

Poisson intensity estimation for tomographic data using a wavelet shrinkage approach

Laurent Cavalier* and Ja-Yong Koo†

September 03, 2002

Abstract

We consider a two-dimensional problem of positron emission tomography where the random mechanism of the generation of the tomographic data is modeled by Poisson processes. The goal is to estimate the intensity function which corresponds to emission density. Using the wavelet-vaguelette-decomposition, we propose an estimator based on thresholding of empirical vaguelette coefficients which attains the minimax rates of convergence on Besov function classes. Furthermore we construct an adaptive estimator which attains the optimal rate of convergence up to a logarithmic term.

Keywords. Adaptation, Besov spaces, intensity estimation, linear estimators, minimax bound, nonlinear estimators, PET, thresholding, WVD.

1 Introduction

This paper deals with the estimation problem in positron emission tomography (PET), when the random mechanism of the generation of the PET data is modeled by Poisson processes. PET deals with the estimation of the amount and location of a radioactively labeled metabolite emitted inside a body using emission measured outside. In order to model the PET problem, we assume that emissions occur according to a spatial Poisson process F_f^n in a certain region \mathcal{Q} (a cross-section of the patient's head) of \mathbb{R}^2 . The process F_f^n has an unknown intensity function f^n , which is usually referred to as the emission density. However, we are not informed of the location of any such emission points, but only that an emission has occurred on a line containing the emission point; the line is randomly and uniformly oriented, independently of the position of the emission

*Centre de Mathématiques et Informatique, Université Aix-Marseille 1, 39 rue Joliot-Curie 13453, Marseille cedex 13, France (email: cavalier@cmi.univ-mrs.fr).

†Department of Statistics, Inha University, Incheon 402-751, Korea (email: jykoo@stat.inha.ac.kr). Koo's research was supported by a grant from the KOSEF (980701-0201-3).

point. In fact, we observe a Poisson process G_g^n with intensity function g^n , where g^n is obtained by Kf^n with K , the Radon operator. For studying asymptotic properties, we adopt the model f^n defined as

$$f^n(x, y) = nf(x, y),$$

where the function f is held fixed and the positive real n increases. The function f will be referred to as the scaled intensity function. The PET problem is to estimate the intensity function f based on the observation of G_g^n .

The appearance of wavelets has exciting implications for nonparametric function estimation [10, 11, 12]. Donoho [9] has shown that wavelet techniques can be a powerful tool for the study of inverse problems. The works in this direction are [1, 14, 24, 42], among others. In this paper, the wavelet-vaguelette decomposition (WVD) [9] is used for the estimation of intensity functions.

Several approaches to studying PET include, among others: Maximum-Likelihood [32, 33, 41]; EM algorithm [36, 37]; Least-Squares [16, 32]; Complexity Regularization [30]; to name just a few, however, as will be discussed below, it appears that our WVD approach is novel and different from the above. Other works on PET include [3, 4, 5, 9, 13, 16, 17, 20, 21, 22, 24, 26, 27, 28, 32, 33, 36, 37, 41]. Deans [8] and Natterer [31] provide a general exposition to computerized tomography.

The main concern of this paper is to show theoretical properties of wavelet methods for PET when the unknown intensity function belongs to a Besov class. We will do so by showing that nonlinear wavelet estimators using thresholding, attain the optimal rate of convergence even for non-smooth functions; while linear estimators, (such as for example filtered backprojection (FBP)) cannot be optimal. We will also obtain adaptability results. It is worthwhile pointing out that theoretical studies, such as that encompassed in this paper, look at what may be called the idealized PET model. In reality, problems such as time-of-flight and accidental coincidences [19, 33] occur and are usually not considered and will not be considered in this paper as well. It is felt however, that the methods of modeling PET in terms of Poisson processes, the approach taken in this paper, provide a solid foundation for analyzing real issues such as the above and is under future consideration. We also have under future consideration, the problem of imposing positivity of the estimator since the Poisson intensity functions are nonnegative by definition. The approaches of [26, 27, 38, 39] appear promising.

