Blind Intensity Estimation from Shot-Noise Data

Raúl E. Sequeira, *Member, IEEE*, and John A. Gubner, *Member, IEEE*

***Abstract*—The estimation of the intensity function of an inhomogeneous Poisson process is considered when the observable data consists of sampled shot noise that results from passing the Poisson process through an unknown linear time-invariant system. The proposed method consists of first estimating a histogram of the underlying point process. The estimated histogram is used to construct a kernel estimate of the intensity function. An estimate of the unknown impulse response of the linear time-invariant system is constructed via a regularized backsubstitution of a discrete-time convolution with the estimated histogram.**

I. Introduction

In a variety of physical applications, including direct-detection optical communications systems and quantum-limited imaging, the mathematical models involve inhomogeneous Poisson processes at some stage in the system [1], [3], [10], [17], [27]. The main elements of these processes are a collection of random events $\{T_\nu\}$ (occurrence times or positions, depending on the application) and an intensity function $\lambda(\cdot)$ governing the statistics of the T_ν [27]. In these applications, λ is the function of interest. Since λ can never be directly observed, inference about it can only be made based on observation of the $\{T_\nu\}$ or on functionals of the $\{T_\nu\}$. For example, in optical communications and quantum-limited imaging, the data can be modeled by the random superposition

$$z(t) := \sum_\nu h(t - T_\nu) \tag{1.1}$$

where h is the impulse response, or point-spread function, resulting from the finite bandwidth of the system. The random process (1.1) is known as *shot noise* (filtered point process).

Two examples of 1-D shot-noise data are shown in Figs. 1 and 2. In Fig. 1, h has a discontinuity at the origin, and hence, z has a discontinuity at each T_ν; in Fig. 2, h is continuous, and hence, z is also continuous.

In our prior paper [24], the problem of estimating λ was addressed under the assumption that the impulse response h is known. Since $\mathsf{E}[z(t)] = \int h(t-\tau)\lambda(\tau)d\tau$, this led to obtaining an approximate solution to the noisy integral equation

$$z(t) = \int h(t-\tau)\lambda(\tau)d\tau + v(t) \tag{1.2}$$

where $v(t) := z(t) - \mathsf{E}[z(t)]$ is a zero-mean "noise" term. In the present paper, we remove the assumption that h is known and address the joint estimation of h and λ based on discrete samples of z, say, $z(t_0), z(2t_0), \ldots, z(Kt_0)$, without making any assumptions on the shape of λ. Our approach is to first obtain an estimate of a histogram of the $\{T_\nu\}$; the estimate of h is then obtained as an approximate solution to a discrete convolution equation, and that of λ is obtained as a *kernel intensity estimate* modified for histogram data.

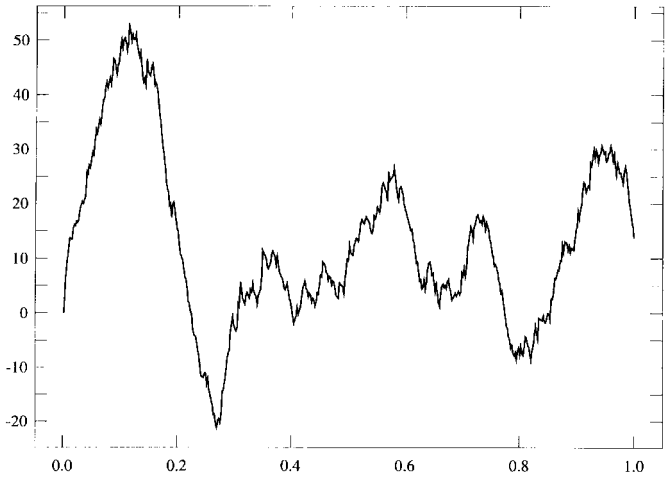


Fig. 1. Shot noise $z(t)$ from Example 1.

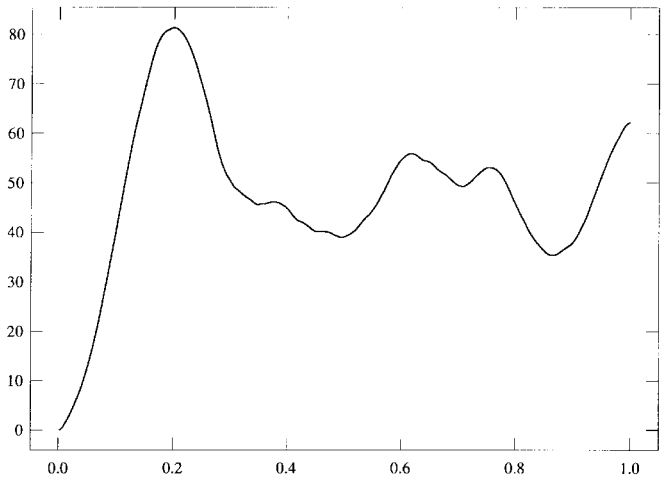


Fig. 2. Shot noise $z(t)$ from Example 2.

Blind deconvolution problems arise in a variety of contexts, and there is a considerable literature on this subject, e.g., [4], [6], [8], [12]–[16], [21], [23], [25], [28], [29]. Because these references consider a different model from the one considered here, the methods developed therein are not tailored for shot-noise data. Therefore, we develop a new method for addressing the blind deconvolution problem when the data can be modeled

Manuscript received June 13, 1995; revised August 5, 1996. This work was supported by the Air Force Office of Scientific Research under Grant F49620–92–J–0305. The associate editor coordinating the review of this paper and approving it for publication was Dr. Ananthram Swami.

The authors are with the Electrical and Computer Engineering Department, University of Wisconsin, Madison, WI 53706–1691 USA (e-mail: gubner@engr.wisc.edu).

Publisher Item Identifier S 1053-587X(97)01190-2.

as shot noise (1.1). When λ is constant (the homogeneous case), z is a stationary process, and the estimation of h and λ can then be addressed via spectral analysis; this special case has been addressed, e.g., in [22] and [30].

This paper is organized as follows. In the next section, we obtain an estimate of a histogram of the $\{T_\nu\}$. We consider the two cases $h(0) > 0$ and $h(0) = 0$. A discrete version of the problem is stated in Section II-C, and the estimate of h is then obtained as an approximate solution to the resulting discrete convolution equation. Numerical examples are presented in Section II-C1. The estimate of λ is obtained in Section II-D, and numerical examples are given in Section II-D3. The results of Section II-D motivate an approach, other than the one presented in [24], for estimating λ for the case when h is known; this is addressed in Section II-E. The conclusions appear in Section III.

II. The Joint Estimation of h and λ

Our only explicit assumption about λ is that $\lambda(t) = 0$ for $t < 0$; hence, the occurrence times T_ν are positive with probability one. For $t \geq 0$, let N_t denote the number of T_ν that occur in the interval $(0, t]$. Then, for $t \geq 0$,

$$z(t) = \sum_{\nu=1}^{N_t} h(t - T_\nu) = \int h(t-\tau)x(\tau)\,d\tau$$

where $x(\tau) := \sum_{\nu=1}^{N_t} \delta(\tau - T_\nu)$. If $N_t = 0$, we take $z(\tau) = x(\tau) = 0$ for $0 \leq \tau \leq t$.

In this paper, we restrict our attention to two classes of impulse response. We say an impulse response h is of class one if it is causal, continuous on $(0, \infty)$, and right continuous at 0 with $h(0) \neq 0$; without loss of generality, we take $h(0) > 0$. We say that h is a class-two impulse response if it is the integral of a class-one impulse response.

Our estimation of h and λ will be based on K samples of z, say, $z(kt_0)$, $k = 1, \ldots, K$, where $t_0 > 0$ is the sampling period. For class-one impulse responses h, we assume that t_0 is chosen small enough that the modulus of continuity $\omega(h; t_0) := \max_{t \geq 0} |h(t+t_0) - h(t)|$ satisfies $\omega(h; t_0) \ll h(0)$. For a class-two impulse response h, we assume that t_0 is chosen small enough that $\omega(\dot{h}; t_0) \ll \dot{h}(0)$, where $\dot{h}$ is the derivative of h.

In order to estimate h and λ, we also need to estimate the impulse train x. Because the data is sampled, what we estimate is, as explained below, a histogram of the T_ν.

Notation: Let $T := Kt_0$. A vector in $\mathbb{R}^K$ is denoted $\mathbf{r} := [\mathbf{r}(1), \ldots, \mathbf{r}(K)]'$. If r is a function of time, then $\mathbf{r}$ denotes the vector of samples, i.e., $\mathbf{r}(k) := r(kt_0)$. There is one exception—the histogram $\mathbf{x}$—where $\mathbf{x}(k) := N_{kt_0} - N_{(k-1)t_0}$ is the number of events from x that occur in the interval $I_k := ((k-1)t_0, kt_0]$. If r is a linear combination of some linearly independent functions $b_1, \ldots, b_M$, we let $\underline{r} := [\underline{r}(1), \ldots, \underline{r}(M)]' \in \mathbb{R}^M$ be the vector of coefficients such that $r = \sum_{m=1}^{M} \underline{r}(m) b_m$; in this case, $\mathbf{r} = [\mathbf{b}_1, \ldots, \mathbf{b}_M]\underline{r}$.

We are interested in estimating h by a function $\hat{h} \in \text{span}\{b_m\}_{m=1}^M$; our task, therefore, is to find its coordinate vector $\underline{\hat{h}}$. The estimate of λ will be expressed as a kernel intensity estimate. We also explain how $\hat{h}$ can be used to estimate λ with the regularization approach of [24].

A. The Estimate of $\mathbf{x}$ When h is of Class One

In this case, we write $h(t) = h(0)u(t) + h_1(t)$, where u is the unit step function, $u(t) = 1$ for $t \geq 0$ and $u(t) = 0$ for $t < 0$, and where $h_1(t) := h(t) - h(0)u(t)$; hence, h_1 is causal and since $h_1(0) = 0$, h_1 is continuous. We can then rewrite $z(t)$ as

$$\begin{aligned} z(t) &= \sum_{\nu=1}^{N_t} h(0)u(t - T_\nu) + \sum_{\nu=1}^{N_t} h_1(t - T_\nu) \\ &= h(0)N_t + y_1(t) \end{aligned} \tag{2.1}$$

where

$$y_1(t) := \sum_{\nu=1}^{N_t} h_1(t - T_\nu) \tag{2.2}$$

and can be viewed as shot noise with impulse response $h_1(t)$. Since h_1 is continuous, one might think of recovering the T_ν by finding the discontinuities of z, i.e., by finding those points where $\dot{z}(t)$ contains an impulse; however, since only finitely many samples of z are available, namely, the entries of $\mathbf{z} = [z(t_0), \ldots, z(Kt_0)]'$, this approach cannot be undertaken. We proceed by taking differences, which are the discrete analog of differentiation.

Consider the first and second differences

$$\mathbf{d}_1(k) := \begin{cases} \mathbf{z}(1), & k = 1, \\ \mathbf{z}(k) - \mathbf{z}(k-1), & k \geq 2, \end{cases}$$

and

$$\mathbf{d}_2(k) := \begin{cases} \mathbf{d}_1(1), & k = 1, \\ \mathbf{d}_1(k) - \mathbf{d}_1(k-1), & k \geq 2. \end{cases}$$

In this case, $\mathbf{d}_1(1) = h(0)\mathbf{x}(1) + y_1(t_0)$, and

$$\mathbf{d}_1(k) = h(0)\mathbf{x}(k) + y_1(kt_0) - y_1((k-1)t_0), \quad k \geq 2.$$

Fig. 3 shows a plot of $\mathbf{d}_1(k)$ from Example 1 in Section II-D3 below; the lower-level pattern corresponds to those k for which $\mathbf{x}(k) = 0$. If the random process $y_1(t)$ in (2.2) satisfies (A.1) in Appendix A, then

$$\mathbf{d}_2(k) \approx \begin{cases} h(0)\mathbf{x}(1), & k = 1, \\ h(0)(\mathbf{x}(k) - \mathbf{x}(k-1)), & k \geq 2 \end{cases} \tag{2.3}$$

and the $\mathbf{d}_2(k)$ appear to be approximately on a lattice. Fig. 4 shows a plot of $\mathbf{d}_2(k)$ from Example 1; in this example, (A.1) is true, and hence, the $\mathbf{d}_2(k)$ satisfy (2.3).

