Laplace deconvolution with noisy observations

Felix Abramovich, Marianna Pensky and Yves Rozenholc

Tel Aviv University, University of Central Florida
and Université Paris, Descartes

Abstract

In the present paper we consider Laplace deconvolution on the basis of discrete noisy data observed on the interval which length may increase with a sample size. Although this problem arises in a variety of applications, to the best of our knowledge, it has not been systematically studied in statistical literature and the present paper contributes to fill this gap. We derive an adaptive kernel estimator of the function of interest, and establish its asymptotic minimaxity over a range of Sobolev classes. A limited simulation study shows that, in addition to providing theoretical asymptotic results, the proposed Laplace deconvolution estimator demonstrates good performance in finite sample examples.

AMS 2010 subject classifications. 62G05, 62G20.

Key words and phrases: adaptivity, kernel estimation, minimax rates, Volterra equation, Laplace convolution

1 Introduction

Mathematical modeling of a variety of problems in population dynamics, mathematical physics, theory of superfluidity and many others leads to the convolution type Volterra equation of the first kind of the form

$$q(t) = \int_0^t g(t-\tau) f(\tau) d\tau, \ t \geq 0 \tag{1.1}$$

where $q(\cdot)$ is the known (observed) function, $g(\cdot)$ is the (known) kernel and $f(\cdot)$ is the unknown function to be solved for (see, e.g. Gripenberg, Londen & Steffans, 1990). Two motivating examples from computed tomography and fluorescence spectroscopy are described below. This problem is also known as Laplace deconvolution problem.

In practice, however, one typically observes discrete data on a finite interval that, in addition, is corrupted by noise which leads to the following discrete noisy version of equation (1.1):

$$y(t_i) = \int_0^{t_i} g(t_i - \tau) f(\tau) d\tau + \sigma \epsilon_i, \quad i = 1, \cdots, n, \tag{1.2}$$

where $0 \leq t_1 \leq ... \leq t_n \leq T_n$, ϵ_i are i.i.d. $N(0,1)$ and T_n may grow with n.

Example 1: Dynamic contrast enhanced computed tomography data. Dynamic Contrast Enhanced Imaging (DCE-imaging) is widely used in medical imaging of brain structures or cancerous tumors (see, e.g., Cao *et al.*, 2010; Goh *et al.*, 2005; Goh and Padhani, 2007; Cuenod *et al.*, 2006; Miles, 2003; Padhani and Harvey, 2005 and Bisdas *et al.*, 2007). DCE-imaging has great potential for cancer detection and characterization, as well as for monitoring *in vivo* the effects of treatments. The experiment follows the evolution of a bolus of contrast agent injected during sequential imaging acquisition.

The data is assumed to consist of observations of quantities (concentrations) of contrast agents at voxels of unit volumes measured at different times:

$$Q_x(t_k) = Q_x^0(t_k) + \sigma \epsilon_x(k), \quad 0 = t_0 < t_1 < \cdots < t_n = T, \tag{1.3}$$

where $Q_x(t_k)$ and $Q_x^0(t)$ are respectively the observed and the true (unknown) quantities of a contrast agent at time t_k in the voxel x and $\epsilon_x(k), k = 1, \cdots, n$, are i.i.d. standard normal variables.

The total amount of the contrast agent $A_x(t)$ arrived into voxel x by the time t is the so called arterial input function which is usually known. If $D_x(t)$ is the amount of contrast agent which has departed by time t from voxel x and S_x is the random lapse of time during which a molecule of contrast agent stays in the voxel x, then $D_x(t)$ can be presented as a Laplace convolution of the density $\alpha_x(s)$ of the rate of arrivals of the contrast agent into the voxel x and unknown function $f_x(t) = P(S_x \geq t)$:

$$D_x(t) = \int_0^t \alpha_x(s) P(S_x \geq t - s) ds = \int_0^t \alpha_x(t - s) f_x(s) ds. \tag{1.4}$$

The function $\alpha_x(t)$ depends on the voxel x only through the voxel dependent factor and, therefore, it cannot be used to describe a particular tissue type. On the contrary, S_x and, hence, function $f_x(t)$ depends on the properties of the tissue and, hence, can be used for its characterization. In order to connect equation (1.4) with the observations $Q_x(t_k)$ in (1.3), note that $D_x(t) = A_x(t) - Q_x^0(t)$ where $A_x(t)$ is known, so that, equation (1.4) can be viewed as a particular case of problem (1.2).

Example 2: Fluorescence spectroscopy data. Equation (1.2) has been extensively used for modeling of time-resolved measurements in fluorescence spectroscopy, particularly, for studies of biological macromolecules and for cellular imaging (see, e.g., Ameloot and Hendrickx, 1983; Ameloot *et al.*, 1984; Gafni, Modlin and Brand, 1975; McKinnon, Szabo and Miller, 1977;

O'Connor, Ware and Andre, 1979, and also the monograph of Lakowicz, 2006 and references therein).

At present, in fluorescence spectroscopy, most of the time-domain measurements are carried out using time-correlated single-photon counting. The measured intensity decay is represented by $N(t_k)$, the number of photons that were detected within the time interval $(t_k, t_k + \Delta t)$ and appears as a convolution of the response function $I(t)$ with the lamp function $L(t)$. One can imagine the excitation pulse to be a series of δ-functions with different amplitudes. Each δ-function excitation is assumed to excite an impulse response $L(t_k)I(t - t_k)\Delta t$ at time $t > t_k$, with the amplitude at time t_k proportional to the excitation intensity $L(t_k)$. The measured decay $N(t)$ is the sum of the impulse responses created by all the individual δ-function excitation pulses occurring until time t: $N(t) = \sum_{t_k=0}^{t_k=t} L(t_k)I(t - t_k)\Delta t$. As $\Delta t \to \infty$, the sum in the right hand side can be replaced by the integral, and with a change of variables $t - s = x$, the last equation can be written as $N(t) = \int_0^t L(t - x)I(x)dx$. The experimental data come in the form of measurements $\hat{N}(t_k)$ which are contaminated by random noise and, therefore, can be modeled by equation (1.2). The objective is to determine the impulse response function $I(x)$ that best matches the experimental data.

Formally, by setting $g(t) = f(t) \equiv 0$ for $t < 0$, equation (1.1) can be viewed as a particular case of the Fredholm convolution equation

$$h(t) = \int_{-\infty}^{\infty} g(t - \tau)f(\tau)d\tau, \tag{1.5}$$

whose discrete stochastic version

$$y(t_i) = \int_a^b g(t_i - \tau)f(\tau)d\tau + \sigma\epsilon_i, \quad i = 1, \cdots, n, \tag{1.6}$$

known also as Fourier deconvolution problem, has been extensively studied in the last thirty years (see, for example, Carroll and Hall, 1988; Delaigle, Hall and Meister, 2008; Diggle and Hall, 1993; Fan, 1991; Fan and Koo, 2002; Johnstone *et al.*, 2004; Pensky and Vidakovic, 1999; Stefanski and Carrol, 1990 among others)

However, such an approach to solving (1.1) and (1.2) is very misleading. In fact, Gripenberg, Londen & Steffans (1990, p. 3) state that "much of the classical theory of Fredholm equations reduces to mere trivialities when applied to Volterra equations. On the other hand, Volterra equations exhibit a variety of phenomena unknown to Fredholm theory." In particular, artificial zero extension of g and f for negative values of the argument evidently affects their regularity at zero. In addition, note that the measurements of the right-hand side of (1.1) are available only on the interval $[0, T_n]$ which makes application of usual discrete Fourier transform typically applied for solving (1.5) impossible since the latter would assume periodicity of a function on $[0, T_n]$. Moreover, for the noisy measurements in (1.5), the solution by Fourier transform may not vanish for $t < 0$ and, as a result, may be different from a true solution f of the equation.

The mathematical theory of (noiseless) convolution type Volterra equations is well developed (see, e.g., Gripenberg, Londen and Staffans, 1990) and the exact solution of (1.1) can be obtained through Laplace transform. However, direct application of Laplace transform for discrete measurements faces serious conceptual and numerical problems. The inverse Laplace transform is usually found by application of tables of inverse Laplace transforms, partial fraction decomposition or series expansion (see, e.g., Polyanin and Manzhirov, 1998), neither of which is applicable in the case of the noisy version of Laplace deconvolution. Only few applied mathematicians took an effort to solve the problem using discrete measurements in the LHS of (1.5) (see, e.g., Ameloot and Hendrickx, 1983; Cinzori and Lamm, 2000; Lamm, 1996; Lien *et al.*, 2008; Maleknejad *et al.*, 2007; Rashed, 2003; Weeks, 1966). Ameloot and Hendrickx (1983) applied Laplace deconvolution for the analysis of fluorescence curves and used a parametric presentation of the solution f as a sum of exponential functions with parameters evaluated by minimizing discrepancy with the right-hand side. In a somewhat similar manner, Maleknejad *et al.* (2007) proposed to expand the unknown solution over a wavelet basis and find the coefficients via the least squares algorithm. Lien *et al.* (2008), following Weeks (1966), studied numerical inversion of the Laplace transform using Laguerre functions. Finally, Cinzori and Lamm (2000) and Lamm (1996) used discretization of the equation (1.1) and applied various versions of the Tikhonov regularization technique. However, in all of the above papers, the noise in the measurements was either ignored or treated as deterministic. The presence of random noise in (1.2) makes the problem even more challenging. To the best of our knowledge, the only paper which considered inverse Laplace transform on the basis of random noisy measurements is Chauveau, van Rooij and Ruymgaart (1994) who studied this problem in the framework of mixture density estimation.

For all these reasons, estimation of function f from noisy observations y in (1.2) requires development of a novel approach. Unlike Fourier deconvolution that has been intensively studied in statistical literature (see references above), Laplace deconvolution received virtually no attention within statistical framework. We can mention Dey, Martin & Ruymgaart (1998) that considered the model essentially equivalent to (1.2) with the exponential kernel $g(t) = be^{-at}$. They proposed an estimator for this very specific type of kernel and derived the convergence rate for its quadratic risk as n increases, where the r-th derivative of f is assumed to be continuous on $(0, \infty)$.

The goal of the present paper is to investigate a general Laplace deconvolution problem (1.2). In Section 2, we re-formulate some known relevant mathematical results for Volterra equations to make them suitable for developing the novel statistical approach for Laplace deconvolution with noisy data measured on an interval. In Section 3, we construct an adaptive estimator of the unknown function f in (1.2) and establish its asymptotic optimality in minimax sense over the entire class of Sobolev balls. In Section 4, we present the results of a limited simulation study of the developed estimator in the paper. Section 5 concludes the paper with discussion. All the proofs are given in the Appendix.