This paper is organized as follows. Section 2 describes the Poisson process model for the PET problem. In Section 3, several properties of WVD for the Radon operator are given. Sections 4 and 5 address the optimal L_2 -rate of convergence over Besov spaces B_{spq} . We prove lower bound properties for linear and nonlinear estimators and remark that when $1 \leq p < 2$, the linear estimator is not optimal. Numerical works using both simulated and real PET data have been carried out to show that wavelet methods outperform FBP [6]. In Section 5, we approximate the Poisson model by a Gaussian white-noise model and use results for threshold estimators in the white-noise model of [9], to construct a nonlinear estimator which attains the optimal rate of convergence

depending on s, p, q . In Section 6 we construct an estimator, using hard-thresholding, which does not depend on s, p, q , hence is adaptive, and which attains the optimal rate of convergence up to a logarithmic term. Thus up to a logarithmic term, we construct an estimator which adapts to unknown smoothness while depending on a regularization parameter. The choice of the regularization parameter needs to be carefully performed in order to obtain quality reconstruction. One way to make such a choice is through simulations, especially in tomographic problems. Another way is to construct adaptive estimators which adapt automatically to unknown smoothness [4, 5, 20], which can be done by WVD for PET. Adaptability is crucially important since an adaptive estimator will automatically make the correct choice and estimate the function f as if smoothness is known, when in fact it is unknown. In this sense, the thresholding wavelet estimator can estimate the function parameter in the “best” possible way, which is of course very important in real medical applications. In Section 7, we present basic properties of Poisson processes such as: moments bounds; large-deviation bound; information inequalities; and Kullback-Leibler divergence between Poisson processes.

2 Poisson processes for PET

Let (x, y) be a point of emission in the unit square $\mathcal{Q} = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq 1\}$ and denote by D_1, D_2 the points of detection on the circle $S_0 = \{u \in \mathbb{R}^2 : u_1^2 + u_2^2 = \sqrt{2}\}$. The unordered pairs $\{D_1, D_2\}$ form the detection space. Assume that the emission points are governed by a Poisson process F_f^n on \mathcal{Q} , where the intensity function f^n of F_f^n reflects the intensity of radioactivity in \mathcal{Q} . The intensity function f^n is the Lebesgue density of the intensity measure λ_f^n of F_f^n which is defined by the expectations $\lambda_f^n(B) = \mathbb{E}F_f^n(B) = \int_B f^n(x, y)dx dy$ for any Borel subset B of \mathcal{Q} . The scaled intensity measure λ_f is defined by $\lambda_f(B) = \int_B f(x, y)dx dy$.

Let ε_x denotes the Dirac measure with mass 1 at x . An explicit representation for the Poisson process F_f^n of emission points is given by

$$F_f^n = \sum_{i=1}^{\tau} \varepsilon_{(X_i, Y_i)},$$

where $\tau, (X_1, Y_1), (X_2, Y_2), \dots$ are independent, conditional on a given τ where the latter is a Poisson random variable with parameter $\lambda_f(\mathcal{Q})$ and $(X_i, Y_i), i \in \mathbb{N}$, are distributed as (X, Y) whose density function is given by $f^n/\lambda_f^n(\mathcal{Q})$. For notational convenience, we assume that $\lambda_f(\mathcal{Q}) = 1$ in which case τ is a Poisson random variable with mean n and f is the density for (X_i, Y_i) .

Given an emission point (X, Y) , the photons fly off on a line that is determined by (X, Y) and the emission angle $\Omega \in [0, \pi)$ which is generated independently of (X, Y) according to the uniform distribution on $[0, \pi)$. The line through D_1 and D_2 is determined by Ω and

$$R = X \cos \Omega + Y \sin \Omega \in [-\sqrt{2}, \sqrt{2}].$$

We are going to describe the Poisson process G_g^n that governs (R, Ω) . We want to find its intensity function g^n . It is well-known [31], that the density g with respect to $\pi^{-1}drd\omega$ in detector space is given by Kf with K the Radon operator