Observe now that if $q \approx h(0)$ is an estimate of $h(0)$, then $h(0)\mathbf{x}(1)$ would result from rounding $\mathbf{d}_2(1)$ to the nearest integral multiple of q. This rounding operation can be expressed as $qQ_q(\mathbf{d}_2(1))$, where

$$Q_q(r) := \lfloor r/q + 0.5 \rfloor, \quad r \in \mathbb{R}.$$

$Q_q(r)$ gives the units for quantizing r to the nearest integral multiple of the quantum q. Hence, if $q \approx h(0)$, $\mathbf{x}$ can be recovered from $\mathbf{d}_2$ since $\mathbf{x}(1) = Q_q(\mathbf{d}_2(1))$, and $\mathbf{x}(k) =$

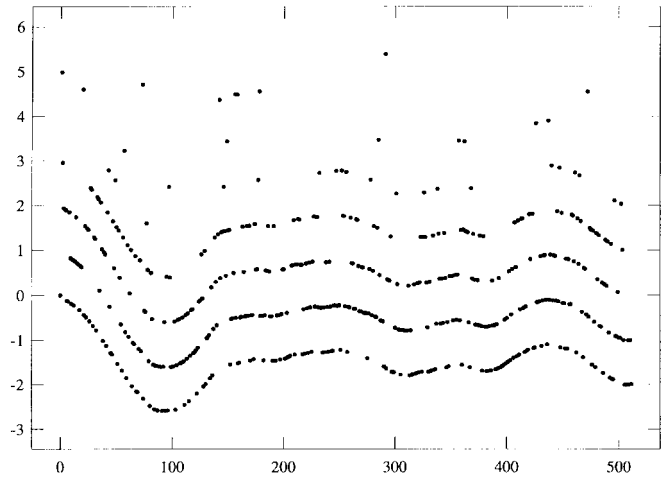


Fig. 3. Example 1: $\mathbf{d}_1(k)$.

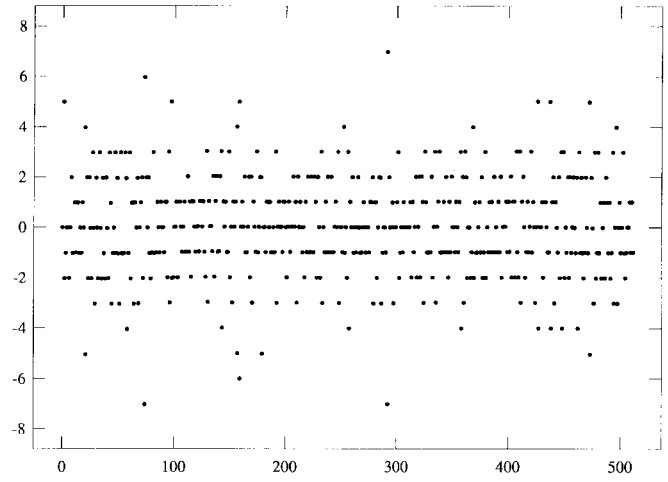


Fig. 4. Example 1: $\mathbf{d}_2(k)$.

$Q_q(\mathbf{d}_2(k))+\mathbf{x}(k-1)$ for $k \geq 2$. This motivates our estimating $\mathbf{x}$ by

$$\hat{\mathbf{x}}(k) := \begin{cases} Q_q(\mathbf{d}_2(1)), & k=1, \\ \max\{Q_q(\mathbf{d}_2(k)) + \hat{\mathbf{x}}(k-1), 0\}, & k \geq 2. \end{cases} \quad (2.4)$$

From (2.3), we see that if $q \approx h(0)/n$, for some $n \in \mathbb{N}$, then $\hat{\mathbf{x}} = n\mathbf{x}$ would result. The details on how to choose q appear in Appendix A.

B. The Estimate of $\mathbf{x}$ *When* h *Is of Class Two*

In this case, h can be written as $h(t) = \int_0^t h_1(\tau)d\tau$ for some class-one function h_1 satisfying $h_1(0) \gg \omega(h_1; t_0)$; in turn, we write $h_1(\tau) = h_1(0)u(\tau) + h_2(\tau)$, where $h_2(\tau) := h_1(\tau) - h_1(0)u(\tau)$. We can then write

$$z(t) = \sum_{\nu=1}^{N_t} h_1(0)(t - T_\nu)u(t - T_\nu) + y_2(t) \quad (2.5)$$

where

$$y_2(t) := \sum_{\nu=1}^{N_t} \int_0^{t-T_\nu} h_2(\tau)d\tau \quad (2.6)$$

and can be viewed as shot noise with impulse response $\int_0^t h_2(\tau)d\tau$. Since h_2 is continuous, one might think of recovering the T_ν by finding the discontinuities in $\dot{z}(t)$, i.e., by finding those points where $\ddot{z}(t)$ contains an impulse; this approach cannot be undertaken since only a finite set of samples from z is available. We again proceed by taking differences. Figs. 5 and 6 show the plots of $\mathbf{d}_1(k)$ and $\mathbf{d}_2(k)$ from Example 2 in Section II-D3; the smooth trend below the $\mathbf{d}_2(k)$ corresponds to $\mathbf{s}_2(k)$, where

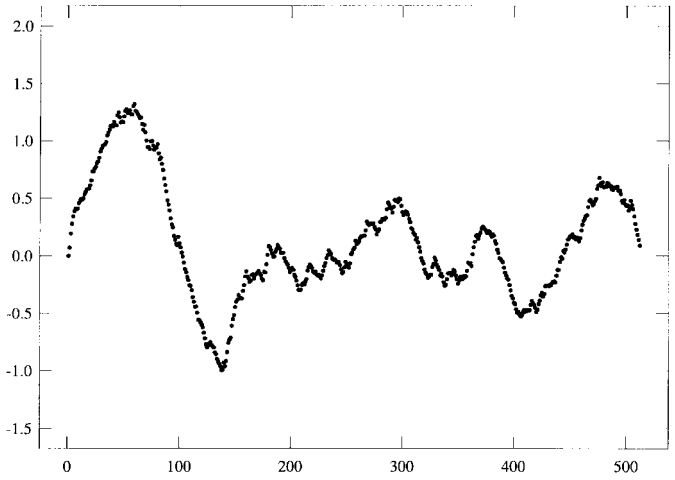


Fig. 5. Example 2: $\mathbf{d}_1(k)$.

$$s_2(t) := y_2(t) - 2y_2(t - t_0) + y_2(t - 2t_0). \quad (2.7)$$

Notice that no pattern is apparent in these plots, in contrast with Figs. 3 and 4, where $h(0) > 0$. This makes the problem of estimating $\mathbf{x}$ more difficult now than in the case of Section II-A. As explained in Appendix B

$$\mathbf{d}_2(k) - \mathbf{s}_2(k) \in [0, t_0 h_1(0)(\mathbf{x}(k-1) + \mathbf{x}(k))).$$

Our approach for obtaining $\hat{\mathbf{x}}$ is to first obtain $\hat{\mathbf{s}}_2$, which is an estimate of $\mathbf{s}_2$, and then take

$$\hat{\mathbf{x}}(k) = Q_q(\mathbf{d}_2(k) - \hat{\mathbf{s}}_2(k)) \quad (2.8)$$

where the quantum q is now an estimate of $t_0 h_1(0)$; notice that $\hat{\mathbf{x}}|_{q/n} \approx n\hat{\mathbf{x}}|_q$ is expected (we write $\hat{\mathbf{x}}|_q$ to stress the dependence of $\hat{\mathbf{x}}$ on q). The details for obtaining $\hat{\mathbf{s}}_2$ and choosing a value for q are given in Appendix B.

C. The Estimate of h

Recall that we are interested in estimating h by

$$\hat{h} = \sum_{m=1}^{M} \hat{\underline{h}}(m) b_m. \quad (2.9)$$

In this section, we find the coordinate vector $\hat{\underline{h}}$. We first obtain a discrete formulation of the problem; we use $*$ to denote the discrete convolution operation

$$(\mathbf{p} * \mathbf{q})(k) := \sum_{\ell=1}^{k} \mathbf{p}(\ell)\mathbf{q}(k - \ell + 1), \quad k = 1, \ldots, K$$

and we use $L(\mathbf{p})$ to denote the lower-triangular $K \times K$ matrix

$$L(\mathbf{p}) := \begin{bmatrix} \mathbf{p}(1) & & \\ \vdots & \ddots & \\ \mathbf{p}(K) & \cdots & \mathbf{p}(1) \end{bmatrix}.$$

Note that $\mathbf{p} * \mathbf{q} = L(\mathbf{p})\mathbf{q}$.

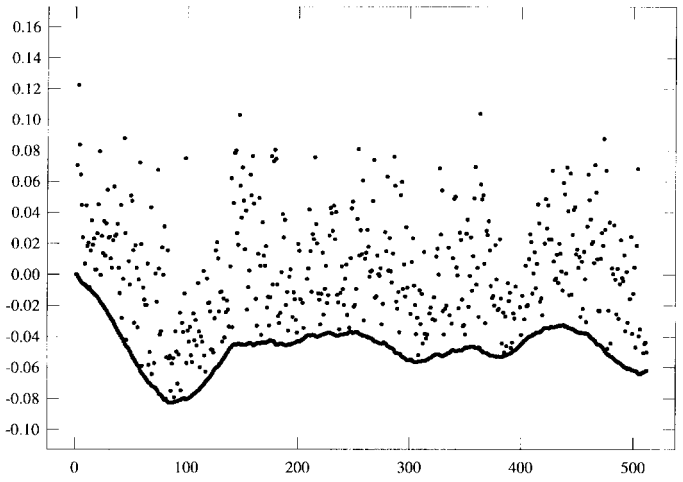


Fig. 6. Example 2: $\mathbf{d}_2(k)$ and $\mathbf{s}_2(k)$ (smooth trend at the bottom).

At the beginning of Section II, we wrote the data as $z(t) = \int h(t-\tau)x(\tau)d\tau$ and $x(\tau) := \sum_{\nu=1}^{N_t} \delta(\tau - T_\nu)$; observe that if the T_ν were integral multiples of t_0, then $\mathbf{z} = \mathbf{x} * \mathbf{h} = L(\mathbf{x})\mathbf{h}$ would hold. In view of (2.9), $\hat{\mathbf{h}} = [\mathbf{b}_1, \ldots, \mathbf{b}_M]\underline{\hat{h}}$; we now obtain $\hat{\mathbf{h}}$ as a function of $\mathbf{z}$ and $\hat{\mathbf{x}}$. Notice that since the T_ν do not occur on a lattice, it is not advisable to let $\hat{\mathbf{h}} \approx L(\hat{\mathbf{x}})^\dagger \mathbf{z}$ even if $\hat{\mathbf{x}} = \mathbf{x}$ (here, $L(\hat{\mathbf{x}})^\dagger$ denotes the pseudoinverse of the matrix $L(\hat{\mathbf{x}})$). Instead, we proceed by obtaining a *regularized solution* to

$$\mathbf{z} = L(\hat{\mathbf{x}})\mathbf{r} = \hat{\mathbf{x}} * [\mathbf{b}_1, \ldots, \mathbf{b}_M]\underline{r} = V\underline{r}$$

where $V := [\hat{\mathbf{x}} * \mathbf{b}_1, \ldots, \hat{\mathbf{x}} * \mathbf{b}_M]$; we take

$$\underline{\hat{h}} = \arg\min_{\underline{r} \in \mathbb{R}^M} \left\{ ||\mathbf{z} - V\underline{r}||_2^2 + \beta ||\underline{r}||_E^2 \right\}$$

for some $\beta \geq 0$ (to be selected in Appendix C), where $||\underline{r}||_E^2 := \underline{r}' E \underline{r}$, and $E_{m,n} := \mathbf{b}_m' \mathbf{b}_n$ (this norm was introduced to approximate the L^2 norm of $r = \sum_{m=1}^M \underline{r}(m) b_m$). An easy calculation shows that

$$\underline{\hat{h}} = (V'V + \beta E)^{-1} V' \mathbf{z}$$

which is the solution to an $M \times M$ system of equations.

1) Numerical Examples: In this section, we present numerical examples of the estimation of the impulse response h for different h and different λ. These examples are continued in Section II-D3 below, where we address the estimation of the intensity λ. The shot-noise data was obtained by generating $\Lambda := \int_0^T \lambda(\tau)d\tau = \mathsf{E}[N_T]$ independent random points T_ν over $[0, T] = [0, 1]$ with density $\lambda(\tau)/\Lambda$, $\tau \in [0, 1]$ and then using (1.1) to obtain z. We took $K = 512$ samples of z over $[0, 1]$ with $t_0 = 1/512$ to form the data vector $\mathbf{z}$.

We took $\hat{h} \in \text{span}\{b_m\}_{m=1}^{13}$, where the b_m are the $M = 13$ cubic B-splines [5] generated by the NAG [18] subroutine E02BCF when there are four coincident knots at both end points of $[0, 1]$ and nine uniformly spaced internal knots at $0.1, \ldots, 0.9$.

