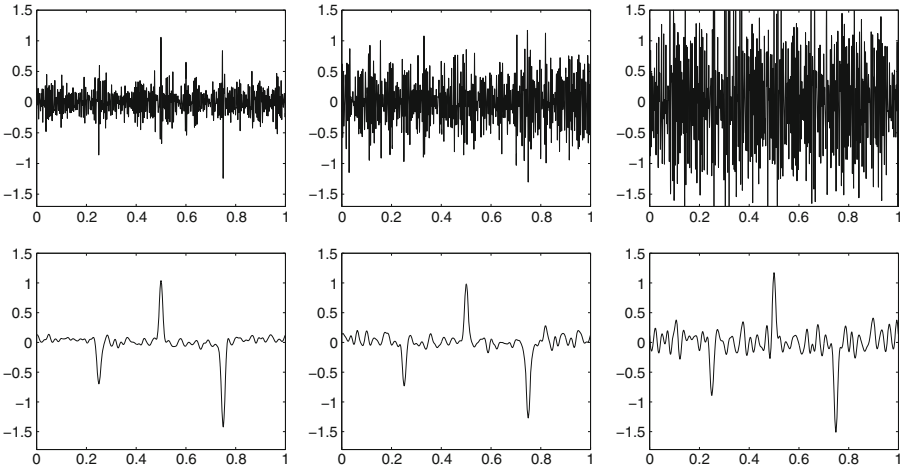

Fig. 5 *Top*: $K_N[\sigma](x)$ with additive Gaussian noise level $\delta = 0.005, 0.01$ and 0.02, respectively. *Bottom*: $T_N[\sigma_{\lambda_N}](x)$ with $\lambda_N = N/\log^{1.3} N$ and same noise levels. The data size is $N = 300$ for all plots

deviation $\delta = 0.005,\ 0.1,\ 0.2$, and 0.3, respectively, which correspond to noise in decibels (dB) and signal-to-noise ratio (SNR) shown below:

Standard dev. δ	Noise in dB	Signal-to-noise ratio
0.005	0.28 dB	12.86 dB
0.01	0.74 dB	7.25 dB
0.02	2.56 dB	1.09 dB

The first three plots show the concentration kernel edge detector $K_N[\sigma]$ under additive Gaussian white noise with $\delta = 0.005, 0.01$ and 0.02, respectively. Oscillations already obscure the results when $\delta = 0.005$, and when $\delta = 0.02$, the results of $K_N[\sigma]$ are indistinguishable from the noise. On the other hand, the peaks from the edges still show up conspicuously in $T_N[\sigma_{\lambda_N}]$, even for $\delta = 0.02$. With suitable thresholding or other techniques, such as the sparsity enforcing technique designed in [13], strong edges can easily be identified in a signal.

Figure 6 demonstrates our method on real data, which is given as a line slice in the famous Lena image, and normalized to domain $[0, 1]$ with $0 \leq f(x) \leq 1$. We compute $\{\widehat{f}(n)\}_{n=-N}^{N}$ for $N = 500$. White complex Gaussian noise of standard deviation $\delta = 0.005$ (0.32 dB, SNR = 11.82 dB) and 0.01 (0.98 dB, SNR = 6.19 dB) are added to $\widehat{f}(n)$, respectively. In this example we apply the mollifier $\widehat{\sigma}(x) = \frac{8}{3\pi^2}(1 + x^2)^{-3}$ (The Gaussian mollifier yields very similar results). Observe that $f(x)$ contains two edges near $x = 0.5$ that are very close. Some of these jump discontinuities may be seen as steep gradients.