$$\begin{aligned}(Kf)(r, \omega) &= \int_{-\infty}^{\infty} f(r \cos \omega - t \sin \omega, r \sin \omega + t \cos \omega) dt \\ &= \int_{\mathbb{R}^2} f(x, y) \delta^D(r - x \cos \omega - y \sin \omega) dx dy,\end{aligned}$$

where δ^D is the Dirac delta function. Thus the value of the density $g(r, \omega)$ in detector space is the integral of the density f , in the emission space, along the line characterized by (r, ω) . It can be seen that $\delta^D(r - x \cos \omega - y \sin \omega)$ acts as a Markov kernel. We have the relation $\lambda_f^n(\mathcal{Q}) = \nu_g^n(\mathcal{T})$ since we suppose that every photon is detected. Thus, the intensity functions are related by

$$g^n(r, \omega) = (Kf^n)(r, \omega).$$

The statistical problem is to estimate the scaled intensity function f based on the observation

$$G_g^n = \sum_{i=1}^{\tau} \varepsilon_{(R_i, \Omega_i)},$$

which is the desired Poisson process of detection pairs.

3 Wavelet-Vaguelette Decomposition

We will assume a certain level of working knowledge concerning wavelets for which there are now excellent books [7, 18, 29]. We will thus proceed accordingly.

3.1 Wavelets and Besov spaces

Here we present the tensor-product wavelet basis of $L_2(\mathbb{R}^2)$. We suppose that we have mother and father wavelets ψ and ϕ which have at least three continuous derivatives and one vanishing moment. Now the index set is $I = (j, k, \epsilon)$, where j is an integer, $k = (k_x, k_y)$ is a member of the integer lattice \mathbb{Z}^2 , and $\epsilon \in \{0, 1, 2, 3\}$. Let

$$\Psi_I(x, y) = \begin{cases} \phi_{j, k_x}(x) \phi_{j, k_y}(y) & \text{if } \epsilon = 0 \\ \phi_{j, k_x}(x) \psi_{j, k_y}(y) & \text{if } \epsilon = 1 \\ \psi_{j, k_x}(x) \phi_{j, k_y}(y) & \text{if } \epsilon = 2 \\ \psi_{j, k_x}(x) \psi_{j, k_y}(y) & \text{if } \epsilon = 3. \end{cases}$$

Define spaces V_j , $j \in \mathbb{Z}$, by $V_0 = \overline{\mathcal{V}_0 \otimes \mathcal{V}_0}$. As in the one-dimensional case, we define, for each $j \in \mathbb{Z}$, the complement space W_j to be the orthogonal complement in V_{j+1} of V_j . We have

$V_{j+1} = V_j \oplus W_j$ so that $W_j = \left(\overline{\mathcal{V}_j \otimes \mathcal{W}_j}\right) \oplus \left(\overline{\mathcal{W}_j \otimes \mathcal{V}_j}\right) \oplus \left(\overline{\mathcal{W}_j \otimes \mathcal{W}_j}\right)$. Let E_j be the associated projection operator onto V_j and $D_j = E_{j+1} - E_j$.

Besov spaces depend on three parameters $s > 0$, $1 \leq p \leq \infty$ and $1 \leq q \leq \infty$ and are denoted B_{spq} . Let $\|\cdot\|_{L_p(B)}$ denote the L_p norm on a set B . We will say that $f \in B_{spq}$ if and only if the norm

$$J_{spq}(f) = \left\| E_{j_0} f \right\|_{L_p(\mathbb{R}^2)} + \left\{ \sum_{j \geq j_0} (2^{js} \|D_j f\|_{L_p(\mathbb{R}^2)})^q \right\}^{1/q} < \infty$$