In Examples 1–3, we used point-process data corresponding to the high-variation λ shown later in Fig. 15 (dotted line). In the first example, we considered a class-one impulse response h with a discontinuity at the origin; in Examples 2 and 3, h was a continuous class-two impulse response. The parameters Λ, r_1, q, $||\hat{\mathbf{x}}||_1$, β, w, c, and α for each example are given in Table I. (The significance of r_1 is explained in Appendix B, and that of w, c, and α is explained in Section II-D.)

TABLE I
PARAMETERS FOR EXAMPLES 1–7

Example	Λ	r_1	q	$\|\hat{\mathbf{x}}\|_1$	β	w	c	α
1	800		1.002	800	0	3.714(−2)	–	–
2	800	2.012(−3)	2.456(−2)	934	98.81	3.327(−2)	–	–
3	800	6.255(−4)	1.478(−2)	949	75.68	3.714(−2)		–
4	480		1.004	480	0	5.265(−2)	10	1.398(−6)
5	480	1.253(−3)	2.912(−2)	502	14.49	5.265(−2)	10	9.265(−7)
6	480	4.004(−4)	1.756(−2)	502	15.81	5.265(−2)	5.6(−2)	2.269(−4)
7	50	–	1.002	50	0	3.327(−2)	10	1.884(−6)

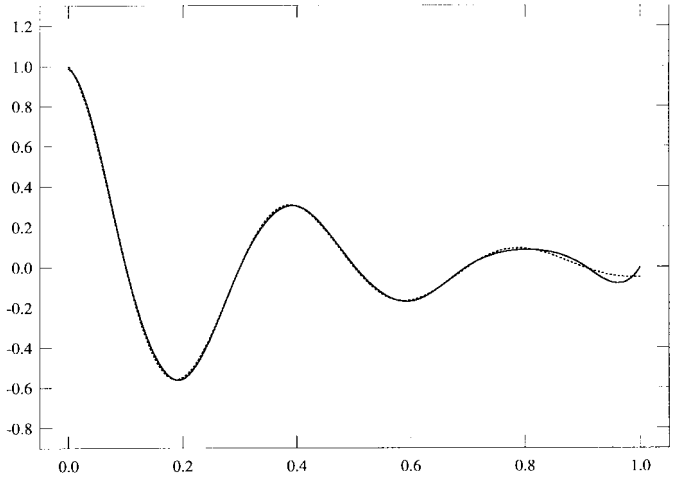


Fig. 7. Example 1: $\hat{h}(t)$ (solid line) and $h(t)$ (dotted line). (High-variation λ.)

Example 1: We begin with a class-one impulse response h with a jump at the origin $h(t) = \exp(-3t)\cos(5\pi t)$, $t \geq 0$. The shot noise data for this example was shown earlier in Fig. 1. We obtained $\hat{\mathbf{x}}$ following the procedure in Section II-A. The first and second differences $\mathbf{d}_1$ and $\mathbf{d}_2$ were shown earlier in Figs. 3 and 4. The estimate $\hat{h}$ is shown in Fig. 7 along with the function h.

Example 2: We now consider the continuous class-two impulse response $h(t) = \exp(-3t)\sin(5\pi t)$, $t \geq 0$. The shot-noise data for this example was shown earlier in Fig. 2. We estimated $\hat{\mathbf{x}}$ as explained in Section II-B. The differences $\mathbf{d}_1$ and $\mathbf{d}_2$ were shown earlier in Figs. 5 and 6. The estimate $\hat{h}$ is shown in Fig. 8 along with the function h.

Example 3: In the previous example, the factor $\sin(5\pi t)$ in $h(t)$ caused h to have a lot of oscillation. In this example, we changed $\sin(5\pi t)$ to $\sin(3\pi t)$; the reduced oscillation allowed us to get a better estimate of h. We took $h(t) = \exp(-3t)\sin(3\pi t)$, $t \geq 0$ (class-two h). The estimate $\hat{h}$ is shown in Fig. 9 along with h. To gain a general sense of the behavior of the estimates to varying the shot-noise data, the estimates of h obtained from five realizations of $z(t)$ with $N_1 = \Lambda = 800$ are shown in Fig. 10. Observe that these estimates show all the features in h, where some have minor oscillations near the tail, and that only one of the estimates is over smoothed.

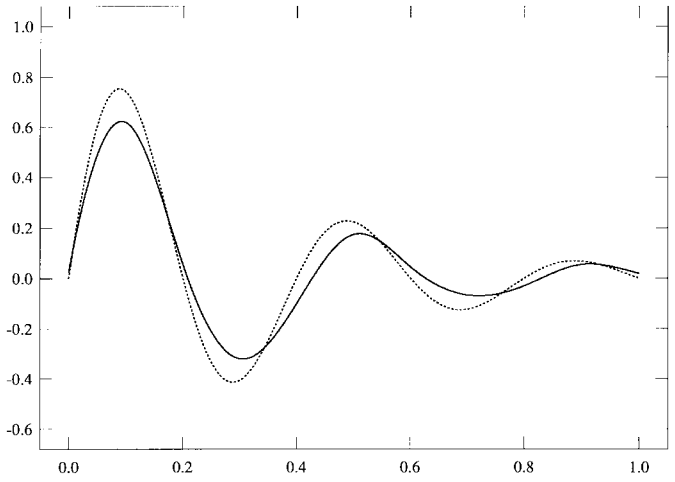


Fig. 8. Example 2: $\hat{h}(t)$ (solid line) and $h(t)$ (dotted line). (High-variation λ.)

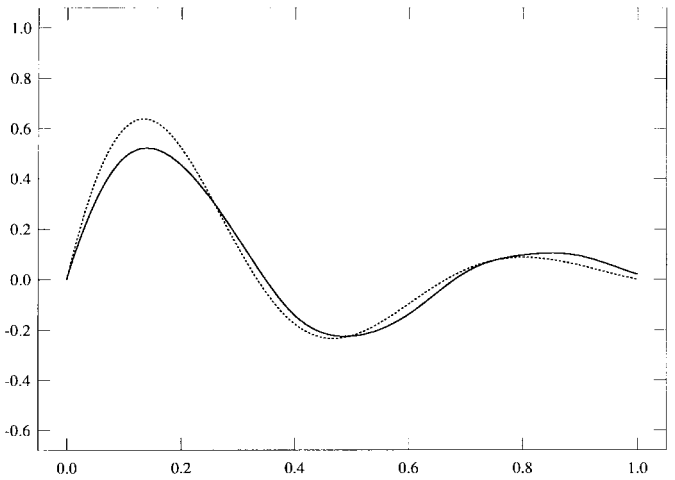


Fig. 9. Example 3: $\hat{h}(t)$ (solid line) and $h(t)$ (dotted line). (High-variation λ.)

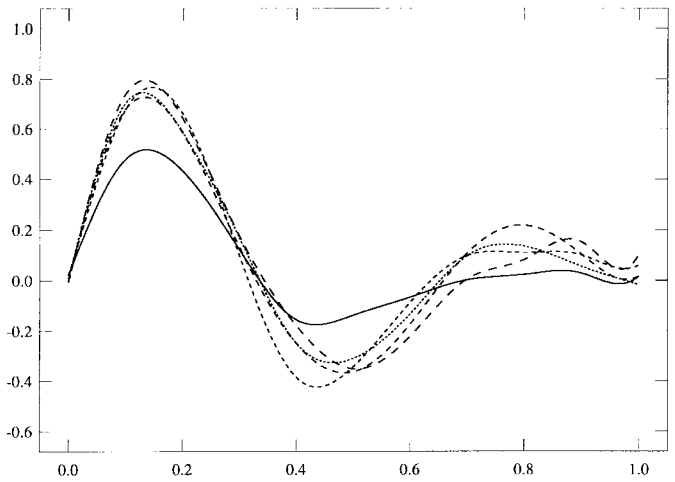


Fig. 10. Example 3: The estimates of h obtained from different realizations of $z(t)$. (High-variation λ.)

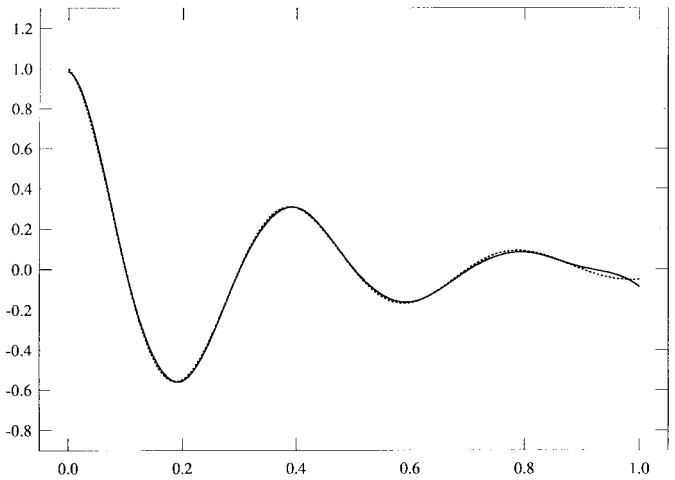


Fig. 11. Example 4: $\hat{h}(t)$ (solid line) and $h(t)$ (dotted line). (Two-peak λ.)

In Examples 4–6 below, the point-process data corresponds to the medium-variation, two-peak λ shown later in Fig. 21 (dotted line). We considered the same impulse responses h as in the previous examples to compare the method of estimating h to varying λ (i.e., to varying the number of random points T_ν as well as their distribution). The shot-noise data is not shown. The relevant parameters are summarized in Table I.

Example 4: In this case, $h(t) = \exp(-3t)\cos(5\pi t)$, $t \geq 0$, as in Example 1. The estimate $\hat{h}$ is shown in Fig. 11, along with the impulse response h. This $\hat{h}$ looks very similar to the one shown in Fig. 7, i.e., varying λ had little effect on the quality of the estimate.

Example 5: In this case, $h(t) = \exp(-3t)\sin(5\pi t)$, $t \geq 0$, as in Example 2. The function h and the estimate $\hat{h}$ are shown in Fig. 12. Comparing this estimate with the one shown in Fig. 8, we see that a better estimate of h resulted when Λ was higher.

Example 6: Here, $h(t) = \exp(-3t)\sin(3\pi t)$, $t \geq 0$, as in Example 3. The estimate $\hat{h}$ is shown in Fig. 13 along with h. This $\hat{h}$, like that in the previous example, shows a spurious oscillation near the tail, which is not present in the estimate obtained in Example 3.

D. The Estimate of λ

We now obtain the estimate $\hat{\lambda}$ of the intensity λ; we consider a *kernel intensity estimate* modified for histogram data. For the purpose of comparison, we also consider a version of the approach presented in [24], which is useful for the case where Λ is small.

1) The Kernel-Estimate Approach: If $\{T_\nu\}_{\nu=1}^{N_T}$ were available, a natural approach for estimating λ would be to write the estimate $\hat{\lambda}$ as a kernel intensity estimate. A modification of this method is carried out below to handle histogram data.

Recall that $T := Kt_0$, and let $n := N_T$. A kernel intensity estimate of λ is a function of the form

$$\hat{\lambda}(\tau) = w^{-1} \sum_{\nu=1}^{n} \eta((\tau - T_\nu)/w), \quad 0 \leq \tau \leq T$$

where the *kernel* η is usually a symmetric probability density function, and $w > 0$ is the *kernel width*.

Note that when η integrates to one, $\int_{-\infty}^{\infty} \hat{\lambda}(\tau)d\tau = n$. However, the true intensity λ satisfies $\int_0^T \lambda(\tau)d\tau = \Lambda = \mathsf{E}[N_T]$. Hence, if the particular Poisson-process realization we are dealing with contains a number of points that is much

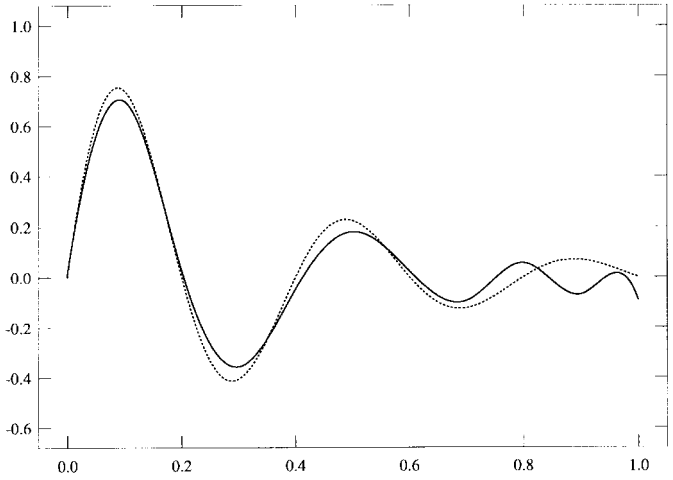


Fig. 12. Example 5: $\hat{h}(t)$ (solid line) and $h(t)$ (dotted line). (Two-peak λ.)