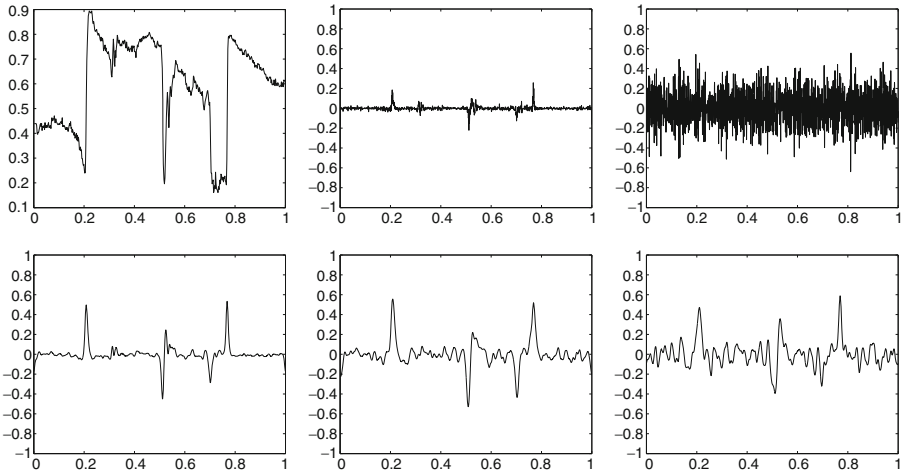

Fig. 6 *Top left*: the data f. *Top middle and right*: concentration kernel edge detection with no noise and additive Gaussian noise $\delta = 0.005$, respectively. *Bottom*: $T_N[\sigma_{\lambda_N}](x)$ with additive Gaussian noise of $\delta = 0$, 0.005, and 0.01, respectively

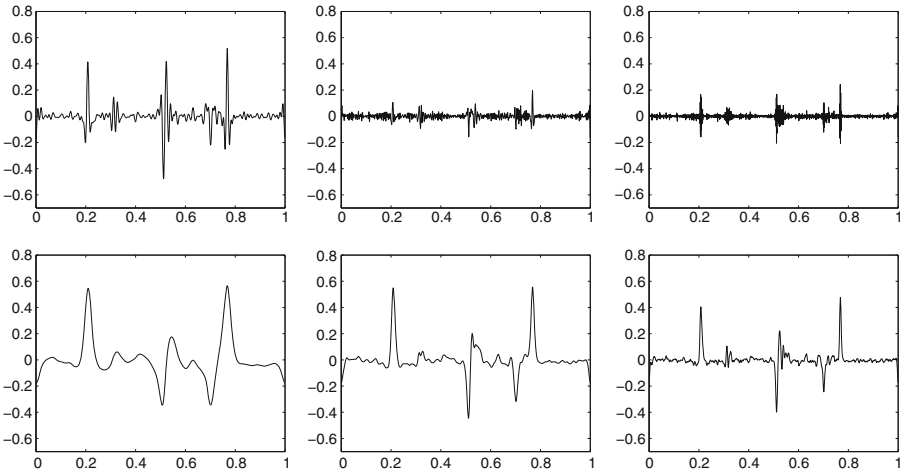

Fig. 7 *Top*: $K_N[\sigma](x)$ with $N = 100, 300, 900$, respectively. *Bottom*: $T_N[\sigma_{\lambda_N}](x)$ with $N = 100, 300, 900$, respectively and $\lambda_N = N/\log N$

In this case, the high order concentration kernel $K_N[\sigma]$, may actually be *suppressing* these values, since it is trying to discern true edges from steep gradients. Moreover, when the edges are close together, the oscillatory kernels cause interfering oscillations between edges, yielding incorrect heights and multiple peaks. This occurs even when no noise is added. With additive noise $\delta = 0.005$, the kernel $K_N[\sigma]$ returns a noisy meaningless profile. In contrast, even with 0.98 dB noise added to $\{\widehat{f}(n)\}$, Fig. 6 (bottom) suggests that edges can be successfully extracted from the calculation of $T_N[\sigma_{\lambda_N}]$ with $\lambda_N = N/\log^{1.3} N$. An interesting fact is that higher resolution (i.e. more Fourier data) does not necessarily yield better results for both methods, particularly for the concentration kernel method. For some applications, such as when the underlying function is piecewise smooth with high variation between the edges, the high order concentration kernels yield substantially better results for smaller N. This is particularly useful in post-processing numerical solutions for partial differential equations that admit shocks, for instance. Figure 7 shows the impact as we increase N from $N = 100$ to $N = 900$. With $T_N[\sigma_{\lambda_N}]$ (bottom row), where we choose $\lambda_N = N/\log N$, increased resolution forces the edge detector to capture finer edges. This isn't necessarily a good thing since in this example the function f itself is digitized so it is a step function. In theory there is a discontinuity between just about any two neighboring pixels. Therefore most of these small peaks do not represent true edges. At $N = 100$ only the significant edges are captured. As we increase the resolution the edge detector picks up more small oscillations. Nevertheless, strong edges stand out in the process. The concentration kernel method $K_N[\sigma]$ using σ given in (3.3) with $\alpha = 4$ (top row) picks up more smaller edges. However, for the same reasons as stated in reference to Fig. 6, the performance deteriorated with increased resolution.