(with usual modification for $q = \infty$). For $j \in \mathbb{Z}$, let $\mathcal{I}_0^{(j)} = \{I : \text{supp } \Psi_{(j,k,0)} \cap \mathcal{Q} \neq \emptyset\}$ and $\mathcal{I}^{(j)} = \{I : \text{supp } \Psi_{(j,k,\epsilon)} \cap \mathcal{Q} \neq \emptyset, \epsilon \in \{1, 2, 3\}\}$. It can be seen that $|\mathcal{I}_0^{(j)}| \asymp 2^{2j}$ and $|\mathcal{I}^{(j)}| \asymp 2^{2j}$, where $a_n \asymp b_n$ means that a_n/b_n is bounded away from zero and infinity. We have the decomposition of f :

$$E_j f = \sum_{I \in \mathcal{I}_0^{(j)}} \alpha_I \Psi_I, \quad D_j f = \sum_{I \in \mathcal{I}^{(j)}} \beta_I \Psi_I,$$

where $\alpha_I = \langle f, \Psi_I \rangle$, $I \in \mathcal{I}_0^{(j)}$ and $\beta_I = \langle f, \Psi_I \rangle$, $I \in \mathcal{I}^{(j)}$, the Besov space B_{spq} can be defined via the equivalent norm

$$\|f\|_{spq} = \|\alpha_{j_0 \cdot}\|_p + \left\{ \sum_{j \geq j_0} \left(2^{j(s+2(1/2-1/p))} \|\beta_{j \cdot}\|_p \right)^q \right\}^{1/q} < \infty$$

where we have set $\|\alpha_{j \cdot}\|_p = (\sum_{I \in \mathcal{I}_0^{(j)}} |\alpha_I|^p)^{1/p}$ and $\|\beta_{j \cdot}\|_p = (\sum_{I \in \mathcal{I}^{(j)}} |\beta_I|^p)^{1/p}$.

Forthwith, let M, C, C_1, C_2, \dots denote positive constants independent of n , where C denotes a generic constant, which may have different values from line to line.

Set $B_{spq}(M) = \{f : \|f\|_{spq} \leq M\}$. The space of scaled intensities is defined by

$$\mathcal{F}_{spq}(M) = \{f : f \geq 0, \|f\|_{spq} \leq M\}. \quad (1)$$

3.2 Vaguelettes

In this section we use the WVD described in [9] for the Radon operator. Define

$$\gamma_I(r, \omega) = \frac{1}{4\pi} \int_{-\infty}^{\infty} |u| \widehat{\Psi}_I(u \cos \omega, u \sin \omega) e^{iru} du.$$

One can see

$$\gamma_I(r, \omega) = 2^j \gamma_{(0,0,\epsilon)}(2^j r - \zeta_k(\omega), \omega),$$

where

$$\zeta_k(\omega) = k_x \cos \omega + k_y \sin \omega \quad \text{for } k = (k_x, k_y) \in \mathbb{Z}^2.$$

Let $\langle \cdot, \cdot \rangle$ and $[\cdot, \cdot]$ denote the inner products on \mathcal{Q} and \mathcal{T} with respect to the measures $dxdy$ and $\pi^{-1}drd\omega$, respectively. Since

$$[\gamma_I, Kf] = \langle \Psi_I, f \rangle, \quad (2)$$

f may be recovered from Kf by

$$f = \sum_{I \in \mathcal{I}} [Kf, \gamma_I] \Psi_I$$

when f can be written as a finite linear combination of wavelets

$$f = \sum_{I \in \mathcal{I}} \langle f, \Psi_I \rangle \Psi_I.$$

We list straightforward lemmas without proof.

Lemma 3.1 *For any ℓ_2 summable sequence $(a_k)_{k \in \mathbb{Z}^2}$,*

$$\left\| \sum_{k \in \mathbb{Z}^2} a_k K \Psi_{(j,k,3)} \right\| \leq C 2^{-j/2} \|(a_k)\|_2.$$

Lemma 3.2 *Suppose that the support of Ψ_I is contained in \mathcal{Q} . Then, for $\epsilon \neq 0$,*

$$\int_{\mathcal{T}} (K \Psi_I)(r, \omega) dr d\omega = 0.$$

Lemma 3.3 *For $\epsilon \in \{0, 1, 2, 3\}$,*

$$|\gamma_{(0,0,\epsilon)}| \leq C_1.$$

4 Linear estimation

In this section, we consider asymptotic behavior of linear estimators in L_2 -norm on the Besov ball $\mathcal{F}_{spq}(M)$. The class \mathcal{C}_L of linear estimators is defined by the representation

$$f_{L,n}^*(x, y) = \int F_n((x, y), (r, \omega)) dG_g^n(r, \omega)$$

where F_n is an arbitrary measurable function on $\mathcal{Q} \times \mathcal{T}$. The class of estimators are wide enough for most practical applications. For example, a direct modification of the orthogonal density estimator [22] admits such a form.