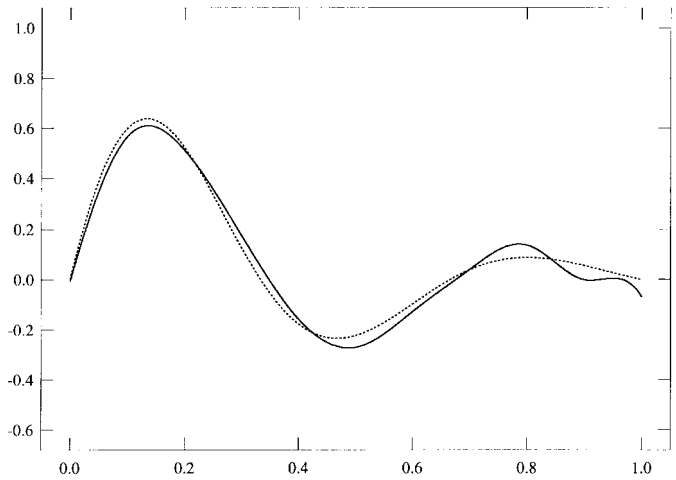


Fig. 13. Example 6: $\hat{h}(t)$ (solid line) and $h(t)$ (dotted line). (Two-peak λ.)

different from the expected number of points, the intensity estimate $\hat{\lambda}$ will lie above or below the true intensity.

A common optimality criterion for choosing η and w, in the context of density estimation, is to minimize the *mean integrated squared error* (MISE) $\mathsf{E}[\int(\lambda-\hat{\lambda})^2]$. On the basis of this criterion, there is a large amount of flexibility in the choice of η, but the choice of w is a delicate matter [26]. Given η, a popular method for choosing w is *least-squares cross-validation* [26], which consists of choosing w as the minimizer of

$$\hat{R}(w) := \frac{1}{n^2 w}\sum_{\nu=1}^{n}\sum_{\rho=1}^{n}\eta^{(2)}((T_\nu - T_\rho)/w) - \frac{2}{n(n-1)w}\sum_{\nu=1}^{n}\sum_{\substack{\rho=1\\ \rho\neq\nu}}^{n}\eta((T_\nu - T_\rho)/w)$$

where $\eta^{(2)}$ denotes the convolution of η with itself. $\hat{R}(w)$ is an unbiased estimate of $\mathsf{E}[\int(\hat{\lambda}^2 - 2\hat{\lambda}\lambda)]$, i.e., of the part of the MISE involving $\hat{\lambda}$.

Since neither n nor $\{T_\nu\}_{\nu=1}^n$ is available, we use instead the estimates $\hat{n} := \|\hat{\mathbf{x}}\|_1$ and the histogram $\hat{\mathbf{x}}$, and we choose w as the minimizer of

$$\tilde{R}(w) := \frac{1}{\hat{n}^2 w}\sum_{k=1}^{K}\sum_{\ell=1}^{K}\hat{\mathbf{x}}(k)\hat{\mathbf{x}}(\ell)\eta^{(2)}((k-\ell)t_0/w) - \frac{2}{\hat{n}(\hat{n}-1)w}\sum_{k=1}^{K}\sum_{\substack{\ell=1\\ \ell\neq k}}^{K}\hat{\mathbf{x}}(k)\hat{\mathbf{x}}(\ell)\eta((k-\ell)t_0/w).$$

In spite of the similarity between $\tilde{R}$ and $\hat{R}$, this method of choosing w should be used with caution when dealing with discretized data; see [26, pp. 51 and 52] for details. A typical $\tilde{R}(w)$ curve is shown in Fig. 14.

Kernel estimates are known to behave poorly near the boundaries of the data interval if the underlying intensity or density is not zero at the boundaries. Since we have data only over the interval $[0,T]$, and we are interested in estimating $\lambda(\tau)$ for $\tau \in [0,T]$, we need to make some boundary adjustments to improve the quality of the estimate near the end points. To this end, and to account for the data being histogram data (as opposed to exact data), we modify slightly the approach taken in [9]: Once w is chosen, we take

$$\hat{\lambda}(\tau) := w^{-1}\sum_{k=1}^{K}\hat{\mathbf{x}}(k)\{\eta((\tau-kt_0)/w)+\eta((\tau+kt_0)/w) + \eta((\tau-2T+kt_0)/w)\} \tag{2.10}$$

for $\tau \in [0,T]$. The end point corrections consist of appending symmetrical versions of $\hat{\mathbf{x}}$ to $\hat{\mathbf{x}}$, the symmetry being with respect to either end point. The first $\eta(\cdot)$ term in the expression for $\hat{\lambda}$ is the usual term; the second and third ones correspond, respectively, to corrections at 0 and at T.

Remark: In our examples, η is symmetric and integrates to one. Hence, the kernel intensity estimate in (2.10) satisfies $\int_0^T \hat{\lambda}(\tau)d\tau = \sum_{k=1}^K \hat{\mathbf{x}}(k) = \|\hat{\mathbf{x}}\|_1$. However, the true intensity λ satisfies $\int_0^T \lambda(\tau)d\tau = \Lambda = \mathsf{E}[N_T]$. Therefore, if $\|\hat{\mathbf{x}}\|_1$ is much different from the expected number of points Λ, $\hat{\lambda}$ will lie above or below the true intensity λ; see, e.g., Figs. 18 and 19.

2) The Regularization Approach: We now consider obtaining $\hat{\lambda}$ with a version of the method presented in [24]. The

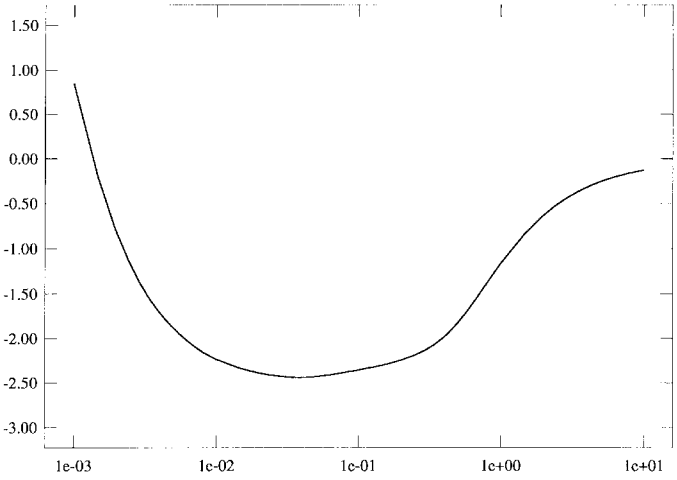


Fig. 14. $\tilde{R}(w)$ for Example 1.

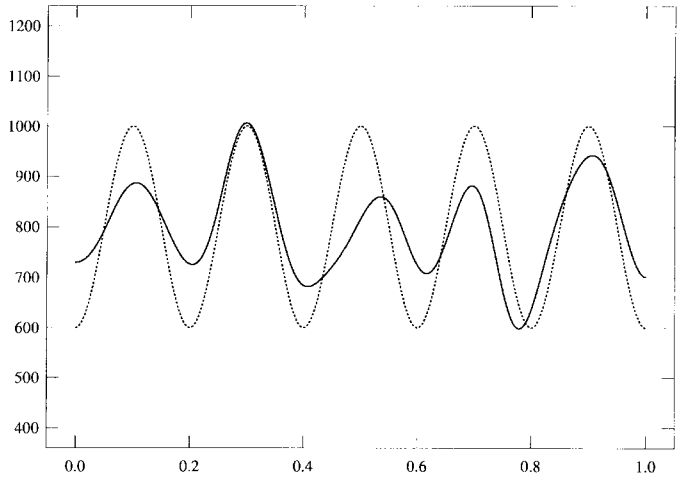


Fig. 15. Example 1 (continued): Kernel estimate $\hat{\lambda}(t)$ (solid line) and $\lambda(t)$ (dotted line).

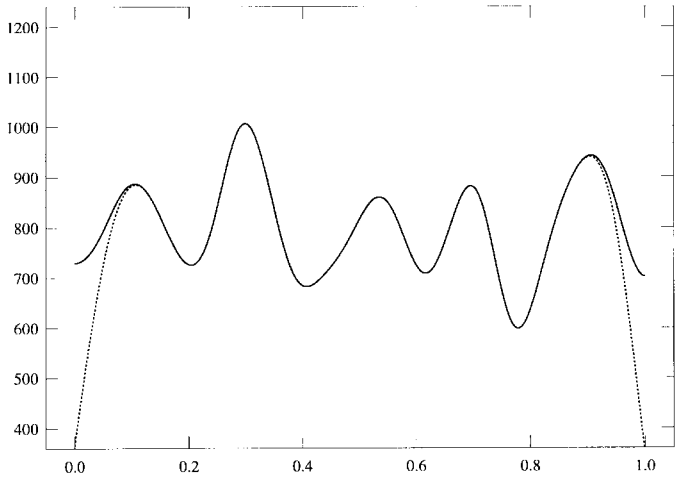


Fig. 16. Example 1 (continued): $\hat{\lambda}(t)$ from Fig. 15 (solid line) and the kernel estimate without end corrections (dotted line).

estimate is of the form

$$\hat{\lambda} = \sum_{m=1}^{M'} \hat{\underline{\lambda}}(m) B_m$$

where the B_m are linearly independent functions; our task is to obtain the coordinate vector $\hat{\underline{\lambda}}$ (we use the underbar notation to denote a vector in either $\mathbb{R}^M$ or $\mathbb{R}^{M'}$; it should be clear from the context what is meant).

The first step for obtaining $\hat{\underline{\lambda}}$ is to choose values for the pair of regularization parameters c and α in [24]. To choose c, we examine a plot of either $P_i(\cdot), i = 1, 3$, of [24, Section VI] and then take $\alpha = P_i(c)$. In our numerical examples below, we examine $P_1(\cdot)$. Since h is not known, we substitute our estimate $\hat{h}$ in place of h in the equations involving h. Hence, we can expect results that are comparable with those in [24] if $\hat{h}$ is close to h.

As in [24], let $\tilde{\underline{\lambda}} \in \mathbb{R}^{M'}$ denote the right-hand side of [24, Eq. (2.3)]; to obtain $\hat{\underline{\lambda}}$, we scale $\tilde{\underline{\lambda}}$ as follows. Recall that $\Lambda := \int_0^T \lambda(\tau) d\tau$, and now, let $\hat{\Lambda} := \int_0^T \hat{\lambda}(\tau) d\tau$. Since $\|\mathbf{x}\|_1 = N_T$ is the natural estimate of Λ, we take $\hat{\underline{\lambda}} = \|\hat{\mathbf{x}}\|_1 \tilde{\underline{\lambda}} \,/\, \int_0^T \tilde{\lambda}(\tau) d\tau$ so that $\hat{\Lambda} = \|\hat{\mathbf{x}}\|_1$.

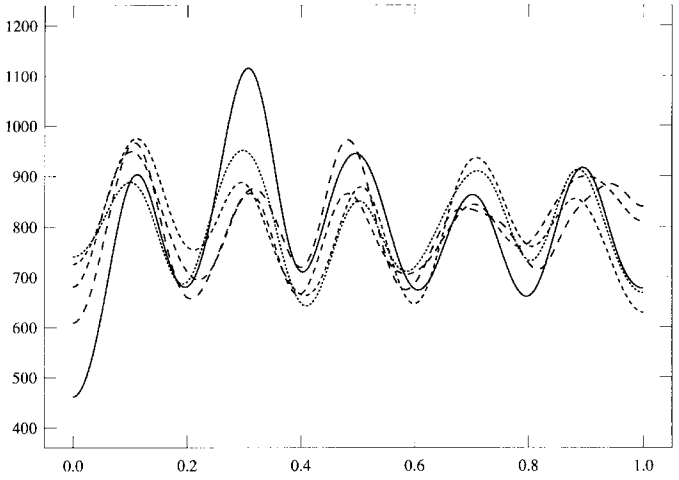


Fig. 17. Example 1 (continued): Kernel estimates of λ obtained from different realizations of $z(t)$.