2 Convolution type Volterra equations

To construct an estimator $\hat{f}_n$ for the unknown f in (1.2) we start from the noiseless Volterra equation (1.1) and find its exact solution.

We first introduce several notations used throughout the paper. We denote the Laplace transform of a function $f(t)$ by $\tilde{F}(s)$, that is, $\tilde{F}(s) = \int_0^\infty e^{-st} f(t) dt$. The $L_k(\mathbb{R}^+)$-norm of the function h is denoted by $\|h\|_k$ and $\|h\|_\infty$ is the supremum norm of h. If $k = 2$ and there is no ambiguity, we shall omit the subscript in the notation of the norm, i.e. $\|h\| = \|h\|_2$. We use the standard notation $W^{s,p}(\mathbb{R}^+)$ for a Sobolev space of functions on $[0, \infty)$ that have s derivatives with finite L_p-norms and, in particular, for $p = 2$, $H^s(\mathbb{R}^+) = W^{s,2}(\mathbb{R}^+)$. In what follows, we shall omit $\mathbb{R}^+$ in the notations of the norms and functional spaces and, unless the opposite is stated and assume that all functions are defined on the nonnegative part of the real line.

Assume now the following conditions on the unknown f and the (known) kernel g in (1.1):

(A1) $f \in H^m$

(A2) There exists an integer $1 \le r < m$ such that

$$g^{(j)}(0) = \begin{cases} 0, & \text{if } \ j = 0, \cdots, r-2 \\ B_r \neq 0, & \text{if } \ j = r-1 \end{cases} \tag{2.1}$$

(A3) $g \in W^{r,1} \cap H^u$ for some $u \ge m + r - 1$ (hence, in particular, $u > 2r - 1$)

Taking derivatives of both sides of (1.1) under Assumption (A2), one obtains

$$\begin{aligned} q^{(j)}(t) &= \int_0^t g^{(j)}(t-\tau) f(\tau) d\tau, \quad j = 1, \cdots, r-1; \\ &\cdots \\ q^{(r)}(t) &= B_r f(t) + \int_0^t g^{(r)}(t-\tau) f(\tau) d\tau, \end{aligned} \tag{2.2}$$

Keeping differentiating q, (2.2) yields

$$\begin{aligned} q^{(r+1)}(t) &= B_r f'(t) + g^{(r)}(t) f(0) + \int_0^t g^{(r)}(\tau) f'(t-\tau) d\tau \\ &\cdots \\ q^{(r+m)}(t) &= B_r f^{(m)}(t) + \sum_{j=0}^{m-1} g^{(r+j)}(t) f^{(j)}(0) + \int_0^t g^{(r)}(\tau) f^{(m)}(t-\tau) d\tau \end{aligned}$$

Then, under Assumptions (A1) and (A3), $q^{(r+m)} \in L_2$ and, hence, $q \in H^{(r+m)}$.

In addition, due to Assumptions (A1) and (A3), (2.2) implies that $q^{(r)}, g^{(r)} \in L_1$ and, therefore, we can use the following known facts from the theory of Volterra equations:

1. there exists a unique solution ϕ of the equation

$$g^{(r)}(t) = B_r\phi(t) + \int_0^t g^{(r)}(t-\tau)\phi(\tau)d\tau \tag{2.3}$$

called a *resolvent* of $g^{(r)}$ (see Theorem 3.1 of Gripenberg, Londen & Staffans, 1990);

2. there exists a unique solution of (1.1) which can be written as

$$f(t) = B_r^{-1}q^{(r)}(t) - B_r^{-1}\int_0^t q^{(r)}(t-\tau)\phi(\tau)d\tau \tag{2.4}$$

(see Theorem 3.5 of Gripenberg, Londen & Staffans, 1990).

Therefore, to solve (1.1) one only needs to determine a resolvent ϕ in (2.3) defined entirely by $g^{(r)}$. We find ϕ using Laplace transform. Taking Laplace transform of both sides of (2.3) yields

$$\widetilde{G^{(r)}}(s) = B_r\tilde{\Phi}(s) + \widetilde{G^{(r)}}(s)\tilde{\Phi}(s)$$

where, due to Assumption (A2), $\widetilde{G^{(r)}}(s) = s^r\tilde{G}(s) - B_r$, and, therefore,

$$\tilde{\Phi}(s) = \frac{s^r\tilde{G}(s) - B_r}{s^r\tilde{G}(s)} \tag{2.5}$$

Behavior of the resolvent function ϕ is thus determined by the properties of the Laplace transform $\tilde{G}$ of the kernel g. Under Assumption (A2), $\tilde{G}$ is analytic, so all its zeros are well separated. However, if $\tilde{G}$ has zeros with positive real parts, the resulting resolvent $\phi(t)$ becomes *unstable*: it grows exponentially as $t \to \infty$. To avoid this potentially very treacherous situation, we impose an additional condition on $\tilde{G}$:

(A4) Let Ω be a collection of distinct zeros s_ω of $\tilde{G}$. Then $s^* = \max_{\omega\in\Omega} Re(s_\omega) < 0$.

Under Assumption (A4), the theory developed in Gripenberg, Londen & Steffans (1990, Chapter 7) leads to the following result:

Theorem 1. *Let Assumption (A4) hold. Then, the resolvent ϕ in (2.3) is of the form*

$$\phi(t) = \sum_{j=0}^{r-1}\frac{a_j}{j!}t^j + \phi_1(t), \tag{2.6}$$

where $\phi_1 \in L_1$, and, hence, from (2.4), f can be recovered as

$$f(t) = B_r^{-1}\left(q^{(r)}(t) - \sum_{j=0}^{r-1} a_{r-1-j}q^{(j)}(t) - \int_0^t q^{(r)}(t-\tau)\phi_1(\tau)d\tau\right). \tag{2.7}$$

In majority of situations, the number of zeros of $\tilde{G}$ is finite and, since it is an analytic function, these zeros are of finite orders. In this case, the solution f of (1.1) in (2.7) can be written explicitly:

Theorem 2. *Let f and g satisfy assumptions (A1)–(A4). Let $\tilde{G}(s)$ have M distinct zeros $s_l < 0$ of orders α_l, respectively, $l = 1, \cdots, M$. Set $s_0 = 0$ and $\alpha_0 = r$. Then, f is of the form*

$$f(t) = B_r^{-1} \left(q^{(r)}(t) - f_0(t) - f_1(t) \right), \tag{2.8}$$

where $f_0(t) = \sum_{j=0}^{r-1} a_{0,r-1-j} q^{(j)}(t)$, $f_1(t) = \int_0^t q^{(r)}(t-\tau)\phi_1(\tau)d\tau$ and

$$\phi_1(t) = \sum_{l=1}^{M} \sum_{j=0}^{\alpha_l - 1} \frac{a_{lj}}{j!} t^j e^{s_l t}, \tag{2.9}$$

$$a_{lj} = \frac{1}{(\alpha_l - 1 - j)!} \frac{d^{\alpha_l - j - 1}}{ds^{\alpha_l - j - 1}} \left[(s - s_l)^{\alpha_l} \tilde{\Phi}(s) \right] \Bigg|_{s = s_l}. \tag{2.10}$$

Note that Assumption (A4) implies $s_l < 0$, $l = 1, \cdots, M$, so that the function $\phi_1(x)$ is a sum of products of polynomials and exponentials with negative powers and, hence, $\phi_1 \in L_1 \cap L_2$.

3 Laplace deconvolution in the presence of noise

3.1 Preamble

We return now to the original problem of recovering an unknown function f from a noisy version of Volterra equation (1.2). We assume that the convolution kernel g is known and satisfies Assumptions (A2)–(A4) while f satisfies Assumption (A1). Our goal is to construct an estimator $\hat{f}_n$ of f from the noisy data $y(t_i)$, $i = 1, ..., n$.

The precision of estimating f by $\hat{f}_n$ is measured by the L_2 risk $E\|\hat{f}_n - f\|^2_{[0,T_n]}$. In particular, we shall be interested in the rate of its convergence as n increases. Note that the proposed setup assumes that both the length of the interval T_n, where the data is observed, and the data density per unit interval may increase as the sample size tends to infinity.

As we have mentioned, the resolvent ϕ of $g^{(r)}$ in (2.3) is completely determined by the (known) convolution kernel g, can be obtained by the methods presented in Section 2 and is not affected by noise. In this sense we can consider it as known. The analysis of solution (2.7) of the noiseless version (1.1) of (1.2) in Section 2 implies that estimation of f essentially reduces to estimating the r-th derivative of q in (1.1) and its Laplace convolution $q^{(r)} * \phi_1 = \int_0^t q^{(r)}(t-\tau)\phi_1(\tau)d\tau$ or, in view of (2.6) and (2.7), to estimating all derivatives $q^{(j)}$, $j = 0, ..., r$ up to order r and $q^{(r)} * \phi_1$, where $\phi_1 \in L_1 \cap L_2$ was defined in (2.9)-(2.10).

The errors of estimating derivatives $q^{(j)}$ are evidently dominated by the estimation error of the highest order derivative $q^{(r)}$. We now show that the latter also dominates the error of estimating Laplace convolution $q^{(r)} * \phi_1$. Indeed, let $\widehat{q_n^{(r)}}$ be any estimator of $q^{(r)}$ and estimate $q^{(r)} * \phi_1$ by the corresponding plug-in estimator $\widehat{q_n^{(r)}} * \phi_1$. Then,

$$\|\widehat{q_n^{(r)}} * \phi_1 - q^{(r)} * \phi_1\|_2 \le \|\phi_1\|_1 \cdot \|\widehat{q_n^{(r)}} - q^{(r)}\|_2 = O\left(\|\widehat{q_n^{(r)}} - q^{(r)}\|_2 \right) \tag{3.1}$$

(see also Theorem 2.2 of Gripenberg, Londen & Staffans, 1990). Thus, if there is a "good" estimator of $q^{(r)}$, one simply plugs it into (2.4) to estimate f.

In what follows, we shall present only the method for estimating the r-th derivative of q, since derivatives of lower orders can be estimated in a similar manner and have smaller risks. There exists a variety of methods to estimate derivatives of the unknown function and in the next section we consider the kernel estimator.

3.2 Adaptive estimation of r-th derivative

Re-write the original model (1.2) as

$$y_i = q(t_i) + \sigma\epsilon_i, \quad i = 1, ..., n,$$

where the unknown $q = g * f$ belongs to a Sobolev ball $H^{(r+m)}(A')$ of radius A', $r < m$ and $m + r - 1 \leq u$ (see Section 2), and we need to estimate $q^{(r)}$.