The following theorem states the best possible rate for the class of linear estimators.

Theorem 1 *Suppose that $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $s > 2/p$. Then as $n \rightarrow \infty$,*

$$\inf_{f_{L,n}^* \in \mathcal{C}_L} \sup_{f \in \mathcal{F}_{spq}(M)} \left(\mathbb{E}_f \|f_{L,n}^* - f\|_2^2 \right)^{1/2} \geq C n^{-s'/(2s'+3)},$$

where $s' = s + 2(1/2 - 1/p_-)$, with $p_- = \min(2, p)$, $\mathcal{F}_{spq}(M)$ is defined in (1) and \mathcal{C}_L denotes the class of linear estimators.

PROOF. Let f_0 be an infinitely differentiable intensity satisfying $f_0(x, y) \geq C$ for $(x, y) \in \mathcal{Q}$ and $\|f_0\|_{spq} \leq M/2$, and

$$\mathcal{I}_3^{(j)} = \{I : I \in \mathcal{I}^{(j)}, \epsilon = 3, \text{ supp } \Psi_I \subset \mathcal{Q}\}.$$

Consider

$$V_3^{(j)} = \left\{ f = f_0 + \sum_{I \in \mathcal{I}^{(j)}} \kappa_I \Psi_I : \kappa_I \leq \Gamma(j; s, p, M) \right\},$$

where

$$\Gamma(j; s, p, M) = \min \left(\frac{C}{2\|\Psi_{(0,0,3)}\|_{L_\infty(\mathcal{Q})}} 2^{-j}, \frac{M}{2} 2^{-2s'j} \right).$$

Choose η such that $f_I^+ = f_0 + \eta \Psi_I$ and $f_I^- = f_0 - \eta \Psi_I$ belong to $V_3^{(j)}$. When $0 \leq \eta \leq \Gamma(j; s, p, M)$, a pyramid

$$\mathcal{P}_j = \{f_0 \pm \eta \Psi_I : I \in \mathcal{I}_3^{(j)}\} \quad \text{for } j \geq 0$$

is included in $\mathcal{F}_{spq}(M)$.

Suppose that f_L^* is such that $\mathbb{E}_f f_L^*(x, y) < \infty$ for all $f \in V_3^{(j)}$ and $(x, y) \in \mathcal{Q}$. $\hat{\alpha}_I = \langle f_L^*, \Psi_I \rangle$. Let

$$\hat{\alpha}_I = \langle f_L^*, \Psi_I \rangle \text{ and } a_I = \int_{\mathcal{Q}} \frac{\partial}{\partial \kappa_I} [\mathbb{E}_f f_L^*(x, y)] \Psi_I(x, y) dx dy.$$

From Lemma 3.2, $\int (Kf)(r, \omega) dr d\omega = \int (Kf_0)(r, \omega) dr d\omega$ for $f \in V_3^{(j)}$. By applying the information inequality in Lemma 7.5 to the model in which G_g^n is the Poisson process with scaled intensity function Kf for $f \in V_3^{(j)}$ with an unknown parameter $\theta = \kappa_I$, we have

$$\text{var}_\theta(\hat{\alpha}_I) \geq \frac{\left(\frac{\partial}{\partial \theta} \mathbb{E}_\theta \hat{\alpha}_I\right)^2}{I(\theta)} = \frac{|a_I|^2}{I(\theta)}. \quad (3)$$

We have

$$I(\theta) = \int \left(\frac{K\Psi_I}{Kf_\theta} \right)^2 d\nu_g^n \leq Cn \int |K\Psi_I(r, \omega)|^2 dr d\omega \leq Cn/2^j. \quad (4)$$

since $0 \leq \kappa_I \leq \Gamma(j; s, p, M)$, then $\inf_{\mathcal{Q}} f \geq C$ for $f \in V_3^{(j)}$ so that $Kf_\theta \geq C$ on \mathcal{T} since the Radon operator is an averaging operator and using Lemma 3.1. Combining (3) and (4), we obtain

$$\mathbb{E}_\theta |\hat{\alpha}_I|^2 \geq \text{var}(\hat{\alpha}_I) \geq C2^j |a_I|^2 / n. \quad (5)$$