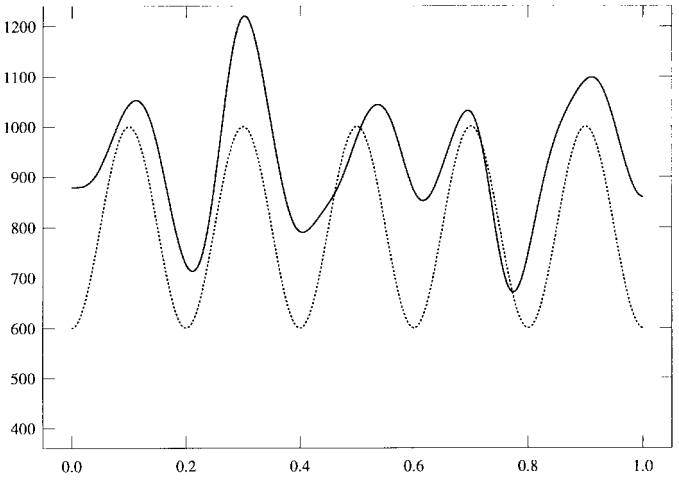


Fig. 18. Example 2 (continued): Kernel estimate $\hat{\lambda}(t)$ (solid line) and $\lambda(t)$ (dotted line).

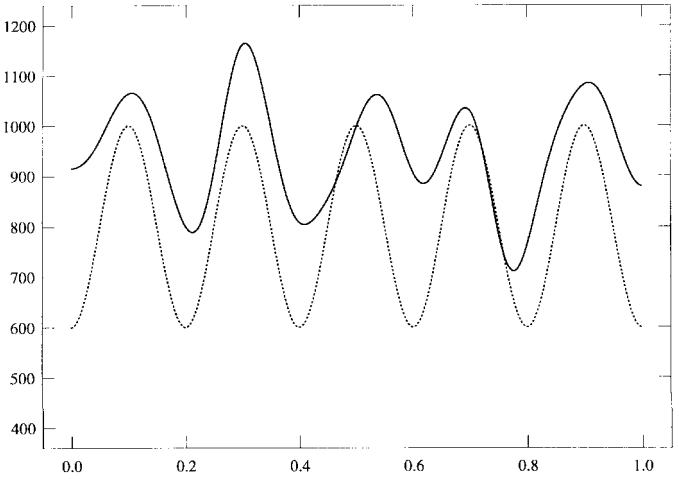


Fig. 19. Example 3 (continued): Kernel estimate $\hat{\lambda}(t)$ (solid line) and $\lambda(t)$ (dotted line).

Typically, a larger number of data samples will be required for the joint estimation of h and λ than for estimating λ when h is known. The matrix $\tilde{\Gamma}$ of [24, Section V] is of size $K \times K$; for computational tractability in the numerical examples below,

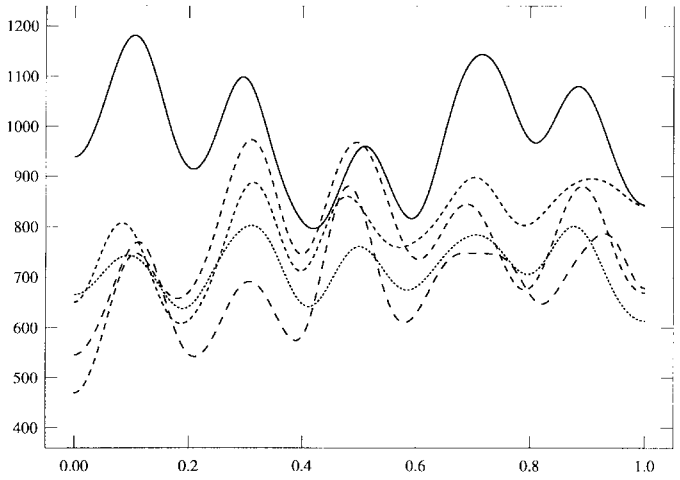


Fig. 20. Example 3: Kernel estimates of λ obtained from different realizations of $z(t)$.

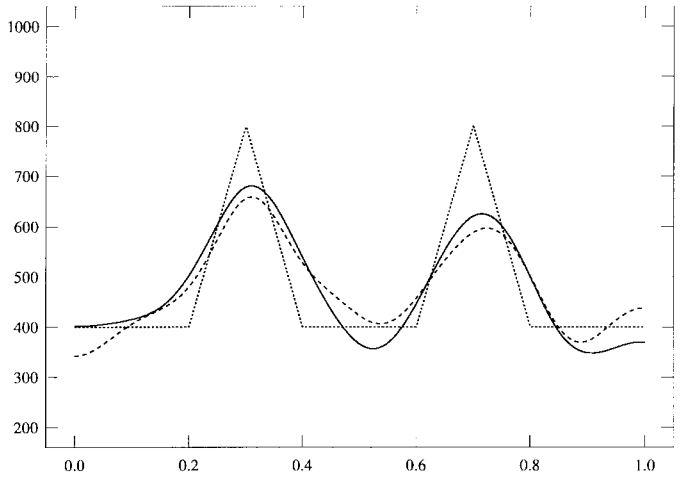


Fig. 21. Example 4: $\hat{\lambda}_{\mathrm{reg}}(t)$ (solid line), $\hat{\lambda}_{\mathrm{ker}}(t)$ (dashed line), and $\lambda(t)$ (dotted line).

once $\hat{\mathbf{x}}$ and $\hat{h}$ are obtained, we use only a few data samples for obtaining $\hat{\lambda}$.

3) Numerical Examples: We now continue the numerical examples in Section II-C1; we obtain the estimate of λ corresponding to the different pairs (h, λ) considered there. The first three examples illustrate the kernel-estimate approach of Section II-D1; the next four illustrate also the approach from [24] adapted in the previous section. Even though the kernel-estimate approach is conceptually simpler and requires fewer computations than the regularization approach, our last example shows that when the number of T_ν is small (low-intensity λ), the former approach may not be as effective as the latter one.

To obtain a kernel estimate of λ, we need to choose a kernel η; for computational convenience, we use the standard normal density $\eta(\tau) = (2\pi)^{-1/2}\exp(-\tau^2/2)$; in this case, $\eta^{(2)}(\tau) = (4\pi)^{-1/2}\exp(-\tau^2/4)$. The kernel width is chosen to minimize $\tilde{R}(\cdot)$, as explained in Section II-D1.

The relevant parameters for these examples are summarized in Table I. We now continue Examples 1–3, in which the point-process data corresponded to the high-variation λ shown in Fig. 15 (dotted line); here, $\lambda(t) = 200\sin((10t - 0.5)\pi) + 800$, $t \geq 0$, and $\Lambda = 800$.

Example 1 (Continued): A plot of $\tilde{R}(\cdot)$ was shown earlier in Fig. 14. The kernel estimate $\hat{\lambda}$ is shown in Fig. 15 along with the function λ. To see the importance of the end corrections for kernel estimates, Fig. 16 shows the two kernel estimates corresponding to $w = 0.037\,14$, with (solid line) and without (dotted line) end corrections. Fig. 17 shows the kernel estimates of λ obtained from five realizations of the shot noise $z(t)$ with $N_1 = \Lambda = 800$. These estimates all show the main features in λ. In each case, $\|\hat{\mathbf{x}}\|_1 = 800$ resulted. The corresponding estimates $\hat{h}$ were all very close to that in Fig. 7 and are not shown.

Example 2 (Continued): The kernel estimate $\hat{\lambda}$ is shown in Fig. 18 along with the function λ. Since our estimate of Λ, $\|\hat{\mathbf{x}}\|_1$ is too high, so is our estimate $\hat{\lambda}$. (Recall the remark following (2.10).)

Example 3 (Continued): The kernel estimate $\hat{\lambda}$ is shown in Fig. 19 along with the intensity λ. Fig. 20 shows the kernel estimates of λ corresponding to the five realizations of z from which resulted the estimates of h shown in Fig. 10. The corresponding $\|\hat{\mathbf{x}}\|_1$ ranged from 688 to 983.

In the next examples, the kernel estimate of λ is called $\hat{\lambda}_{\mathrm{ker}}$, whereas the estimate obtained via the regularization approach of Section II-D2 is called $\hat{\lambda}_{\mathrm{reg}}$. The latter is expressed as a linear combination of the $M' = 21$ cubic B-splines generated by the NAG subroutine E02BCF when there are four coincident knots at both end points of $[0, 1]$ and 17 uniformly spaced internal knots at $0.1, 0.15, \ldots, 0.85, 0.9$.

Since $K = 512$ is too large for the numerical computations in Section II-D2, we used *only* the 64 data samples obtained by subsampling $\mathbf{z}$ every $8k$ units $k = 1, \ldots, 64$ to plot the $P_1(\cdot)$-curve (not shown) as in [24] to choose the pair (c, α). The original 512 data samples were used for obtaining $\hat{\mathbf{x}}$.

Example 4 (Continued): Recall that in this example, h had a jump at the origin and that a very good estimate $\hat{h}$ was obtained; $\hat{h}$ and h were shown in Fig. 11. The two estimates $\hat{\lambda}_{\mathrm{reg}}$ (solid line) and $\hat{\lambda}_{\mathrm{ker}}$ (dashed line) are shown in Fig. 21, along with the intensity λ (dotted line). The two methods for estimating λ gave comparable results, but $\hat{\lambda}_{\mathrm{reg}}$ behaves better near the end points, and it shows better the two peaks of λ.

Example 5 (Continued): The estimate $\hat{h}$ in this case was shown earlier in Fig. 12 (solid line). The two estimates $\hat{\lambda}_{\mathrm{reg}}$ (solid line) and $\hat{\lambda}_{\mathrm{ker}}$ (dashed line) are shown in Fig. 22 along with the intensity λ (dotted line). Again, the two estimates look similar, but $\hat{\lambda}_{\mathrm{reg}}$ behaves better near the end points.

Example 6 (Continued): The estimate $\hat{h}$ in this example was shown earlier in Fig. 13 (solid line). Fig. 23 shows the estimates $\hat{\lambda}_{\mathrm{reg}}$ (solid line) and $\hat{\lambda}_{\mathrm{ker}}$ (dashed line) as well as the intensity λ (dotted line). In this case, $\hat{\lambda}_{\mathrm{reg}}$ still behaves better than $\hat{\lambda}_{\mathrm{ker}}$ near the end points, but it is oversmooth when compared with $\hat{\lambda}_{\mathrm{ker}}$.

The next example shows that the regularization approach can be more effective than the kernel-estimate approach when the number of T_ν is small. The relevant parameters for this example appear in Table I.

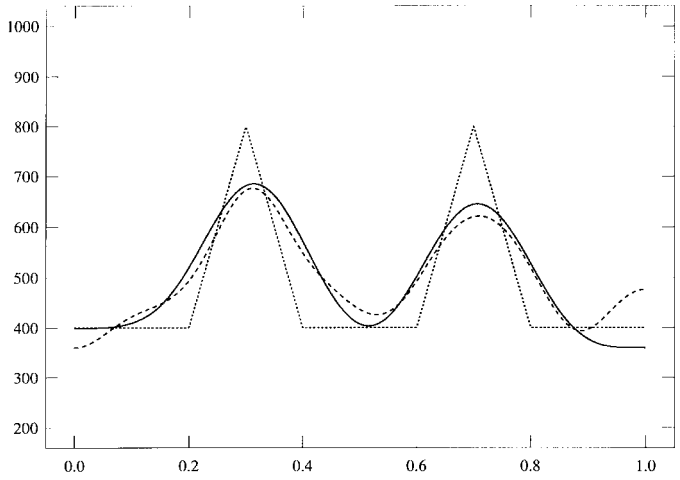


Fig. 22. Example 5: $\hat{\lambda}_{\mathrm{reg}}(t)$ (solid line), $\hat{\lambda}_{\mathrm{ker}}(t)$ (dashed line), and $\lambda(t)$ (dotted line).

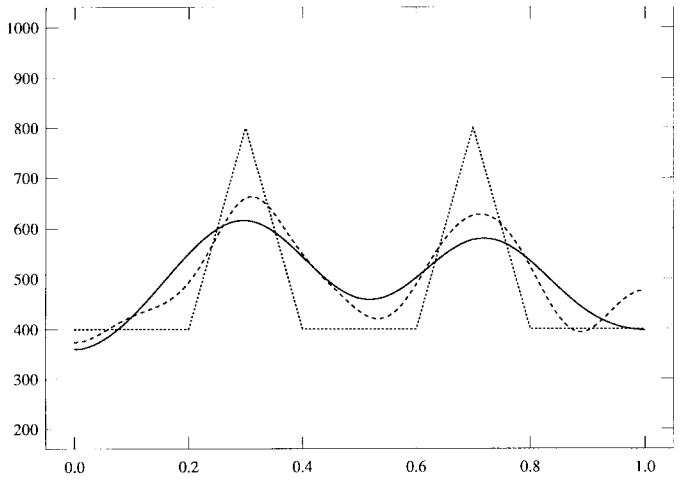


Fig. 23. Example 6: $\hat{\lambda}_{\mathrm{reg}}(t)$ (solid line), $\hat{\lambda}_{\mathrm{ker}}(t)$ (dashed line), and $\lambda(t)$ (dotted line).