Let K be a kernel function (not to be confused with the convolution kernel g) of order $(r + m - 1, r)$, that is,

(K1) $supp(K) = [-1, 1]$, twice continuously differentiable and $\int K^2(t)dt < \infty$.

(K2) $\int t^j K(t)dt = \begin{cases} 0, & j = 0, ..., r-1, r+1, ..., m+r-1 \\ (-1)^r r!, & j = r \end{cases}$

The construction of such kernels is described in Appendix.

Define a well-known Priestley-Chao type kernel estimator of $q^{(r)}$ with a bandwidth λ:

$$\widehat{q_\lambda^{(r)}}(t) = \frac{1}{\lambda^{r+1}} \sum_{i=1}^{n} K\left(\frac{t - t_i}{\lambda}\right)(t_i - t_{i-1})y_i \tag{3.2}$$

In order to construct a consistent estimator of $q^{(r)}$, we impose the following restriction on the design:

(A5) Let T_n be such that $n^{-1}T_n^2 \to 0$ as $n \to \infty$ and there exist $1 \leq \mu < \infty$ such that $\max_i |t_i - t_{i-1}| \leq \mu T_n/n$.

By the standard asymptotic calculus for kernel estimation (see, e.g., Gasser & Müller, 1984) for an interior point t one has

$$Var\left(\widehat{q_{n,\lambda}^{(r)}}(t)\right) = \frac{\sigma^2}{\lambda^{2(r+1)}} \sum_{i=1}^{n}(t_i - t_{i-1})^2 K^2\left(\frac{t_i - t}{\lambda}\right) = \frac{\sigma^2}{\lambda^{2r+1}}\frac{T_n}{n}\int K^2(u)du\,(1 + o(1))$$

Certain boundary corrections are required to get the same order of error for t close to the boundaries (Gasser & Müller, 1984) and the integrated variance then is

$$V(\lambda) = \int_0^{T_n} Var\left(\widehat{q_{n,\lambda}^{(r)}}(t)\right) dt = V_0\,\frac{T_n^2}{\lambda^{2r+1}n}\,(1 + o(1)) \tag{3.3}$$

where $V_0 = \sigma^2 \|K\|^2$. Similarly,

$$E\left(\widehat{q_\lambda^{(r)}}(t)\right) = \lambda^{-r} \int K(u) f(t - u\lambda) du \; (1 + o(1)) \tag{3.4}$$

Expanding (3.4) into Taylor series and exploiting the moment assumptions on the kernel, one has

$$\begin{aligned} E\left(\widehat{q_\lambda^{(r)}}(t)\right) &= \frac{1}{\lambda^r} \int K(u) \left(q(t) + \ldots + \frac{(-1)^r q^{(r)}(t)}{r!} (\lambda u)^r + \ldots \right. \\ &+ \left. \int_0^{\lambda u} \frac{[(-1)(\lambda u - \tau)]^{r+m-1}}{(r+m-1)!} q^{(r+m)}(t+\tau) d\tau \right) du (1 + o(1)) \\ &= \left(q^{(r)}(t) + \frac{1}{\lambda^r (r+s-1)!} \int K(u) \int_0^{\lambda u} (\lambda u - \tau)^{r+m-1} q^{(r+m)}(t+\tau) d\tau du \right) (1 + o(1)) \end{aligned}$$

Changing the order of integration and applying Hölder's inequality, by straightforward calculus, the integrated squared bias can be written then as

$$B^2(\lambda, q) = \int_0^{T_n} \left(E\left(\widehat{q_{n,\lambda}^{(r)}}(t)\right) - q^{(r)}(t) \right)^2 dt = B_0 \lambda^{2m} (1 + o(1)), \tag{3.5}$$

where $B_0 = 2||q^{(m+r)}||^2 ||K\|^2 [(r+m)!]^{-2} (r+m)^{-1}$. Hence,

$$\sup_{q \in H^{m+r}(A)} E||\widehat{q_\lambda^{(r)}} - q^{(r)}||^2_{[0,T_n]} = \sup_{q \in H^{m+r}(A)} [V(\lambda) + B^2(\lambda, q)] = O\left(\frac{T_n^2}{\lambda^{2r+1} n}\right) + O(\lambda^{2m})$$

The asymptotically optimal bandwidth λ_n^* that minimizes $E||\widehat{q_{n,\lambda}^{(r)}} - q^{(r)}||^2_{[0,T_n]}$ is then

$$\lambda_n^* = O\left(\left(\frac{T_n^2}{n}\right)^{\frac{1}{2(m+r)+1}}\right) \tag{3.6}$$

and the corresponding optimal risk

$$\sup_{q^{(r)} \in H^{m+r}(A')} E\|\widehat{q_{\lambda_n^*}^{(r)}} - q^{(r)}\|^2_{[0,T_n]} = O\left(\left(\frac{T_n^2}{n}\right)^{\frac{2m}{2(m+r)+1}}\right). \tag{3.7}$$

The optimal bandwidth λ_n^* in (3.6) is evidently not so helpful in practice since it involves the unknown $q^{(m+r)}$ but the corresponding ideal global risk (3.7) can be used as a benchmark for assessment of estimation accuracy. In addition, λ_n^* depends on the regularity m of the unknown f in (1.2) which is rarely known precisely. We would like to construct a kernel estimator with a data-driven bandwidth that would be also adaptive to m. For this goal we utilize a general methodology developed by Lepski (e.g., Lepski, 1991) for data-based adaptive selection of the bandwidth $\hat{\lambda}$ such that the quadratic risk of the resulting kernel estimator $\widehat{q_{n,\hat{\lambda}}^{(r)}}(t)$ achieves the optimal rates (3.7) *simultaneously* over the entire range of m. In particular, we apply the global bandwidth version of Lepski, Mammen & Spokoiny's (1997) procedure and modify it also for estimating derivatives. The resulting procedure for estimating $q^{(r)}$ in (3.2) can be described as follows.

Consider a kernel K of order $(r+m_0-1,r)$, where $r<m_0\le u-r+1$ and the geometric grid of bandwidths Λ, where

$$\Lambda=\{\lambda_j\in[(n^{-1}T_n^2)^{\frac{1}{2r+1}},T_n]:\lambda_j=T_n a^{-j},\ j=0,1,...,J_n\}, \tag{3.8}$$

and $a>1$ is an arbitrary constant. Note that cardinality of Λ is at most $\mathrm{card}\,(\Lambda)=1+(2r+1)^{-1}\log_a(n/T_n^2)\le\log_a n$. Define

$$\hat{\lambda}_n=\max\{\lambda\in\Lambda:\|\widehat{q_\lambda^{(r)}}-\widehat{q_h^{(r)}}\|^2_{[0,T_n]}\le C_0^2 n^{-1}\sigma^2T_n^2h^{-(2r+1)}\text{ for all }h\in\Lambda,(n^{-1}T_n^2)^{\frac{1}{2r+1}}\le h<\lambda\}, \tag{3.9}$$

where

$$C_0^2>\mu^2\|K\|^2 \tag{3.10}$$

and estimate $q^{(r)}$ by $\widehat{q^{(r)}_{\hat{\lambda}_n}}$.

Proposition 1. *Let assumptions (A1)-(A5) hold and K with $supp(K)=[-1,1]$ be a square integrable, twice continuously differentiable inside $(-1,1)$ kernel of order $(r+m_0-1,r)$ where $r<m_0\le u-r+1$. Then,*

$$\sup_{q\in H^{m+r}(A')}E\|\widehat{q^{(r)}_{\hat{\lambda}_n}}-q^{(r)}\|^2=O\left(\left(\frac{T_n^2}{n}\right)^{\frac{2m}{2m+2r+1}}\right) \tag{3.11}$$

for all $r<m\le m_0$ and $A'>0$.

Note that the original Lepski, Mammen & Spokoiny's (1997) procedure is based on *locally* adaptive kernel estimation with a locally chosen bandwidth that necessarily yields an extra logarithmic factor in the rate of the *global* quadratic risk (3.7). The use of a (adaptive) global bandwidth allows one to remove this logarithmic factor and to achieve the sharp optimal rate.

3.3 Adaptive estimation of Laplace deconvolution

As we have argued in Section 3.1, the resulting estimator $\hat{f}_n$ for f in (1.2) is obtained by plugging the estimates of derivatives of q into (2.7). Following the ideas of the previous section, we choose twice continuously differentiable, square integrable kernels $K_j,\ j=0,...,r$ of respective orders $(r+m_0-1,j)$, where $r<m_0\le u-r+1$. We estimate the corresponding derivatives by

$$\widehat{q^{(j)}_{\hat{\lambda}_{j,n}}}(t)=\frac{1}{\hat{\lambda}_{j,n}^{j+1}}\sum_{i=1}^{n}K_j\left(\frac{t-t_i}{\hat{\lambda}_{j,n}}\right)(t_i-t_{i-1})y_i,\ j=0,...,r,$$

where, similarly to (3.9),

$$\hat{\lambda}_{n,j}=\max\{\lambda\in\Lambda:\|\widehat{q_\lambda^{(j)}}-\widehat{q_h^{(j)}}\|^2_{[0,T_n]}\le C_0^2n^{-1}\sigma^2T_n^2h^{-(2j+1)}\text{ for all }h\in\Lambda,(n^{-1}T_n^2)^{\frac{1}{2j+1}}\le h<\lambda\}$$

and C_0 is given in (3.10). Construct an estimator $\hat{f}_n$ of f as

$$\hat{f}_n(t) = B_r^{-1} \left(\widehat{q^{(r)}_{\hat{\lambda}_{n,r}}}(t) - \sum_{j=0}^{r-1} a_{0,r-1-j} \widehat{q^{(j)}_{\hat{\lambda}_{n,j}}}(t) - \int_0^x \widehat{q^{(r)}_{\hat{\lambda}_{n,r}}}(t-\tau)\phi_1(\tau) d\tau \right), \qquad (3.12)$$

where the function ϕ_1 and the coefficients a_{lj} were defined respectively in (2.9) and (2.10). Application of (3.1) and Proposition 1 with r replaced by j, $j = 0, \cdots, r$, yields the following theorem:

Theorem 3. *Let assumptions (A1)-(A5) hold and kernels K_j satisfy the above conditions. Then, for all $r < m \le m_0$ and $A > 0$,*

$$\sup_{f \in H^m(A)} E\|\hat{f}_n - f\|^2_{[0,T_n]} = O\left(\left(\frac{T_n^2}{n} \right)^{\frac{2m}{2(m+r)+1}} \right).$$

The proof is a direct consequence of Proposition 1 and the fact that that $f \in H^m(A)$ implies $q \in H^{m+r}(A')$ for a certain A' (see Section 2).