Now the arguments used to prove Theorem 1 of [11] can be used to show

$$\sup_{f \in \mathcal{F}_{spq}(M)} \mathbb{E}_f \|f_L^* - f\|_2^2 \geq \frac{1}{|\mathcal{P}_j|} \sum_{f \in \mathcal{P}_j} \mathbb{E}_f \|f_L^* - f\|_2^2 \geq C \frac{2^{3j}}{n}.$$

Choosing j so that $2^j \asymp n^{1/(2s'+3)}$, we have the desired result. \square

5 Nonlinear estimation

The aim of this section is to establish the optimal rate for any estimator of f based on the observation G_{Kf}^n . This is accomplished by first establishing a lower bound and then construct an estimator which attains the optimal rate of convergence. Therefore, we will remark that in some cases, when $1 \leq p < 2$, linear estimators do not achieve the optimal rate of convergence.

5.1 Lower bound

In this section we obtain a lower bound for all estimators. If one would use Sobolev space, then rate (5.3) in [40] could be obtained. Comparison between Theorem 1 and the following Theorem 2 shows that when $1 \leq p < 2$ linear estimators have suboptimal rate of convergence [9, 11, 18]. This means that in the case of functions which are not very smooth, we need some kind of nonlinearity in order to achieve the optimal rate of convergence up to a logarithmic term.

Theorem 2 *If $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and $s > 2/p$, then*

$$\inf_{f_n^* \in \mathcal{A}} \sup_{f \in \mathcal{F}_{spq}(M)} \mathbb{E}_f \|f_n^* - f\|_2 \geq Cn^{-\alpha},$$

where $\alpha = s/(2s+3)$, $\mathcal{F}_{spq}(M)$ is defined in (1) and \mathcal{A} denotes the class of all possible estimators.

PROOF. Let f_0 and $\mathcal{I}_3^{(j)}$ be the same as those defined in the proof of Theorem 1. Consider the set of vertices of a hypercube defined by

$$\mathcal{F}_j = \left\{ f = f_0 + f_\tau : f_\tau = \sum_{I \in \mathcal{I}_3^{(j)}} \kappa \tau_I \Psi_I, \tau_I = 0, 1 \right\},$$

where

$$0 \leq \kappa \leq \min \left(\frac{C}{2\|\Psi_{(0,0,3)}\|_{L_\infty(\mathcal{Q})}} 2^{-j}, \frac{M}{2} 2^{-(s+1)j} \right). \quad (6)$$

Let $|B|$ denote the number of elements in a set B . By Lemma 3.1 of [25] and the orthonormality of $\{\Psi_I\}$, there is a subset \mathcal{F}_j^* of \mathcal{F}_j such that

$$\|f_1 - f_2\|_2 \geq C2^{-sj} \text{ for } f_1 \neq f_2 \in \mathcal{F}_j^*, \text{ and } \log(|\mathcal{F}_j^*| - 1) > 0.272 \log(|\mathcal{F}_j|) \quad (7)$$

when $|\mathcal{F}_j| > 8$. By Lemma 3.1,

$$\|Kf_1 - Kf_2\|_2 \leq C2^{-(s+1/2)j} \quad \text{for all } f_1, f_2 \in \mathcal{F}. \quad (8)$$

Applying Lemma 3.2, we get that

$$\int (Kf)(r, \omega) dr d\omega = \int (Kf_0)(r, \omega) dr d\omega \quad \text{for all } f \in \mathcal{F}_j. \quad (9)$$