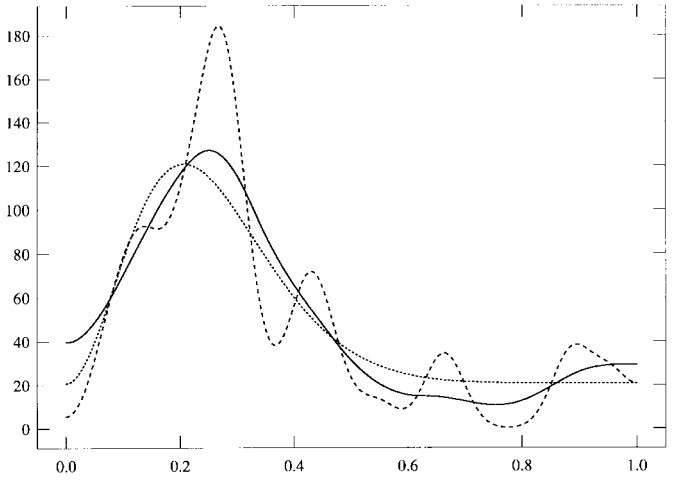


Fig. 24. Example 7: $\hat{\lambda}_{\mathrm{reg}}(t)$ (solid line), $\hat{\lambda}_{\mathrm{ker}}(t)$ (dashed line), and $\lambda(t)$ (dotted line).

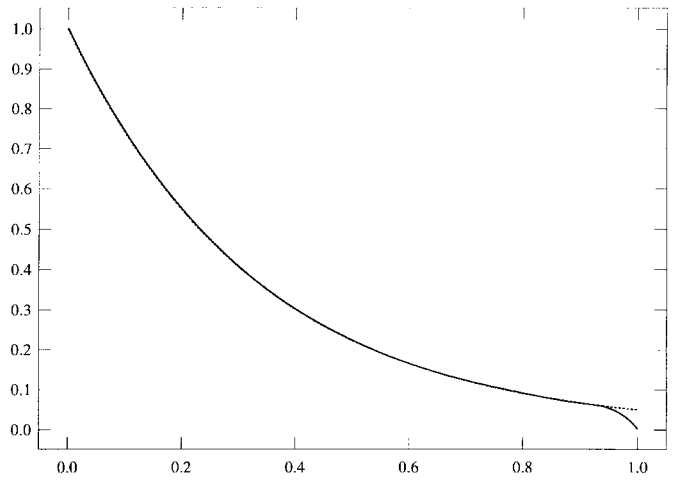


Fig. 25. Example 7: $\hat{h}(t)$ (solid line) and $h(t)$ (dotted line).

Example 7: We considered the low-variation, low-intensity λ shown in Fig. 24 (dotted line); here

$$\lambda(t) = 50\sin((2(1-t)^3 - 0.5)\pi) + 1 + 20.7, \quad t \geq 0$$

and $\Lambda = 50$. We took $h(t) = \exp(-3t)$, $t \geq 0$. Fig. 24 shows the two estimates of λ, $\hat{\lambda}_{\mathrm{reg}}$ (solid line) and $\hat{\lambda}_{\mathrm{ker}}$ (dashed line), and Fig. 25 shows the estimate $\hat{h}$. In this example, $\hat{\lambda}_{\mathrm{reg}}$ is a good estimate of λ, whereas $\hat{\lambda}_{\mathrm{ker}}$ shows too much oscillation.

E. A Kernel Estimate of λ When h Is Known

The results above suggest another approach to the problem considered in [24] of estimating λ when h is known. Since h is known, it should be easy to obtain a good estimate of the histogram $\mathbf{x}$, and $\hat{\lambda}$ can then be obtained as a kernel estimate. To this end, recall that $\mathbf{h}(k) := h(kt_0)$; we let $\hat{\mathbf{x}}$ be the approximate solution, over $\mathbf{r} \in \mathbb{N}^K$, to $\mathbf{z} = \mathbf{h} * \mathbf{r}$ (since h is known, no estimate $\hat{\mathbf{h}}$ is involved) obtained by "back substitution," as explained at the end of Section II-C; this is our estimate of $\mathbf{x}$. For given kernel η, we minimize $\tilde{R}(\cdot)$, with $\tilde{\mathbf{x}}$ substituted in place of $\hat{\mathbf{x}}$, as explained in Section II-D1, to choose the kernel width w and thus obtain $\hat{\lambda}$.

III. Conclusion

We have addressed the problem of jointly estimating the intensity $\lambda(\cdot)$ of an inhomogeneous Poisson process and the system's impulse response h when the data are samples from the shot noise (1.1).

Making only some general assumptions on h, we exploited the structure of the data z to obtain the estimate $\hat{\mathbf{x}}$ of the histogram $\mathbf{x}$ of the underlying point process. The estimate $\hat{h}$ was then obtained so that $\mathbf{z} \approx \hat{\mathbf{x}} * \hat{\mathbf{h}}$. We obtained very good estimates of h when $h(0) > 0$, in spite of h having a lot of oscillation. Examples 1 and 4 are two representative examples. When $h(0) = 0$, the problem of estimating h becomes harder. In Examples 2 and 3, good estimates $\hat{h}$ were obtained; in Examples 5 and 6, the estimates $\hat{h}$ are fairly good, showing only some spurious oscillations near the tail.

In Examples 1–6, we considered the kernel-estimate approach for obtaining $\hat{\lambda}$. We obtained satisfactory estimates of λ for a reasonable amount of computation. In Examples 4–6,

we also considered a version of the regularization approach introduced in [24]; good estimates of λ were obtained in Examples 4 and 5, but the estimate obtained in Example 6 is oversmooth when compared with the other two. These estimates behaved better near the end points than the corresponding kernel estimates.

The kernel-estimate approach of Section II-D1 suggested a similar approach to the problem of estimating λ when h is known. Even though this approach is conceptually simpler and requires fewer computations than the regularization approach presented in [24], its performance was poor when the number of random points T_ν was small (cf. Example 7). Furthermore, one does not expect the kernel-estimate method to be effective when the number of available data samples is small when compared with the detail in λ.

A. Measurement Noise

In a more general scenario, measurement noise would be present in the system; hence, the available data samples would be of the form $z(kt_0)+\varepsilon_k$, where the ε_k are i.i.d. Gaussian random variables with mean 0 and variance σ^2. Because of the dependence of $\hat{\mathbf{x}}$ on $\mathbf{d}_2$ (cf. (2.4) and (2.8)), the proposed method for obtaining $\hat{\mathbf{x}}$ may not work if σ^2 is large. In fact, by taking successive differences, the contribution of the ε_k terms on the $\mathbf{d}_2(k)$ is equivalent to that of a random variable with variance $6\sigma^2$. Even if this is the case, we have obtained results comparable with those presented here for the case where h has a jump at the origin when $\sigma \le 0.06h(0)$; the quality of the estimates deteriorated when $\sigma \ge 0.08h(0)$. Similarly, we have obtained good results for the case when h is continuous at the origin when $\sigma \le 0.025t_0h_1(0)$; the quality of the estimates deteriorated when $\sigma \ge 0.06t_0h_1(0)$.

We are now investigating the following approach to overcome this limitation: Since the ε_k are uncorrelated and are statistically independent of the random points T_ν, one can preprocess the measured data in an attempt to recover the shot noise data and then follow the procedure in this paper for estimating h and λ. If σ is known, then the *unbiased risk method* can be used for this purpose [19]; if σ is not known, then the *generalized cross-validation method* is appropriate [7].

Appendix A
The Estimate of $\mathbf{x}$ When h Is of Class One

Recall from (2.1) that $z(t)=h(0)N_t+y_1(t)$; then

$$\mathbf{d}_1(1) := \mathbf{z}(1) = h(0)\mathbf{x}(1)+y_1(t_0)$$

and for $k \ge 2$

$$\begin{aligned}\mathbf{d}_1(k) &:= \mathbf{z}(k)-\mathbf{z}(k-1)\\ &= h(0)\mathbf{x}(k)+y_1(kt_0)-y_1((k-1)t_0).\end{aligned}$$

The condition $h(0) \gg \omega(h;t_0)$ implies that $y_1(t)$ is a slowly varying process when compared with $h(0)N_t$, and we expect to be able to recover $\mathbf{x}$ under rather general conditions on t_0, h, and λ. To continue the analysis, write

$$\mathbf{d}_2(1) := \mathbf{d}_1(1) = h(0)\mathbf{x}(1)+y_1(t_0)$$

and for $k \ge 2$

$$\begin{aligned}\mathbf{d}_2(k) &:= \mathbf{d}_1(k)-\mathbf{d}_1(k-1)\\ &= h(0)(\mathbf{x}(k)-\mathbf{x}(k-1))\\ &\quad + y_1(kt_0)-2y_1((k-1)t_0)+y_1((k-2)t_0).\end{aligned}$$

We assume that

$$|y_1(kt_0)-2y_1((k-1)t_0)+y_1((k-2)t_0)| \ll h(0) \quad \text{(A.1)}$$

for $k \ge 1$; in this case, (2.3) holds. A careful analysis of (A.1) is carried out below.

For $\mathbf{p} \in \mathbb{R}^K$, let

$$S_q(\mathbf{p}) := \sum_{k=1}^{K} qQ_q(\mathbf{p}(k)).$$

The following result is the key to our method for choosing q:

Proposition A.1: For $r \in \mathbb{R}$ and $q>0$, $\lim_{q\to 0} S_q(r\mathbf{x}) = r\|\mathbf{x}\|_1$. Furthermore, $S_q(r\mathbf{x}) = r\|\mathbf{x}\|_1$ if $q = |r|n_0/n$, where n_0 is the greatest common divisor of the $\mathbf{x}(k)$, and $n \in \mathbb{N}$. In the vicinity of $|r|n_0/n$, $S_q(r\mathbf{x})$ is a linear function of q.

Proof: Observe that $Q_1(r)$ rounds r to the nearest integer; hence, $|Q_1(r)-r| \le 1/2$. Since $Q_q(r)=Q_1(r/q)$,

$$|qQ_q(r)-r| = q|Q_1(r/q)-r/q| \le q/2.$$

Therefore, $\lim_{q\to 0} qQ_q(r) = r$, and the first assertion of the proposition follows from the linearity of $\lim$ and the fact that $\mathbf{x}(k) \ge 0$ for each k. To prove the second one, write $\mathbf{x}(k)=n_kn_0$, where $n_k := \mathbf{x}(k)/n_0 \in \mathbb{Z}_+$, and take $q=|r|n_0/n$ for any $n \in \mathbb{N}$. If $n_k=0$, $qQ_q(r\mathbf{x}(k))=0$; otherwise

$$\begin{aligned}qQ_q(r\mathbf{x}(k)) &= |r|n_0/n\lfloor rn_kn_0(n/|r|n_0)+0.5\rfloor\\ &= rn_kn_0 = r\mathbf{x}(k)\end{aligned}$$

and the second assertion follows by adding the K terms. To show the third one, it suffices to show that as a function of q, $qQ_q(r)$ is linear in the vicinity of q-points satisfying $qQ_q(r)=r$ since those of the form $q=|r|n_0/n$ satisfy $qQ_q(r\mathbf{x}(k)) = r\mathbf{x}(k)$ simultaneously for each k, and the superposition of K linear terms is also linear. To this end, observe that $qQ_q(r)=r \Leftrightarrow Q_q(r)=r/q \Leftrightarrow r/q \in \mathbb{Z} \Leftrightarrow q=|r|/m$ for $m \in \mathbb{N}$; moreover, $qQ_q(r)$ is piecewise linear (as a function of $q>0$) with discontinuities at and only at $q=2|r|/(2m-1)$ for $m \in \mathbb{N}$, and any q-point of the form $q=|r|/m$ lies *strictly* between adjacent discontinuities. □