Under the additional conditions on f and T_n, the results of Theorem 3 can be easily extended to the entire positive half-line:

Corollary 1. *Suppose that there exists $\rho \ge 1$ such that $\int_0^\infty t^{2\rho} f^2(t) dt < \infty$ and $\lim_{n\to\infty} T_n^{-2\rho} n < \infty$. Let $\hat{f}_n$ be estimator (3.12) of f for $t \le T_n$ and $\hat{f}_n \equiv 0$ for $t > T_n$. Then,*

$$\sup_{f \in H^m(A)} E\|\hat{f}_n - f\|^2_{[0,\infty)} = O\left(\left(\frac{T_n^2}{n} \right)^{\frac{2m}{2(m+r)+1}} \right)$$

for all $r < m \le m_0$ and $A > 0$.

3.4 Lower bounds for the minimax risk

To prove the optimality (in the minimax sense) of the rates established in Theorem 3, below we derive the corresponding minimax lower bounds for the $L^2[0, T_n]$-risk:

Theorem 4. *Let assumptions (A1)–(A5) hold. Then, there exists a universal constant $C > 0$ such that*

$$\inf_{\tilde{f}_n} \sup_{f \in H^m(A)} E||\tilde{f}_n - f||^2_{[0,T_n]} \ge C \left(\frac{T_n^2}{n} \right)^{\frac{2m}{2(m+r)+1}}, \qquad (3.13)$$

where the infimum is taken over all possible estimators $\tilde{f}_n$ of f, and therefore

$$\inf_{\tilde{f}_n} \sup_{f \in H^m(A)} E||\tilde{f}_n - f||^2_{[0,\infty)} \ge C \left(\frac{T_n^2}{n} \right)^{\frac{2m}{2(m+r)+1}}.$$

Theorem 4 implies that the proposed adaptive Laplace deconvolution estimator is asymptotically minimax over the entire range of Sobolev classes.

4 Simulation study

In this section we present results of a simulation study to illustrate finite sample performance of the Laplace deconvolution procedure developed above. The data were simulated according to the model (1.2) with the true function $f(t) = e^{-t/3}$ and the convolution kernel $g(t) = e^{-t}$ (hence, $r = 1$ and $B_1 = 1$ in (A2)). The resulting $q = f * g$ in (1.2) is then $q(t) = 3(e^{-t/3} - e^{-t})/2$. We generated $N = 400$ samples by adding independent Gaussian noise $\mathcal{N}(0, \sigma^2)$ to $q(t)$ at $n = 200$ equally spaced points on $[0, T]$ with $T = 5$. Instead of considering various values of n and T_n, we fixed them and used a series of different values of $\sigma = 0.1, 0.05, 0.01, 0.005$ and 0.001. The examples of the resulting generated noisy data for various noise levels are presented in Figure 1.

The kernel K of order $(3, 1)$ used for estimating q' was constructed according to a general scheme described in the Appendix with the boundary corrections which follow Gasser & Müller (1984) procedure. In each simulation the bandwidth was adaptively selected by Lepski-type method described in Section 3.2 with the tuning parameters $a = 1.1$ in (3.8) and $C_0^2 = 1.5||K||^2$ in (3.10) (for equally spaced design $\mu = 1$). We also used a slightly modified Gasser-Müller version of the original Priestley-Chao kernel estimator (3.2) (see Gasser & Müller, 1984) which has the same asymptotic rates of convergence as the Priestley-Chao estimator but usually performs somewhat better in practice.

Figure 1 provides examples of deconvolution estimators based on single samples for four different values of σ. It is easy to see that the estimation precision increases rapidly when σ decreases. The estimator manages to capture basic features of the unknown f even for a high noise level ($\sigma = 0.1$). It demonstrates good behavior for quite high noise ($\sigma = 0.05$) and excellent performance when $\sigma = 0.01$. For $\sigma \leq 0.005$, the estimator $\hat{f}$ cannot essentially be distinguished from the true function f on the plot. Note also that despite the boundary corrections, the boundary effects are still quite significant especially for large noise.

The precision of the deconvolution estimator $\hat{f}$ of f was measured by the integrated mean squared error (IMSE). To reduce the impact of boundary effects on the overall IMSE (see comments above), we focused on interior points and calculated IMSE on a slightly smaller interior subinterval $[0.25, 4.75]$ excluding 10% of points on the boundaries, that is, IMSE is the average value over $N = 400$ runs of $||\hat{f} - f||^2_{[0.25,4.75]}$.

The resulting IMSE and their standard errors for various values of σ are given in Table 1. Figure 2 presents the corresponding boxplots of IMSE on the log-scale. One can see a clear linear decreasing tendency of ln(IMSE).

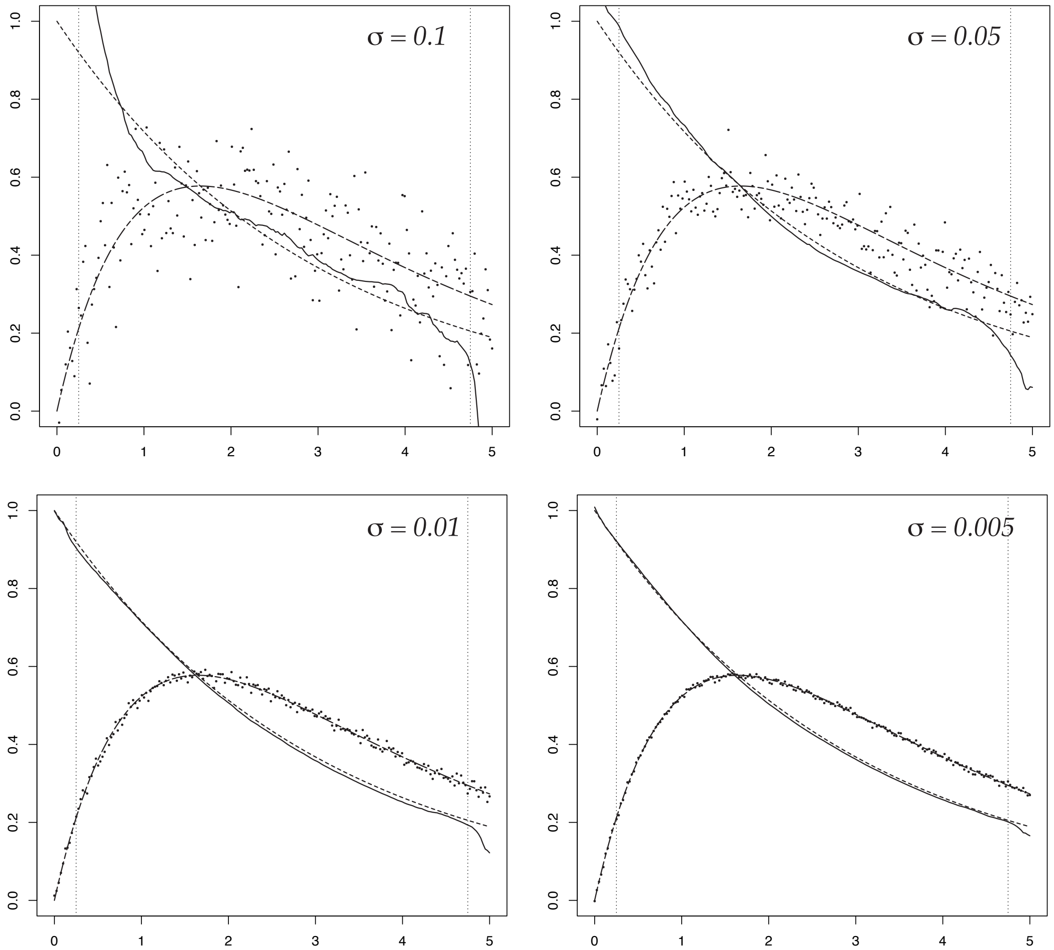

Figure 1: The noisy data, the true function f (dotted), $q = f * g$ (dot-dashed) and estimates $\hat{f}$ (solid) for $\sigma = 0.1, 0.05, 0.01, 0.005$. Vertical lines mark the subinterval of 90% of interior points used in calculating the overall IMSE.

σ	0.1	0.05	0.01	0.005	0.001
IMSE	0.027	0.0070	5.80e-4	2.29e-4	2.37e-5
SE	(0.0016)	(0.00040)	(2.46e-5)	(9.26e-6)	(7.34e-7)

Table 1: IMSE and their standard errors (in brackets) for various σ.

5 Discussion

In the present paper we consider Laplace deconvolution on the basis of discrete noisy data observed on the interval which length may increase with a sample size. Although this problem arises in a variety of applications, to the best of our knowledge, it has not been systematically studied in statistical literature and the paper contributes to fill this gap. We show that the original Laplace

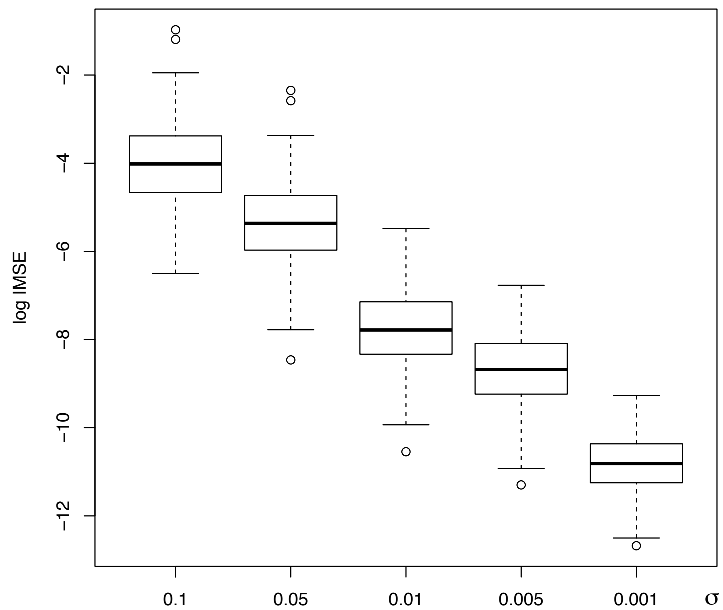

Figure 2: Boxplots of IMSE on the log-scale for various σ.

deconvolution problem can be essentially reduced to estimating derivatives. We derive an adaptive kernel estimator of the function of interest and established its asymptotic minimaxity over a range of Sobolev classes. The choice of the bandwidth in estimating derivatives is carried out by a version of the Lepski procedure (e.g., Lepski, 1991) applied globally which, as far as we know, has never been done previously. In fact, one can apply any other type of estimators (e.g., local polynomial regression, splines or wavelets) that allows an adaptive estimation of function derivatives. In particular, we believe that the use of wavelet-based methods can extend the adaptive minimaxity range from Sobolev to more general Besov classes.