4 Concluding remarks

In this study we introduced the notion of *uniform convergence to sharp peaks* for a piecewise smooth function given a finite number of Fourier coefficients. Our new method, based on a spectral mollifier, converges to zero away from the jump discontinuities *without* introducing oscillations near them. There are several important advantages to our method. Specifically, it is more amenable to detecting edges when there is noise in the given Fourier data since there are no artificial oscillations already occurring in the approximation. Furthermore, when there are multiple edges close proximity and there is not enough Fourier coefficents to resolve them sufficiently, our method is able to extract them. By contrast, previous edge detection methods based on Fourier coefficients have difficulties in these situations because of the oscillations that are produced near the jumps. These oscillations tend to bleed into neighboring jumps.

Our method is straightforward to apply, as an FFT can be used directly on the data. Moreover, the convergence rate can be improved using iterative techniques such as those that enforce sparsity or other types of regularization [13].

Finally, our technique may be used to determine other features in the physical domain from Fourier data, such as pulses or square waves. This will be considered in future investigations.

References

1. Archibald, R., Gelb, A.: Reducing the effects of noise in image reconstruction. J. Sci. Comput. **17**(1–4), 167–180 (2002)
2. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **8**, 679–698 (1986)
3. Engelberg, S., Tadmor, E.: Recovery of edges from spectral data with noise—a new perspective. SIAM J. Numer. Anal. **46**(5), 2620–2635 (2008)
4. Gelb, A., Cates, D.: Detection of edges in spectral data III—improvements in the presence of noise. J. Sci. Comput. **36**(1), 1–43 (2008)
5. Gelb, A., Tadmor, E.: Detection of edges in spectral data. Appl. Comput. Harmon. Anal. **7**, 101–135 (1999)
6. Gelb, A., Tadmor, E.: Detection of edges in spectral data II. Nonlinear enhancement. SIAM J. Numer. Anal. **38**(4), 1389–1408 (2000)
7. Gelb, A., Tadmor, E.: Adaptive edge detectors for piecewise smooth dara based on the minmod limiter. J. Sci. Comput. **28**(2–3), 279–306 (2006)
8. Gottlieb, D., Shu, C.W.: On the Gibbs phenomenon and its resolution. SIAM Rev. **39**(4), 644–668 (1997)
9. Hesthaven, J.S., Gottlieb, S., Gottlieb, D.: Spectral Methods for Time-dependent Problems. Cambridge University Press (2007)
10. Shu, C.-W.: Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. Advanced Numerical Approximation of Nonlinear Hyperbolic Equations, vol. 1687 of Lecture Notes in Mathematics, pp. 325–432 (1998)
11. Petersen, A., Gelb, A., Eubank, R.: Hypothesis testing for Fourier based edge detection methods. J. Sci. Comput. (2011, accepted)
12. Sobel, I.: An isotropic 3×3 image gradient operator. In: Freeman, H. (ed.) Machine Vision for Three-dimensional Scenes. Academic Press, Boston (1990)

13. Stefan, W., Viswanathan, A., Gelb, A., Renaut, R.: Sparsity enforcing edge detection for blurred and noisy Fourier data. (2011, submitted)
14. Stein, E., Weiss, G.: Introduction to Fourier Analysis on Euclidean Spaces. Princeton University Press, Princeton (1971)
15. Tadmor, E.: Filters, mollifiers and the computation of the Gibbs phenomenon. Acta Numer. **16**, 305–378 (2007)