Recall (2.3); if we neglect the y_1 terms in $\mathbf{d}_2(k)$, then $|\mathbf{d}_2(k)| = h(0)n_k$ for some $n_k \in \mathbb{Z}_+$. Let n_0 denote the greatest common divisor of the n_k, and let $|\mathbf{d}_2|(k) := |\mathbf{d}_2(k)|$. From Proposition A.1, it follows that $S_q(|\mathbf{d}_2|) \approx \|\mathbf{d}_2\|_1$ for small q, that $S_q(|\mathbf{d}_2|) = \|\mathbf{d}_2\|_1$ if $q=h(0)n_0/n$ for $n \in \mathbb{N}$, and that $S_q(|\mathbf{d}_2|)$ is linear in the vicinity of $q=h(0)n_0/n$. See, e.g., Fig. 26. To choose a value for q, we plot $S_q(|\mathbf{d}_2|)$ to identify graphically the intervals over which $S_q(|\mathbf{d}_2|)$ crosses the constant $\|\mathbf{d}_2\|_1$ and is a linear function of q; we then choose q as the abscissa over the right-most such interval for which $S_q(|\mathbf{d}_2|) = \|\mathbf{d}_2\|_1$. We expect this approach to give $q \approx h(0)n_0$, and thus, $\hat{\mathbf{x}} = \mathbf{x}/n_0$, but the chances of $n_0 \ne 1$ are slim. Fig. 26 shows $S_q(|\mathbf{d}_2|)$ from Example 1 in Section

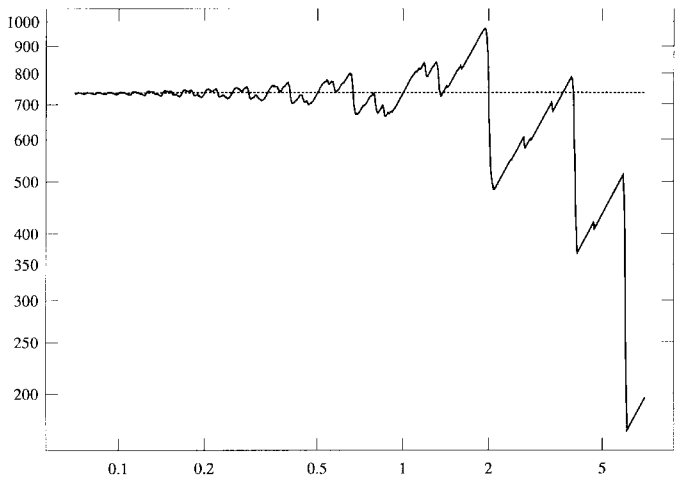


Fig. 26. Example 1: $S_q(|\mathbf{d}_2|)$ (solid line) andits limiting value $\|\mathbf{d}_2\|_1$ (dotted line).

II-D3 along with its limiting value $\|\mathbf{d}_2\|_1$. It is clear from the plot that $S_q(|\mathbf{d}_2|) \approx \|\mathbf{d}_2\|_1$ for $q \approx 1/n$, $n \in \mathbb{N}$ and that $S_q(|\mathbf{d}_2|)$ is linear in the vicinity of those q-points. These observations led us to choose $q = 1.002$.

Comments on Assumption A.1: Previously, we imposed the deterministic condition $h(0) \gg \omega(h; t_0)$. We also assumed that (A.1) holds for $k \geq 1$; this condition, however, involves random variables and requires some justification. To this end, let U_k denote the event (A.1). What we want is that (2.3) hold with probability close to one, for then $\mathsf{P}(\hat{\mathbf{x}} = \mathbf{x}) \approx 1$. This will be indeed the case if $\mathsf{P}\left(\bigcap_{k=2}^{K} U_k\right) \approx 1$.

Let

$$s_1(t) := y_1(t) - 2y_1(t - t_0) + y_1(t - 2t_0)$$

where y_1 is defined in (2.2); this allows the rewriting of U_k more compactly as $U_k = \{|s_1(kt_0)| \ll h(0)\}$. As we pointed out in Section II-A, $y_1(t)$ can be viewed as shot noise with impulse response $h_1(t)$; similarly, $s_1(t)$ can be viewed as shot noise with impulse response $r(t) := h_1(t) - 2h_1(t - t_0) + h_1(t - 2t_0)$. Let $\mu_1(t) := \mathsf{E}[s_1(t)]$ and $\sigma_1^2(t) := \mathsf{E}[|s_1(t) - \mu_1(t)|^2]$; since [20, eq. (16-84)]

$$\mu_1(t) = \int_0^t \lambda(\tau) r(t - \tau) d\tau,$$

and

$$\sigma_1^2(t) = \int_0^t \lambda(\tau) |r(t - \tau)|^2 d\tau$$

we can expect $\mathsf{P}(U_k)$ to be large if both $\mu_1(t)$ and $\sigma_1^2(t)$ are small when $t = kt_0$. Heuristically, r is t_0^2 times the second derivative of h; therefore, if that derivative is "well behaved," we can expect $s_1(t)$ to be small if t_0 is small and, consequently, $\mathsf{P}(U_k)$ to be large.

Appendix B
The Estimate of $\mathbf{x}$ When h Is of Class Two

Recall (2.5) and (2.6); observe that the sum on the right-hand side of (2.5) is a superposition of ramps with slope $h_1(0)$ originating at the T_ν. Recall that $h_1(0) > 0$ (the case $h_1(0) < 0$ is similar), and recall that $I_k := ((k-1)t_0, kt_0]$; then

$$\mathbf{d}_1(k) = t_0 h_1(0) N_{(k-1)t_0} + \nu_k + y_2(kt_0) - y_2((k-1)t_0)$$

where

$$\nu_k := h_1(0) \sum_{\nu: T_\nu \in I_k} (kt_0 - T_\nu) \in (0, t_0 h_1(0) \mathbf{x}(k)]. \quad \text{(B.1)}$$

Recall (2.7); in this case, $\mathbf{d}_2(1) = \nu_1 + \mathbf{s}_2(1)$, and for $k \geq 2$,

$$\mathbf{d}_2(k) = t_0 h_1(0) \mathbf{x}(k-1) - \nu_{k-1} + \nu_k + \mathbf{s}_2(k).$$

Even though $\mathbf{x}$ is integer valued and y_2 is continuous, the presence of the ν_k-terms makes it difficult to recover $\mathbf{x}$ from $\mathbf{d}_2$. Were it not for these terms, we could introduce the obvious vector of third differences and mimic the procedure in Section II-A to recover $\mathbf{x}$: If the ν_k were equal to 0, a plot of $S_q(|\mathbf{d}_3|)$ would look like a plot of $S_q(|\mathbf{d}_2|)$ in Section II-A. As explained in Section II-B, our approach, instead, is to obtain an estimate $\hat{\mathbf{s}}_2$ of $\mathbf{s}_2$ and then use (2.8) to get $\hat{\mathbf{x}}$. This approach can lead to $\hat{\mathbf{x}} \neq \mathbf{x}$ even if $\hat{\mathbf{s}}_2 = \mathbf{s}_2$ and $q = t_0 h_1(0)$ because

$$\begin{aligned} \mathbf{d}_2(k) - \mathbf{s}_2(k) &= t_0 h_1(0) \mathbf{x}(k-1) - \nu_{k-1} + \nu_k \\ &\in [0, t_0 h_1(0)(\mathbf{x}(k-1) + \mathbf{x}(k))]. \end{aligned} \quad \text{(B.2)}$$

For $k < K$, though, the term ν_k affects $\mathbf{d}_2(k)$ and $\mathbf{d}_2(k+1)$ in a complementary way so that on average, its effect gets canceled. To be more explicit, observe that since $\mathbf{d}_2(k) - \mathbf{s}_2(k) \geq 0$

$$\begin{aligned} \|\mathbf{d}_2 - \mathbf{s}_2\|_1 &= \sum_{k=1}^{K} (\mathbf{d}_2(k) - \mathbf{s}_2(k)) \\ &= \nu_1 + \sum_{k=2}^{K} \{t_0 h_1(0) \mathbf{x}(k-1) - \nu_{k-1} + \nu_k\} \\ &= t_0 h_1(0) \|\mathbf{x}\|_1 - t_0 h_1(0) \mathbf{x}(K) + \nu_K. \end{aligned}$$

In general, we expect $\mathbf{x}(K) \ll \|\mathbf{x}\|_1$, and hence, $\|\mathbf{d}_2 - \mathbf{s}_2\|_1 \approx t_0 h_1(0) \|\mathbf{x}\|_1$; in particular, recalling (B.1), $\|\mathbf{d}_2 - \mathbf{s}_2\|_1 = t_0 h_1(0) \|\mathbf{x}\|_1$ whenever $\mathbf{x}(K) = 0$.

Let n_0 denote again the greatest common divisor of the $\mathbf{x}(k)$. Recalling Proposition A.1, we have $S_q(\mathbf{d}_2 - \mathbf{s}_2) \approx \|\mathbf{d}_2 - \mathbf{s}_2\|_1$ for small q. As a function of q, $S_q(\mathbf{d}_2 - \mathbf{s}_2)$ is still piecewise linear, but because of the ν_k-terms, neither do the discontinuities occur necessarily at $t_0 h_1(0) n_0/(2n-1)$, $n \in \mathbb{N}$, nor does the function take on the limiting value $\|\mathbf{d}_2 - \mathbf{s}_2\|_1$ at those points. This situation renders the estimation of $t_0 h_1(0)$ harder than it was to estimate $h(0)$ in Section II-A. Furthermore, only the estimate $\hat{\mathbf{s}}_2$ is available and not $\mathbf{s}_2$.

From the discussion in Appendix A, it follows that a plot of $S_q(\mathbf{d}_2 - \mathbf{s}_2)$ should look like a "noisy" version of the curve in Fig. 26. Given $\hat{\mathbf{s}}_2$, $\hat{\mathbf{x}}$ can be readily obtained from (2.8) for each q. Suppose that $\hat{\mathbf{s}}_2$ is at hand; to choose a value for q, we first plot $S_q(\mathbf{d}_2 - \hat{\mathbf{s}}_2)$ to identify graphically a plausible interval for q: We look for an interval whose lower and higher endpoints correspond, respectively, to approximate global minimum and global maximum around the limiting

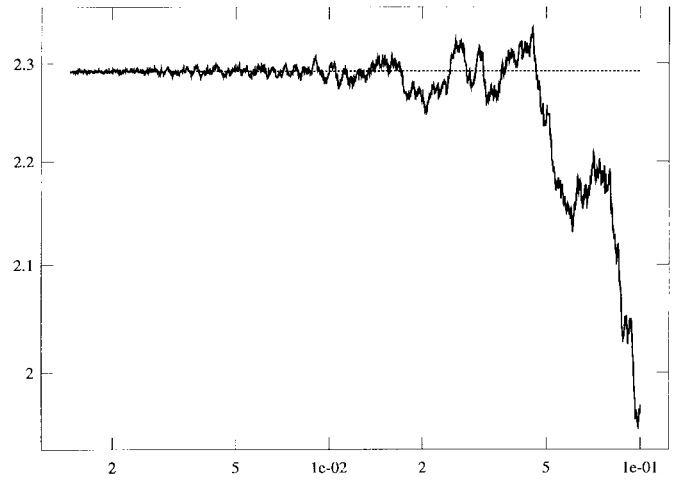


Fig. 27. Example 2: $S_q(\mathbf{d}_2 - \hat{\mathbf{s}}_2)$ (solid line) and its limiting value $\|\mathbf{d}_2 - \hat{\mathbf{s}}_2\|_1$ (dotted line).

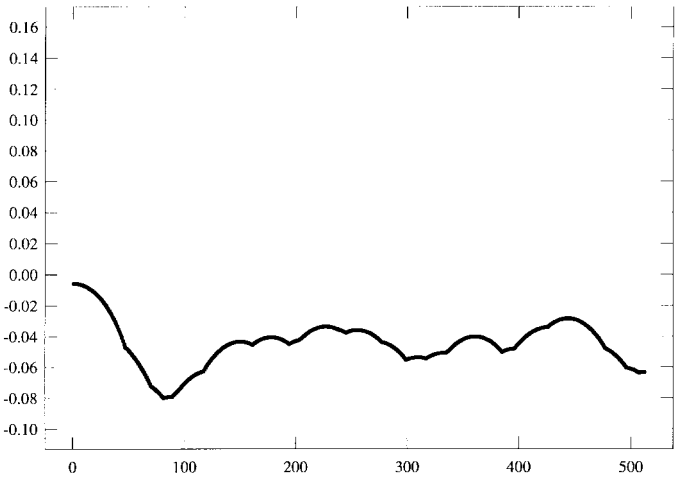


Fig. 28. Example 2: $\hat{\mathbf{s}}_2(k)$.

value $||\mathbf{d}_2 - \hat{\mathbf{s}}_2||_1$ (cf. Fig. 26). Inside this interval, we choose q to minimize the *ad hoc* "consistency cost"

$$C(q) := \sum_{n=1}^{n_{\max}} | \, ||\mathbf{d}_2 - \hat{\mathbf{s}}_2||_1 - q/n||\hat{\mathbf{x}}|_{q/n}||_1|.$$

The definition of $C(q)$ is motivated by $S_q(\mathbf{d}_2 - \mathbf{s}_2) \approx t_0 h_1(0)||\mathbf{x}||_1$ for small q and by (2.8): We expect $\hat{\mathbf{x}}|_{q/n} \approx n\hat{\mathbf{x}}|_q$, $n \in \mathbb{N}$.