A limited simulation study shows, that in addition to providing theoretical asymptotic results, the proposed Laplace deconvolution estimator demonstrates good performance in finite sample examples.

On the other hand, there is a number of open questions which remain unsolved. In particular, an interesting challenge would be to study Laplace deconvolution with unstable resolvents, where Assumption (A4) does not hold. Another important problem would be to study the equation (1.2) when the kernel g is not completely known and is estimated from observations. It is easy to see that methodology developed above relies heavily on the knowledge of the value of g and its derivatives at zero and, thus, cannot be automatically extended to this case.

Acknowledgments

Marianna Pensky was partially supported by National Science Foundation (NSF), grant DMS-1106564. We would like to thank Alexander Goldenshluger and Oleg Lepski for fruitful discussions of the paper.

6 Appendix

Throughout the proofs we use C to denote a generic positive constant, not necessarily the same each time it is used, even within a single equation.

Proof of Theorem 1

To prove Theorem 1 we utilize the following Lemma 1 which is essentially Theorem 7.2.4 of Gripenberg, Londen & Staffans (1990, Chapter 7) adapted to our notations.

Lemma 1. *Let s_g be such that*

$$\inf_{Re(s)=s_g} |\tilde{G}(s)| > 0 \quad \textit{and} \quad \lim_{\substack{|s|\to\infty \\ Re(s)\geq s_g}} |s^r \tilde{G}(s)| > 0. \tag{6.1}$$

Then, solution $\phi(\cdot)$ of equation (2.3) can be presented as

$$\phi(t) = \sum_{l=0}^{L} \sum_{j=0}^{\alpha_l - 1} \frac{a_{lj}}{j!} t^j e^{s_l t} + \phi_1(t) \tag{6.2}$$

where L is the total number of distinct zeros s_l of $s^r \tilde{G}(s)$ such that $Res_l > Res_g$, α_l is the order of zero s_l and $\phi_1 \in L_1$.

Choose s_g such that $s^* < s_g < 0$. Then, the first condition in (6.1) immediately follows from Assumption (A4). To validate the second assumption in (6.1), note that for $s = s_1 + is_2$ conditions $Re(s) \geq s_g$ and $|s| \to \infty$ imply that either $s_1 \to \infty$ or $|s_2| \to \infty$, or both. Recall that $s^r \tilde{G}(s) = B_r + \widetilde{G^{(r)}}(s)$. If $s_1 \to \infty$, no matter whether s_2 is finite or $s_2 \to \infty$, one has

$$\lim_{Re(s)\to\infty} |s^r \tilde{G}(s)| = \lim_{Re(s)\to\infty} |B_r + \int_0^\infty g^{(r)}(t) e^{-st} dt| = |B_r| > 0. \tag{6.3}$$

If s_1 is finite, $s_1 \geq s_g$, and $|s_2| \to \infty$, then Laplace transform $\widetilde{G^{(r)}}(s) = \int_0^\infty g^{(r)}(t) e^{-st} dt$ is equal to Fourier transform $\mathcal{F}[g^{(r)}(t) e^{-s_1 t}](s_2)$ of function $g^{(r)}(t) e^{-s_1 t}$ at the point s_2. Since $g^{(r)}(t) e^{-s_1 t} \in L_1(\mathbb{R}^+)$, one obtains

$$\lim_{|s_2|\to\infty} \int_0^\infty g^{(r)}(t) e^{-st} dt = \lim_{|s_2|\to\infty} \mathcal{F}[g^{(r)}(t) e^{-s_1 t}](s_2) = 0,$$

and (6.3) holds again. Hence, the second assumption in (6.1) is valid, and Lemma 1 can be applied.

Note that, under Assumption (A4), $\tilde{G}(s)$ has no zeros with $Re(s) > s_g$ and, therefore, has a single zero of r-th order at $s = 0$. Lemma 1 yields then that $\phi(t) = \phi_0(t) + \phi_1(t)$, where

$$\phi_0(t) = \sum_{j=0}^{r-1} \frac{a_j}{j!} t^j, \quad a_j = \phi^{(j)}(0), \tag{6.4}$$

and integrating by parts, one has

$$\int_0^t q^{(r)}(t-\tau)\phi_0(\tau)d\tau = \sum_{j=0}^{r-1} \phi_0^{(r-j-1)}(0) q^{(j)}(t), \tag{6.5}$$

that completes the proof.

□

Proof of Theorem 2

From (2.5), $\tilde{\Phi}(s)$ has poles s_l, $l = 0, \cdots, M$, of respective orders α_l, where $s_0 = 0$ and $\alpha_0 = r$. Note that $\lim_{\substack{|s|\to\infty \\ Re(s)\geq s_g}} |s^r \tilde{G}(s)| > 0$ (see the proof of Theorem 1) and, therefore, $\tilde{\Phi}$ does not have a pole at infinity. Then, $\tilde{\Phi}$ is a rational function and, consequently, can be represented using Cauchy integral formula

$$\tilde{\Phi}(s) = -\frac{1}{2\pi i} \sum_{l=0}^{M} \int_{C_l} \frac{\tilde{\Phi}(s)}{z-s} dz$$

where C_l, $l = 0, \cdots, M$, is a circle around the pole s_l such that this circle does not enclose any other pole of $\tilde{\Phi}$ (see LePage, 1961, Section 5.14). Using Laurent expansion of $\tilde{\Phi}(z)$ around s_l, we have

$$I_l(s) \quad = \quad -\frac{1}{2\pi i} \frac{\tilde{\Phi}(s)}{z-s} dz = \sum_{j=0}^{\alpha_l - 1} \frac{1}{(s-s_l)^{j+1}} \frac{1}{(\alpha_l - 1 - j)!} \frac{d^{\alpha_l - j - 1}}{ds^{\alpha_l - j - 1}} \left[(s-s_l)^{\alpha_l} \tilde{\Phi}(s) \right] \Bigg|_{s=s_l}$$

Combining the last two expressions and taking inverse Laplace transform of $\tilde{\Phi}(s)$ yields

$$\phi(t) = \sum_{l=0}^{M} \sum_{j=0}^{\alpha_l - 1} \frac{a_{lj}}{j!} t^j e^{s_l t}.$$

To validate the explicit expression (2.8) for $f(t)$, recall that $f(t)$ is of the form (2.4), i.e. (2.8) holds with

$$f_l(t) \quad = \quad \int_0^t q^{(r)}(t-\tau)\phi_l(\tau)d\tau, \quad l = 0, 1,$$

where ϕ_0 is given by (6.4), same as before, and ϕ_1 is defined in (2.9) and $\phi_1 \in L_1 \cap L_2$. To complete the proof, we just need to repeat calculations in (6.5).

□

Construction of a kernel of order (p, r)

To construct a kernel K of order (p, r) we use the orthonormal basis of Legendre polynomials $\{\psi_l(\cdot)\}_{l=0}^{\infty}$ in $L_2([-1,1])$, where

$$\psi_0(t) = \frac{1}{\sqrt{2}}, \quad \psi_l(t) = \sqrt{\frac{2l+1}{2}}\frac{1}{2^l l!}\frac{d^l}{dt^l}\left[(t^2-1)^l)\right],\ l = 1, 2, ...$$

that can be also obtained using the recursion formula

$$(l+1)\psi_{l+1}(t) = (2l+1)t\psi_l(t) - l\psi_{l-1}(t),\ l \geq 1, \quad \psi_0(u) = \frac{1}{\sqrt{2}},\ \psi_1(t) = \sqrt{\frac{3}{2}}t$$

Let K be the p-th degree polynomial of the form

$$K(t) = \sum_{l=0}^{p} \kappa_l \psi_l(t),\ t \in [-1,1]$$

For any $j = 0, 1, ..., p$, $t^j = \sum_{q=0}^{j} b_{jq}\psi_q(t)$, where $b_{jq} = \int t^j \psi_q(t)dt$. To satisfy (K2) we then have

$$\int t^j K(t)dt = \sum_{l=0}^{p}\sum_{q=0}^{j} \kappa_l b_{jq} \int \psi_l(t)\psi_q(t)dt = \sum_{l=0}^{j} \kappa_l b_{jl} = \begin{cases} (-1)^r r! & j = r \\ 0 & j \leq p,\ j \neq r \end{cases} \tag{6.6}$$

that essentially defines a system of $p+1$ linear equations with a triangular matrix. However, since ψ_l are symmetric functions for even l and antisymmetric otherwise, $b_{jl} = 0$ for even j, odd l and odd j, even l. Thus, (6.6) immediately yields $\kappa_l = 0$ for odd l (symmetric K) when r is even, and for even l (antisymmetric K) when r is odd, where half of the equations in (6.6) are, in fact, trivial. It can be also shown that the resulting kernel K of order (p, r) is the so-called minimal variance kernel that minimizes $V_0 = \sigma^2 ||K||^2$ (see (3.3)) subject to (K1)-(K2) (Gasser, Müller & Mammitzsch, 1985).

Proof of Proposition 1

Denote $d = (C_0 - \mu\|K\|)/(\sqrt{2}\|K\|)$ and set λ_n^* in (3.6) to be

$$\lambda_n^* = \left(d^2\ \frac{\sigma^2[(r+m)!]^2(r+m)}{2(A')^2}\ \frac{T_n^2}{n}\right)^{\frac{1}{2r+2m+1}}.$$

Note that

$$E\|\widehat{q_{\hat\lambda_n}^{(r)}} - q^{(r)}\|^2 = E\left\{\|\widehat{q_{\hat\lambda_n}^{(r)}} - q^{(r)}\|^2 I(\hat\lambda_n \geq \lambda_n^*)\right\} + E\left\{\|\widehat{q_{\hat\lambda_n}^{(r)}} - q^{(r)}\|^2 I(\hat\lambda_n < \lambda_n^*)\right\} = \Delta_1 + \Delta_2.$$

For $\hat\lambda_n \geq \lambda_n^*$, (3.7) and (3.9) imply

$$\begin{aligned} \Delta_1 &\leq 2E\left\{\|\widehat{q_{\hat\lambda_n}^{(r)}} - q_{\lambda_0^*}^{(r)}\|^2 I(\hat\lambda_n > \lambda_n^*)\right\} + 2E\left\{\|\widehat{q_{\hat\lambda_n^*}^{(r)}} - q^{(r)}\|^2 I(\hat\lambda_n > \lambda_n^*)\right\} \\ &= O\left(n^{-1}T_n^2(\lambda_n^*)^{-1/(2r+1)}\right) + O\left((n^{-1}T_n^2)^{-\frac{2m}{2m+2r+1}}\right) = O\left((n^{-1}T_n^2)^{-\frac{2m}{2m+2r+1}}\right) \end{aligned} \tag{6.7}$$

uniformly over $q \in H^{m+r}(A')$.

For $(n^{-1}T_n^2)^{\frac{1}{2r+1}} \le \hat{\lambda}_n < \lambda_n^*$, by direct calculus similar to that in Section 3.2, one can show that

$$\sup_{q\in H^{m+r}(A)} E\|\widehat{q^{(r)}_{\hat{\lambda}_n}} - q^{(r)}\|^4 \quad = \quad O((\lambda_n^*)^{4m}) + O\left((\lambda_n^*)^{-(4r+2)}n^{-2}T_n^4\right) = O(1).$$

Hence,

$$\begin{aligned}\sup_{q\in H^{m+r}(A)} \Delta_2 \quad &\le \quad \sup_{q\in H^{m+r}(A)} \sqrt{E\|\widehat{q^{(r)}_{\hat{\lambda}_n}} - q^{(r)}\|^4}\;\sqrt{P(\hat{\lambda}_n < \lambda_n^*)} \\ &= \quad \sup_{q\in H^{m+r}(A)} O\left(\sqrt{P(\hat{\lambda}_n < \lambda_n^*)}\right). \qquad (6.8)\end{aligned}$$

From the definition (3.9) of $\hat{\lambda}_n$, for $\lambda_n^* > \hat{\lambda}_n$ there exists $\tilde{h} < \lambda_n^*$ such that $\|\widehat{q^{(r)}_{\hat{\lambda}_n^*}} - q^{(r)}_{\tilde{h}}\|^2 > C_0^2 n^{-1}\sigma^2 T_n^2 \tilde{h}^{-(2r+1)}$, where, by (3.10) and definition of d, we have $C_0 = \|K\|(\mu + \sqrt{2}d)$. It follows from (3.3) and (3.5) that for all $h < \lambda_n^*$, the variance term dominates over the squared bias, that is,

$$\sup_{q\in H^{m+r}(A')} \|E\widehat{q^{(r)}_h} - q^{(r)}\|^2 \le d^2\sigma^2\|K\|^2 n^{-1}T_n^2 h^{-(2r+1)}.$$

Hence, for all $\tilde{h} < \lambda_n^*$ and $q \in H^{m+r}(A')$

$$P\left(\|\widehat{q^{(r)}_{\lambda_n^*}} - q^{(r)}_h\|^2 > \sigma^2 C_0^2 n^{-1}T_n^2 h^{-(2r+1)}\right) < P\left(\|\widehat{q^{(r)}_{\lambda_n^*}} - Eq^{(r)}_h\|^2 > \sigma^2\|K\|^2(\mu+d)^2 n^{-1}T_n^2 h^{-(2r+1)}\right) \qquad (6.9)$$

due to $C_0^2 - d^2 > \|K\|^2(\mu+d)^2$. Thus, uniformly over $q \in H^{s+r}(A')$, one has

$$\begin{aligned}P(\hat{\lambda}_n < \lambda_n^*) \quad &\le \quad \sum_{\substack{h\in\Lambda \\ h\le\lambda_n^*}} P(\tilde{h} = h)\; P\left(\|\widehat{q^{(r)}_{\lambda_n^*}} - q^{(r)}_h\|^2 > \sigma^2 C_0^2 n^{-1}T_n^2 h^{-(2r+1)}\right) \qquad (6.10) \\ &\le \quad \sum_{\substack{h\in\Lambda \\ h\le\lambda_n^*}} P(\tilde{h} = h)\; P\left(\|\widehat{q^{(r)}_h} - E\widehat{q^{(r)}_h}\|^2 \ge \sigma^2\|K\|^2(\mu+d)^2 n^{-1}T_n^2 h^{-(2r+1)}\right).\end{aligned}$$

Note that

$$\|\widehat{q^{(r)}_h} - E\widehat{q^{(r)}_h}\|^2 = \left\|\sum_{i=1}^n h^{-(r+1)}K\left(\frac{t-t_i}{h}\right)(t_i - t_{i-1})\epsilon_i\right\|^2 = h^{-(2r+1)}n^{-2}T_n^2\; \boldsymbol{\epsilon}^T\boldsymbol{Q}\boldsymbol{\epsilon},$$

where $\boldsymbol{Q}$ is an $n \times n$ symmetric nonnegative-definite matrix with elements

$$Q_{ij} = \frac{n^2}{T_n^2}(t_i - t_{i-1})(t_j - t_{j-1}) \int_{-1}^1 K(z)K\left(z + \frac{t_i - t_j}{h}\right)dz. \qquad (6.11)$$

Then,

$$P\left(\|\widehat{q^{(r)}_h} - E\widehat{q^{(r)}_h}\|^2 \ge \sigma^2\|K\|^2(\mu+d)^2 n^{-1}T_n^2 h^{-(2r+1)}\right) = P\left(\boldsymbol{\epsilon}^T\boldsymbol{Q}\boldsymbol{\epsilon} \ge n\|K\|^2(\mu+d)^2\right). \qquad (6.12)$$

Applying a χ^2-type inequality which initially appeared in Laurent and Massart (1998), was improved by Comte (2001) and furthermore by Gendre (2010) in his Ph.D. Thesis, we have for any $x > 0$

$$P\left(\sigma^{-2}\boldsymbol{\epsilon}^T\boldsymbol{Q}\boldsymbol{\epsilon} \geq \left[\sqrt{\mathrm{Tr}(\boldsymbol{Q})} + \sqrt{x\rho^2_{\max}(\boldsymbol{Q})}\right]^2\right) \leq e^{-x}, \tag{6.13}$$

where $\mathrm{Tr}(\boldsymbol{Q})$ is the trace of $\boldsymbol{Q}$ and $\rho^2_{\max}(\boldsymbol{Q})$ is the maximal eigenvalue of $\boldsymbol{Q}$. Note that

$$\mathrm{Tr}(\boldsymbol{Q}) = \frac{n^2}{T_n^2}\sum_{i=1}^n (t_i - t_{i-1})^2\|K\|^2 \leq n\mu^2\|K\|^2.$$

and $\rho^2_{\max}(\boldsymbol{Q})$ is the spectral norm of matrix $\boldsymbol{Q}$ which is dominated by any other norm. In particular,

$$\rho^2_{\max}(\boldsymbol{Q}) \leq \max_k \sum_{l=1}^n |Q_{kl}| = \frac{n^2}{T_n^2}\max_k(t_k - t_{k-1})\int_{-1}^1 |K(z)|\left[\sum_{l=1}^n \left|K\left(z + \frac{t_k - t_l}{h}\right)\right|(t_l - t_{l-1})\right]dz.$$

Since

$$\sum_{l=1}^n \left|K\left(z + \frac{t_k - t_l}{h}\right)\right|(t_l - t_{l-1}) = \int_{-1}^1 \left|K\left(z + \frac{t_k - t}{h}\right)\right| dt(1+o(1)) = h\int_{-1}^1 \left|K\left(z + \frac{t_k}{h} - y\right)\right| dt(1+o(1)),$$

we have

$$\begin{aligned}\rho^2_{\max}(\boldsymbol{Q}) &\leq \frac{n^2}{T_n^2}\max_k(t_k - t_{k-1})\, h \int_{-1}^1\int_{-1}^1 |K(z)||K(z + t_k/h - y)|dzdy \\ &\leq \mu\frac{nh}{T_n}\left[\int_{-1}^1 |K(z)|dz\right]^2 \leq 2\mu\|K\|^2\frac{nh}{T_n}.\end{aligned}$$

Using inequality (6.13) with $x = d^2T_n/(2\mu h)$ and $h < \lambda_n^*$ obtain

$$P\left(\|\widehat{q_h^{(r)}} - E\widehat{q_h}^{(r)}\|^2 \geq \frac{\sigma^2\|K\|^2(\mu+d)^2T_n^2}{nh^{2r+1}}\right) \leq \exp\left(-\frac{d^2T_n}{2\mu h}\right) \leq \exp\left(-Cn^{\frac{1}{2r+2m+1}}T_n^{\frac{2r+2m-1}{2r+2m+1}}\right) \tag{6.14}$$

where C depends on r,m, A', μ and d. Combination of (6.7), (6.8), (6.10) and (6.14) complete the proof.

□

Proof of Theorem 4

The main idea of the proof is to find a subset of functions $\mathcal{F} \subset H^m(A)$ such that for any pair $f_1,\ f_2 \in \mathcal{F}$,

$$\|f_1 - f_2\|^2_{[0,T_n]} \geq 4C(T_n^2 n^{-1})^{2m/(2(m+r)+1)} \tag{6.15}$$

and the Kullback-Leibler divergence

$$\mathbb{K}(\mathbb{P}_{f_1}, \mathbb{P}_{f_2}) = \frac{||\boldsymbol{q}_1 - \boldsymbol{q}_2||^2_{\mathbb{R}^n}}{2\sigma^2} \leq \frac{\log \mathrm{card}(\mathcal{F})}{16}, \tag{6.16}$$

where log stands for natural logarithm and $\boldsymbol{q}_{ji} = (g * f_j)(t_i)$, $i = 1, ..., n$, $j = 1, 2$. The result will then follow immediately from Lemma A.1 of Bunea, Tsybakov and Wegkamp (2007).

Without loss of generality, let us assume that the points are equally spaced, i.e. $t_i - t_{i-1} = T_n/n$, $i = 1, \cdots, n$. To construct such a subset $\mathcal{F}$, define integers $M_n \geq 8$ and $N = \left[\frac{n}{M_n}\right]$, the largest integer which does not exceed n/M_n. Let $\lambda_n = NT_n/n$ and define points $z_l = l\,\lambda_n$, $l = 0, 1, \cdots, M_n$. Note that the latter implies that points of observation $t_j = j\,T_n/n$ in equation (1.2) are related to z_l as $z_l = t_j$ where $j = Nl$ for $l = 1, \cdots, M_n$ and $j \leq NM_n$. Note also that $\frac{T_n}{2M_n} \leq \lambda_n \leq \frac{T_n}{M_n}$.

Let $k(\cdot)$ be an infinitely differentiable function with $supp(k) = [0, 1]$ and such that

$$\int_0^1 x^j k(x) dx = 0, j = 0, \cdots, r-1, \quad \int_0^1 x^r k(x) dx \neq 0. \tag{6.17}$$

Introduce functions

$$\varphi_j(x) = L \frac{\lambda_n^s}{\sqrt{T_n}} k\left(\frac{x - z_{j-1}}{\lambda_n}\right) \quad l = 1, \cdots, M_n,$$

where the constant $L > 0$ will be defined later. Note that φ_j have non-overlapping supports, where $supp(\varphi_j) = [z_{j-1}, z_j]$.