Fig. 27, which corresponds to Example 2 in Section II-C1, shows the plot of $S_q(\mathbf{d}_2 - \hat{\mathbf{s}}_2)$ along with its limiting value $||\mathbf{d}_2 - \hat{\mathbf{s}}_2||_1$. This plot shows the difficulty of choosing q now that $h(0) = 0$. The curve $S_q(\mathbf{d}_2 - \hat{\mathbf{s}}_2)$ is not as "clean" as is $S_q(|\mathbf{d}_2|)$ in Fig. 26, and there does not appear to be a pattern of points $q, q/2, q/3, \ldots$ at which the curve hits the limiting value $||\mathbf{d}_2 - \hat{\mathbf{s}}_2||_1$. The curve in Fig. 27 suggests looking for q in the plausible interval $[0.02, 0.045]$; the minimizer of $C(\cdot)$ over this interval is $q = 0.024\,56$.

We now estimate $\mathbf{s}_2$. From (B.2), it follows that $\mathbf{d}_2(k) \geq \mathbf{s}_2(k)$, with equality if $\mathbf{x}(k) = \mathbf{x}(k-1) = 0$; the condition $h_1(0) \gg \omega(h_1; t_0)$ implies $\max_k |\mathbf{d}_2(k) - \mathbf{d}_2(k-1)| \gg \max_k |\mathbf{s}_2(k) - \mathbf{s}_2(k-1)|$. The smooth trend shown in Fig. 6 corresponds to $\mathbf{s}_2(k)$ from Example 2. It turns out that a slight modification of the discrete *morphological opening* of $\mathbf{d}_2$ by a "smooth" *structuring vector* $\mathbf{f}$ is a convenient way for obtaining $\hat{\mathbf{s}}_2$ (see [11, ch. 4]—Figs. 4.15 and 4.16 in particular). The morphological opening consists of performing a *morphological erosion*, or Minkowski difference, followed by a *morphological dilation*, or Minkowski sum (see also [11, Figs. 4.12 and 4.13]). To this end, we consider $\mathbf{f}(k) := r_1(k^2 - k_0^2)^{1/2}$, $|k| \leq k_0$; the width $k_0 \in \mathbb{N}$ and the scale $r_1 > 0$ are chosen below. We first perform

$$\mathbf{p}(k) = \min_{i \leq |k_0|, 1 \leq k+i \leq K} \{\mathbf{d}_2(k+i) - \mathbf{f}(i)\}, \quad 1 \leq k \leq K$$

(were it not for the constraint $1 \leq k + i \leq K$, this would be a morphological erosion), and then, take

$$\hat{\mathbf{s}}_2(k) := \max_{i \leq |k_0|, \; 1 \leq k-i \leq K} \{\mathbf{p}(k-i) + \mathbf{f}(i)\}, \quad 1 \leq k \leq K$$

(were it not for the constraint $1 \leq k - i \leq K$, this would be a morphological dilation). The modification in the morphological opening was done to improve the result near the end points. The width k_0 and the scale r_1 are chosen to achieve a good tradeoff between a small $|\hat{\mathbf{s}}_2(1)|$ and a small $\max_k |\hat{\mathbf{s}}_2(k) - \hat{\mathbf{s}}_2(k-1)|$. For the numerical examples in Section II-D3, we set $k_0 = 60$ and chose r_1 to minimize the max-jump in $\hat{\mathbf{s}}_2$. There is one exception (Example 2), where a local minimum in the max-jump, with a much smaller value of $|\hat{\mathbf{s}}_2(1)|$, was chosen. Fig. 28 shows $\hat{\mathbf{s}}_2(k)$ from Example 2 (cf. Fig. 6).

Appendix C
Choosing β

To compute $\hat{\mathbf{h}}$, we first need to obtain $\hat{\mathbf{x}}$; for each β, $\hat{\mathbf{h}}$ is then obtained as an approximate solution to $\mathbf{z} = \hat{\mathbf{x}} * \mathbf{r}$ viewed as a function of $\mathbf{r}$. One can then ask whether $\hat{\mathbf{x}}$ would result from "solving" $\mathbf{z} = \hat{\mathbf{h}} * \mathbf{r}$ over $\mathbb{N}^K$; this is the key to our approach for choosing β. Recall that $\hat{\mathbf{h}} * \mathbf{r} = L(\hat{\mathbf{h}})\mathbf{r}$ and that $L(\hat{\mathbf{h}})$ is a lower-triangular matrix. Suppose that $\hat{\mathbf{x}}$ is at hand, and let $\hat{\mathbf{h}}$ vary as β varies; for each β, let $\ell_0 = \min\{k : \hat{\mathbf{h}}(k) \neq 0\}$, and let $q = |\hat{\mathbf{h}}(\ell_0)|$. We let $\tilde{\mathbf{x}} \in \mathbb{N}^K$ denote the approximate solution to $\mathbf{z} = L(\hat{\mathbf{h}})\mathbf{r}$ obtained by "back substitution," $\tilde{\mathbf{x}}(\ell_0) = Q_q(\mathbf{z}(\ell_0))$ and for $\ell_0 < k \leq K$,

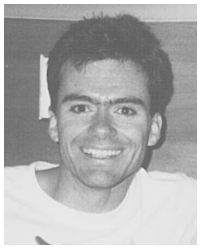
$$\tilde{\mathbf{x}}(k) = \max\left\{Q_q\left(\mathbf{z}(k) - \sum_{\ell=\ell_0+1}^{k} \hat{\mathbf{h}}(\ell)\tilde{\mathbf{x}}(k-\ell+1)\right), 0\right\}.$$

If $\ell_0 \geq 2$, we let $\tilde{\mathbf{x}}(k) = 0$ for $k < \ell_0$. Of course, we would like a β for which $\tilde{\mathbf{x}} = \hat{\mathbf{x}}$ is true; we choose β as the smallest minimizer of $| \, ||\hat{\mathbf{x}}||_1 - ||\tilde{\mathbf{x}}||_1|$.

References

[1] N. Antoniadis and A. O. Hero, "Time-delay estimation for filtered Poisson processes using an EM-type algorithm," *IEEE Trans. Signal Processing*, vol. 42, pp. 2112–2123, 1994.
[2] G. R. Ayers and J. C. Dainty, "Iterative blind deconvolution method and its applications," *Opt. Lett.*, vol. 13, pp. 547–549, 1988.
[3] H. H. Barret and W. Swindell, *Radiological Imaging: The Theory of Image Formation, Detection, and Processing*. New York: Academic, 1981, vols. 1 and 2.

[4] R. H. T. Bates, B. Q. Quek, and C. R. Parker, "Some implications of zero sheets for blind deconvolution and phase retrieval," *J. Opt. Soc. Amer. A*, vol. 7, pp. 468–479, 1990.
[5] C. de Boor, *A Practical Guide to Splines*. New York: Springer-Verlag, 1978.
[6] M. Cannon, "Blind deconvolution of spatially invariant image blurs with phase," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 58–63, 1976.
[7] P. Craven and B. Wahba, "Smoothing noisy data with spline functions," *Numer. Math.*, vol. 31, pp. 377–403, 1979.
[8] B. L. K. Davey, R. G. Lane, and R. H. T. Bates, "Blind deconvolution of noisy complex-valued images," *Opt. Commun.*, vol. 69, pp. 353–356, 1989.
[9] P. Diggle and J. S. Marron, "Equivalence of smoothing parameter selectors in density and intensity estimation," *J. Amer. Statist. Assoc.*, vol. 83, pp. 793–800, 1988.
[10] R. M. Gagliardi and S. Karp, *Optical Communications*. New York: Wiley, 1976.
[11] H. J. A. M. Heijmans, *Morphological Image Operators*. San Diego, CA: Academic, 1994.
[12] T. J. Holmes, "Blind deconvolution of quantum-limited incoherent imagery: Maximum-likelihood approach," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1052–1061, 1992.
[13] R. G. Lane and R. H. T. Bates, "Automatic multidimensional deconvolution," *J. Opt. Soc. Amer. A*, vol. 4, pp. 180–188, 1987.
[14] R. G. Lane, W. R. Fright, and R. H. T. Bates, "Direct phase retrieval," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 520–526, 1987.
[15] B. C. McCallum, "Blind deconvolution by simulated annealing," *Opt. Commun.*, vol. 75, pp. 101–105, 1990.
[16] N. Miura and N. Baba, "Extended-object reconstruction with sequential use of iterative blind deconvolution method," *Opt. Commun.*, vol. 89, pp. 375–379, 1992.
[17] G. M. Morris, "Pattern recognition using photon-limited images," in *Optical Processing and Computing*, H. H. Arsenault, T. Szoplik, and B. Macukow, Eds. Boston, MA: Academic, 1989, pp. 343–390.
[18] *The NAG Fortran Library Manual, Mark 15*, The Numerical Algorithms Group Limited, vol. 3, 1991.
[19] F. O'Sullivan, "A statistical perspective on ill-posed inverse problems," *Statist. Sci.*, vol. 1, pp. 502–527, 1986.
[20] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1991.
[21] T. J. Schulz, "Multi-frame blind deconvolution of astronomical images," *J. Opt. Soc. Amer. A*, vol. 10, pp. 1064–1073, 1993.
[22] J. R. Segal, B. Ceccarelli, R. Fesce, and W. P. Hurlbut, "Miniature endplate potential frequency and amplitude determined by an extension of Campbell's theorem," *Biophys. J.*, vol. 47, pp. 183–202, 1985.
[23] J. H. Seldin and J. R. Fienup, "Iterative blind deconvolution algorithm applied to phase retrieval," *J. Opt. Soc. Amer. A*, vol. 7, pp. 428–433, 1990.
[24] R. E. Sequeira and J. A. Gubner, "Intensity estimation from shot-noise data," *IEEE Trans. Signal Processing*, vol. 43, pp. 1527–1531, 1995.
[25] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inform. Theory*, vol. 36, pp. 312–321, 1990.
[26] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall, 1986 (1990 reprint).
[27] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*. New York: Springer-Verlag, 1991.
[28] T. G. Stockham, T. M. Cannon, and R. B. Ingebretsen, "Blind deconvolution through digital signal processing," in *Proc. IEEE*, vol. 63, pp. 678–692, 1975.
[29] A. M. Tekalp, H. Kaufman, and J. W. Woods, "Identification of image and blur parameters for the restoration of noncausal blurs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 963–972, 1986.
[30] K. Yana, H. Marushima, H. Mino, and N. Takeuchi, "Bispectral analysis of filtered impulse processes with applications to the analysis of bioelectric phenomena," in *Workshop on Higher-Order Spectral Anal.*, Vail, CO, 1989, pp. 140–145.



Raúl E. Sequeira (M'95) was born in San José, Costa Rica. He received the B.S. degree in electrical engineering from Universidad de Costa Rica in 1988 and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin, Madison, in 1990 and 1995, respectively.

From 1989 to 1995, he was a Research Assistant at the Electrical and Computer Engineering (ECE) Department, University of Wisconsin, Madison. His work included the use of detection and estimation theory, statistical analysis, numerical methods, and point process theory, to solve problems arising in communication systems. From June 1995, to March 1996, he worked as a Post Doctoral Fellow at the Département Signal et Image of the Institut National des Télécommunications, Evry, France. His work involved the use of mathematical morphology for image processing applications. Since April 1996, he has been with the ECE Department, University of Wisconsin, Madison, working as a research associate on blind deconvolution problems and as an associate lecturer, teaching a course on random signal analysis.

John A. Gubner (M'88) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1983, 1985, and 1988, respectively.

In 1988, he joined the faculty of the University of Wisconsin, Madison, where he is now an Associate Professor in the Electrical and Computer Engineering Department. In 1985, he received an IEEE Frank A. Cowan Scholarship for graduate study in communications, and from 1986 to 1988, he was a Graduate Fellow with the Systems Research Center at the University of Maryland. His research interests include non-Poisson point processes, shot-noise random processes, Markov chain Monte Carlo, information theory, distributed estimation, and wavelets.