Consider the set of all binary sequences of the length M_n:

$$\Omega = \left\{\boldsymbol{\omega} = (\omega_1, \cdots, \omega_{M_n}), \ \ \omega_j = \{0, 1\}\right\} = \{0, 1\}^{M_n}$$

For sufficiently large n, $M_n \geq 8$ and Varshamov-Gilbert bound (see, e.g. Lemma 2.9 of Tsybakov (2009)) ensures the existence of a subset $\tilde{\Omega} \subset \Omega$ such that $\log_2 \text{card}(\tilde{\Omega}) \geq M_n/8$ and the Hamming distance $\rho(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) = \sum_{j=1}^{M_n} \mathbb{I}\{\boldsymbol{\omega}_{1j} \neq \boldsymbol{\omega}_{2j}\} \geq M_n/8$ for any pair $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2 \in \tilde{\Omega}$. Consider then the corresponding subset of functions

$$\mathcal{F} = \{f_\omega : f_\omega(t) = \sum_{j=1}^{M_n} w_j \varphi_j(t), \ \omega \in \tilde{\Omega}\}. \tag{6.18}$$

We now need to show that $\mathcal{F}$ in (6.18) is exactly the required set. Note first that since the supports of φ_j are non-overlapping, for any $f_\omega \in \mathcal{F}$ a straightforward calculus yields

$$||f_\omega||^2_{[0,T_n]} \leq \sum_{j=1}^{M_n} ||\varphi_j||^2 = L^2 \frac{\lambda_n^{2s+1}}{T_n} M_n ||k||^2 = L^2 \lambda^{2m} ||k||^2 \leq L^2 ||k||^2$$

Similarly,

$$||f_\omega^{(m)}||^2_{[0,T_n]} \leq \sum_{j=1}^{M_n} ||\varphi_j^{(m)}||^2 = \frac{L^2}{T_n} m \lambda_n ||k^{(s)}||^2 = L^2 ||k^{(m)}||^2 < \infty$$

and therefore $f_\omega \in H^{(m)}(A)$, where $A = L||k||_{H^m}$. Furthermore,

$$||f_{\omega_1} - f_{\omega_2}||^2_{[0,T_n]} = L^2 \frac{\lambda_n^{2m+1}}{T_n} ||k||^2 \rho(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) \geq L^2 \frac{\lambda_n^{2m+1}}{T_n} \frac{M_n}{8} \geq 4C\lambda_n^{2m}$$

and (6.15) holds provided $\lambda_n \geq C(T_n^2 n^{-1})^{-1/(2(m+r)+1)}$ for some positive constant C.

To verify (6.16), note that

$$\mathbb{K}(P_1, P_2) = \frac{1}{2\sigma^2} \sum_{i=1}^{n} [q_1(t_i) - q_2(t_i)]^2 \le \frac{1}{\sigma^2} \sum_{j=1}^{2} Q(f_j) \tag{6.19}$$

where, suppressing index j, we write

$$Q(f) \;=\; \sum_{i=1}^{n} \left[\int_0^{t_i} g(t_i - x) f(x) dx \right]^2 = \frac{L^2 \lambda_n^{2m}}{T_n} \sum_{i=1}^{n} \left[\sum_{l=1}^{M_n} \omega_l^{(j)} \int_0^{t_i} g(t_i - x) k\left(\frac{x - z_{l-1}}{\lambda_n} \right) dx \right]^2 .$$

In order to obtain an upper bound for $Q(f)$ we need the following supplementary lemma, the proof of which is presented at the end of the section.

Lemma 2. *Introduce functions $K_j(x)$ using the following recursive relation*

$$K_1(x) = \int_0^x k(t)dt, \quad K_j(x) = \int_0^x K_{j-1}(t)dt, \quad j = 2, \cdots, r. \tag{6.20}$$

Then, under condition (6.17), functions $K_j(x)$, $j = 1, \cdots, r$, are uniformly bounded and $K_j(1) = 0$, $j = 1, \cdots, r$. Moreover,

$$\begin{aligned} \int_0^{t_i} g(t_i - x) k\left(\frac{x - z_{l-1}}{\lambda_n} \right) dx \;&=\; \lambda_n^r \left[B_r K_r\left(\frac{t_i - z_{l-1}}{\lambda_n} \right) \mathbb{I}(z_{l-1} \le y_i \le z_l) \right. \\ &+ \left. \int_{\min(z_{l-1}, t_i)}^{\min(z_l, t_i)} g^{(r)}(t_i - x) K_r\left(\frac{x - z_{l-1}}{\lambda_n} \right) dx \right]. \end{aligned} \tag{6.21}$$

Applying equation (6.21) to the integral in $Q(f)$, obtain

$$Q(f) \le 2L^2 \lambda_n^{2m+2r} T_n^{-1} (\Delta_1 + \Delta_2) \tag{6.22}$$

where

$$\begin{aligned} \Delta_1 \;&=\; \sum_{i=1}^{n} \left[\sum_{l=1}^{M_n} B_r \; K_r\left(\frac{t_i - z_{l-1}}{\lambda_n} \right) \mathbb{I}(z_{l-1} \le y_i \le z_l) \right]^2 , \\ \Delta_2 \;&=\; \sum_{i=1}^{n} \left[\sum_{l=1}^{M_n} \int_{\min(z_{l-1}, t_i)}^{\min(z_l, t_i)} g^{(r)}(t_i - x) K_r\left(\frac{x - z_{l-1}}{\lambda_n} \right) dx \right]^2 . \end{aligned}$$

Observe that for any t and any l_1 and l_2 such that $l_1 \ne l_2$, one has $K_r(\lambda_n^{-1}(t - z_{l_1})) K_r(\lambda_n^{-1}(t - z_{l_2})) = 0$. Also, for each i, $K_r(\lambda_n^{-1}(ti - z_l)) \ne 0$ for only one value of l, namely, for $l = [i/N] + 1$ where $[x]$ is the largest integer which does not exceed x. Therefore,

$$\Delta_1 \le B_r^2 \sum_{i=1}^{n} K_r^2 \left(\frac{t_i - z_{[i/N]}}{\lambda_n} \right) \le n \; B_r^2 \|K_r\|_\infty^2, \tag{6.23}$$

where $\| \cdot \|_\infty$ is the supremum norm. In order to obtain an upper bound for Δ_2, observe that for any nonnegative function $F(x)$ one has

$$\int_{\min(z_{l-1}, t_i)}^{\min(z_l, t_i)} F(x) dx \le \int_{z_{l-1}}^{z_l} F(x) dx.$$

Hence, we derive

$$
\begin{aligned}
\Delta_2 &\le \sum_{i=1}^{n} \left[\sum_{l=1}^{M_n} \int_{z_{l-1}}^{z_l} \left| g^{(r)}(t_i - x) K_r \left(\frac{x - z_{l-1}}{\lambda_n} \right) \right| dx \right]^2 \\
&\le \sum_{i=1}^{n} \|K_r\|_\infty^2 \left[\sum_{l=1}^{M_n} \int_{z_{l-1}}^{z_l} |g^{(r)}(t_i - x)| dx \right]^2 \le n \, \|g^{(r)}\|^2 \|K_r\|_\infty^2. \qquad (6.24)
\end{aligned}
$$

Combining formulae (6.19)–(6.24), we obtain that, in order to satisfy condition (6.16), we need the following inequality to hold

$$
\mathbb{K}(\mathbb{P}_{f_1}, \mathbb{P}_{f_2}) \le \frac{2L^2 \lambda_n^{2m+2r} n}{\sigma^2 T_n} \, \|K_r\|_\infty^2 [B_r^2 + \|g^{(r)}\|_2^2] \le \frac{1}{16} \, \frac{M_n \log 2}{8}. \qquad (6.25)
$$

Since $\lambda_n \le T_n/M_n$, inequalities in (6.25) hold whenever $M_n \ge C n^{1/(2(m+r)+1)} T_n^{(2(m+r)-1)/(2(m+r)+1)}$. Taking $M_n = C n^{1/(2(m+r)+1)} T_n^{(2(m+r)-1)/(2(m+r)+1)}$ obtain $\lambda_n \ge T_n/(2M_n) \ge C(T_n^2 n^{-1})^{1/(2(m+r)+1)}$. Therefore, both conditions (6.15) and (6.16) hold and theorem is proved.

□

Proof of Lemma 2. Definitions (6.20) imply that $k(x) = K_1'(x)$, $K_{j-1}'(x) = K_j(x)$ and $K_j(0) = 0$, $j = 1, \cdots, r$. Observe that condition $K_j(1) = 0$, $j = 1, \cdots, r$, is equivalent to

$$
\int_0^1 K_j(x) dx = 0, \quad j = 0, \cdots, r-1, \qquad (6.26)
$$

where $K_0(x) = k(x)$. It is easy to see that (6.26) is valid for $j = 0$. For $j \ge 1$, note that, by formula (4.631) of Gradshtein and Ryzhik (1980),

$$
K_j(x) = \int_0^x dz_{j-1} \int_0^{z_{j-1}} dz_{j-2} \cdots \int_0^{z_1} k(z) dz = \frac{1}{(j-1)!} \int_0^x (x-z)^{j-1} k(z) dz. \qquad (6.27)
$$

Then, for any $x \in [0,1]$, one has $|K_j(x)| \le [(j-1)!]^{-1} \, \|k\|_\infty \int_0^x (x-z)^{j-1} dz \le \|k\|_\infty$. Moreover, by (6.27), for $j = 1, \cdots, r-1$, one has

$$
\begin{aligned}
&\int_0^1 K_j(x) dx = \frac{1}{(j-1)!} \int_0^1 dx \int_0^x (x-z)^{j-1} k(z) dz \\
&= \frac{1}{(j-1)!} \int_0^1 k(z) dz \int_z^1 (x-z)^{j-1} dx = \frac{1}{(j-1)! j!} \int_0^1 (1-z)^j \, k(z) dz = 0.
\end{aligned}
$$

Now, it remains to prove formula (6.21). Note that support of the function $k(u/\lambda_n - (l-1))$ coincides with (z_{l-1}, z_l), so that

$$
I(i,l) = \int_0^{t_i} g(t_i - x) k \left(\frac{x - z_{l-1}}{\lambda_n} \right) dx = \int_{\min(z_{l-1}, t_i)}^{\min(z_l, t_i)} g(t_i - x) k \left(\frac{x - z_{l-1}}{\lambda_n} \right) dx. \qquad (6.28)
$$

Formula (6.28) implies that $I(i,l) = 0$ whenever $z_{l-1} \geq y_i$. If $z_{l-1} < y_i \leq z_l$, it follows from (6.28) that

$$I(i,l) = \int_{z_{l-1}}^{t_i} g(t_i - x) k\left(\frac{x - z_{l-1}}{\lambda_n}\right) dx.$$

Introduce new variable $t = x - z_{l-1}$ and denote $u_{il} = t_i - z_{l-1}$. Then, recalling condition (A2) and using integration by parts, we derive

$$\begin{aligned}
I(i,l) &= \int_0^{u_{il}} g(u_{il} - t) k\left(\frac{t}{\lambda_n}\right) dt = \lambda_n \int_0^{u_{il}} g(u_{il} - t) dK_1\left(\frac{t}{\lambda_n}\right) \\
&= \lambda_n g(u_{il} - t) K_1\left(\frac{t}{\lambda_n}\right)\Bigg|_0^{u_{il}} + \lambda_n \int_0^{u_{il}} g'(u_{il} - t) K_1\left(\frac{t}{\lambda_n}\right) dt \\
&= \cdots = \lambda_n^r g^{(r-1)}(u_{il} - t) K_r\left(\frac{t}{\lambda_n}\right)\Bigg|_0^{u_{il}} + \lambda_n^r \int_0^{u_{il}} g^r(u_{il} - t) K_r\left(\frac{t}{\lambda_n}\right) dt.
\end{aligned}$$

Changing variables back to x, we arrive at

$$I(i,l) = \lambda_n^r \left[B_r\, K_r\left(\frac{t_i - z_{l-1}}{\lambda_n}\right) + \int_{z_{l-1}}^{t_i} g^{(r)}(t_i - x) K_r\left(\frac{x - z_{l-1}}{\lambda_n}\right) dx \right]. \tag{6.29}$$

Finally, consider the case when $z_l \leq y_i$. Then, using relation $z_l = z_{l-1} + \lambda_n$, integration by parts and the fact that $K_j(0) = K_j(1) = 0$ for $j = 1, \cdots, r$, we obtain

$$\begin{aligned}
I(i,l) &= \int_{z_{l-1}}^{z_l} g(t_i - x) k\left(\frac{x - z_{l-1}}{\lambda_n}\right) dx = \lambda_n \int_0^1 g(t_i - z_{l-1} - \lambda_n t) k(t) dt \\
&= \lambda_n \int_0^1 g(t_i - z_{l-1} - \lambda_n t) dK_1(t) = \lambda_n^2 \int_0^1 g'(t_i - z_{l-1} - \lambda_n t) K_1(t) dt \\
&= \cdots = \lambda_n^{r+1} \int_0^1 g^r(t_i - z_{l-1} - \lambda_n t) K_r(t) dt = \lambda_n^r \int_{z_{l-1}}^{z_l} g^{(r)}(t_i - x) K_r\left(\frac{x - z_{l-1}}{\lambda_n}\right) dx
\end{aligned}$$

which, in combination with (6.29), completes the proof.

References

[1] Ameloot, M., Hendrickx, H. (1983) Extension of the performance of Laplace deconvolution in the analysis of fluorescence decay curves. *Biophys. Journ.*, **44**, 27 - 38.

[2] Ameloot, M., Hendrickx, H., Herreman, W., Pottel, H., Van Cauwelaert, F., and van der Meer, W. (1984) Effect of orientational order on the decay of the fluorescence anisotropy in membrane suspensions. Experimental verification on unilamellar vesicles and lipid/alpha-lactalbumin complexes. *Biophys. Journ.*, **46**, 525 - 539.

[3] Bisdas, S., Konstantinou, G.N., Lee, P.S., Thng, C.H., Wagenblast, J., Baghi, M. and Koh, T.S. (2007) Dynamic contrast-enhanced CT of head and neck tumors: perfusion measurements using a distributed-parameter tracer kinetic model. Initial results and comparison with deconvolution- based analysis. *Physics in Medicine and Biology*, **52**, 6181-6196.

[4] Bunea, F., Tsybakov, A. and Wegkamp, M.H. (2007). Aggregation for Gaussian regression. *Ann. Statist.* **35**, 1674–1697.

[5] Cao, M.,Liang, Y., Shen, C., Miller, K.D. and Stantz, K.M. (2010) Developing DCE-CT to quantify intra-tumor heterogeneity in breast tumors with differing angiogenic phenotype. *IEEE Trans. Medic. Imag.*, **29**, 1089 - 1092.

[6] Carroll, R. J., and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184-1186.

[7] Chauveau, D.E., van Rooij, A.C.M. and Ruymgaart, F.H. (1994). Regularized inversion of noisy Laplace transform. *Adv. Applied Math.* **15**, 186–201.

[8] Cinzori, A.C., and Lamm, P.K. (2000) Future polynomial regularization of ill-posed Volterra equations. *SIAM J. Numer. Anal.*, **37**, 949979.

[9] Comte, F. (2001) Adaptive estimation of the spectrum of a stationary Gaussian sequence. *Bernoulli*, **7**, 267-298.

[10] Cuenod, C.A., Fournier, L., Balvay, D. and Guinebretire, J.M. (2006) Tumor angiogenesis: pathophysiology and implications for contrast-enhanced MRI and CT assessment. *Abdom. Imaging*, **31**, 188-193.

[11] Delaigle, A., Hall, P. and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.*, **36**, 665-685.

[12] Dey, A.K., Martin, C.F. and Ruymgaart, F.H. (1998). Input recovery from noisy output data, using regularized inversion of Laplace transform. *IEEE Trans. Inform. Theory*, **44**, 1125–1130.

[13] Diggle, P. J., and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser.* **B**, **55** 523–531.

[14] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist.*, **19**, 1257-1272.

[15] Fan, J. and Koo, J. (2002). Wavelet deconvolution. *IEEE Trans. Inform. Theory*, **48**, 734–747.

[16] Gasser, T. and Müller, H-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, **11**, 171–185.

[17] Gasser, T., Müller, H-G., and Mammitzsch (1985). Kernels for nonparametric kernel estimation. *J. Roy. Statist. Soc. Ser.* **B**, **47**, 238–252.

[18] Gafni, A., Modlin, R. L. and Brand, L. (1975) Analysis of fluorescence decay curves by means of the Laplace transformation. *Biophys. J.*, **15**, 263–280.

[19] Goh, V., Halligan, S., Hugill, J.A., Gartner, L. and Bartram, C.I. (2005) Quantitative colorectal cancer perfusion measurement using dynamic contrastenhanced multidetector-row computed tomography: effect of acquisition time and implications for protocols. *J. Comput. Assist. Tomogr.*, **29**, 59 - 63.

[20] Goh, V. and Padhani, A. R. (2007) Functional imaging of colorectal cancer angiogenesis. *Lancet Oncol.*, **8**, 245 - 255.

[21] Gradshtein, I.S. and Ryzhik, I.M. (1980) *Tables of Integrals, Series, and Products.* Academic Press, New York.

[22] Gripenberg, G., Londen, S.O., and Staffans, O. (1990) *Volterra Integral and Functional Equations.* Cambridge University Press, Cambridge.

[23] Johnstone, I.M., Kerkyacharian, G., Picard, D. and Raimondo, M. (2004) Wavelet deconvolution in a periodic setting. *J. Roy. Statist. Soc. Ser.* **B**, **66**, 547–573 (with discussion, 627–657).

[24] Lakowicz, J.R. (2006) *Principles of Fluorescence Spectroscopy.* Kluwer Academic, New York.

[25] Lamm, P. (1996) Approximation of ill-posed Volterra problems via predictor-corrector regularization methods. *SIAM J. Appl. Math.*, **56**, 524-541.

[26] Laurent, B. and Massart, P. (1998), Adaptive estimation of a quadratic functional by model selection, Technical report, Universite de Paris-Sud, Mathematiques.

[27] LePage, W.R. (1961) *Complex Variables and the Laplace Transform for Engineers.* Dover, New-York.

[28] Lepski, O.V. (1991). Asymptotic mimimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.*, **36**, 654-659.

[29] Lepski, O.V., Mammen, E., and Spokoiny, V.G. (1997) Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors *Ann. Statist.*, **25**, 929-947.

[30] Lien, T.N., Trong, D.D. and Dinh, A.P.N. (2008) Laguerre polynomials and the inverse Laplace transform using discrete data *J. Math. Anal. Appl.*, **337**, 1302–1314.

[31] Maleknejad, K., Mollapourasl, R. and Alizadeh, M. (2007) Numerical solution of Volterra type integral equation of the first kind with wavelet basis. *Appl. Math.Comput.*, **194**, 400405.

[32] McKinnon, A. E., Szabo, A. G. and Miller, D. R. (1977) The deconvolution of photoluminescence data. *J. Phys. Chem.*, **81**, 1564-1570.

[33] Miles, K. A. (2003) Functional CT imaging in oncology. *Eur. Radiol.*, 13 - suppl. 5, **M134-8**.

[34] O'Connor, D. V., Ware, W. R. and Andre, J. C. (1979) Deconvolution of fluorescence decay curves. A critical comparison of techniques. *J. Phys. Chem.*, **83**, 1333-1343.

[35] Padhani, A. R. and Harvey, C. J. (2005) Angiogenesis imaging in the management of prostate cancer. *Nat. Clin. Pract. Urol.*, **2**, 596-607.

[36] Pensky, M., and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, **27**, 2033–2053.

[37] Polyanin, A.D., and Manzhirov, A.V. (1998) *Handbook of Integral Equations*, CRC Press, Boca Raton, Florida.

[38] Rashed, M.T. (2003) Numerical solutions of the integral equations of the first kind *Appl. Math. Comput.*, **145**, 413420.

[39] Stefanski, L., and Carrol, R. J. (1990). Deconvoluting kernel density estimators. *Statistics*, **21**, 169-184.

[40] Tsybakov, A.B. (2009) *Introduction to Nonparametric Estimation*, Springer, New York.

[41] Weeks, W.T. (1966) Numerical Inversion of Laplace Transforms Using Laguerre Functions. *J. Assoc. Comput. Machinery*, **13**, 419 - 429.

Felix Abramovich
Department of Statistics & Operations Research
Tel Aviv University
Tel Aviv 69978, Israel
felix@post.tau.ac.il

Marianna Pensky
Department of Mathematics
University of Central Florida
Orlando FL 32816-1353, USA
Marianna.Pensky@ucf.edu

Yves Rozenholc
Université Paris, Descartes,
MAP5, UMR CNRS 8145, France
yves.rozenholc@univ-paris5.fr