Since $C^{-1} \leq Kf \leq C$ for $f \in \mathcal{F}_j$, it follows from (8), Lemma 7.3 and Lemma 7.4 with (9) that

$$\Delta(G_{Ku}^n, G_{Kv}^n) \leq \int \frac{(Ku - Kv)^2}{Kv} d\nu_g^n \leq Cn2^{-(2s+1)j} \text{ for } u, v \in \mathcal{F}_j. \quad (10)$$

By Fano's lemma [2], if f_n^* is any estimator of f , then for $2^j \asymp n^{1/(2s+3)}$,

$$\sup_{f \in \mathcal{F}_{spq}(M)} 2^{j/2} \mathbb{E}_f \|f_n^* - f\|_2 \geq \sup_{f \in \mathcal{F}_j^*} 2^{j/2} \mathbb{E}_f \|f_n^* - f\|_2 \geq C^{-1} n^{-\alpha}. \quad \square$$

5.2 Upper bound

5.2.1 Gaussian approximation

The goal of this section is to show that, in some sense, the model with Poisson intensity is not 'too far' from the usual Gaussian white-noise model.

Consider the standard Gaussian white-noise model:

$$Y(du) = Kf(u) + \epsilon W(du),$$

where $u \in Q \subset R^2$. Then, we can make the projection on the vaguelette basis $\{\Psi_I\}$ and obtain the following Gaussian white-noise model in sequence space:

$$y_I = \theta_I + \epsilon z_I, \quad I \in \mathcal{I} = \cup_{j \geq 0} \mathcal{I}^{(j)}, \quad (11)$$

where $\mathcal{I}^{(0)} = \mathcal{I}_0^{(0)} \cup \mathcal{I}^{(0)}$, $z_I = [\gamma_I, W]$ and $\theta = (\theta_I) = \langle f, \psi_I \rangle$ is unknown. We may remark that the z_I are $\mathcal{N}(0, \sigma_j^2)$ where $\sigma_j = 2^{j/2}$.

Now, we only consider the sequence space model (11). Assume that $\theta \in \Theta_{spq}(M) = \{\theta : \|\theta\|_{spq} \leq M\}$, where, in this section,

$$\|\theta\|_{spq}^q = \sum_{j \geq 0} \left(2^{js'} \|\theta_j\|_p \right)^q$$

and $\|\theta_j\|_p^p = \sum_{I \in \mathcal{I}^{(j)}} |\theta_I|^p$.

Define a soft threshold rule $\hat{\theta}^\tau$ [9]. Then Donoho [9] showed that in model (11) we can define some wavelet thresholding estimator for which there exist absolute constants A_{spq} such that for correct choice of $\tau = (\tau_j)$, as $\epsilon \rightarrow 0$,

$$\sup_{\Theta_{spq}(M)} \mathbb{E}_\theta \|\hat{\theta}^\tau - \theta\|_2^2 \leq A_{spq} R_z \left(\epsilon, \Theta_{spq}(M) \right) (1 + o(1)), \quad (12)$$

where the nonlinear minimax risk verifies

$$R_z \left(\epsilon, \Theta_{spq}(M) \right) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta_{spq}(M)} \mathbb{E} \|\hat{\theta} - \theta\|_2^2 \asymp (\epsilon^2)^{2\alpha}, \quad (13)$$

with $\alpha = s/(2s+3)$.

The following approximation lemma is the key tool in bounding the Poisson intensity estimation risk by a corresponding white-noise model risk.

Lemma 5.1 Suppose that $\nu(S) = 1$ and $\int h^2 d\nu = 1$. Let $\|h\|_{L^\infty(S)} \leq H$ and $S_n = \int h d\bar{N}^n$, where N^n is a Poisson process with intensity measure $\nu^n = n\nu$. Then, there exist constants C_1 and C_2 such that, whenever $H^2 n^{-1} \log^3 n \leq C_1$, then there exists a standard Gaussian variable Z such that

$$\mathbb{E}\left(n^{-1/2}S_n - Z\right)^2 \leq C_2 H^2 n^{-1}.$$

PROOF. Let F_n denote the distribution of $W_n = n^{-1/2}S_n$. Since $\mathbb{E}\exp(itW_n)$ is absolutely integrable, F_n is absolutely continuous so that the quantile transformation $Z = \Phi^{-1}(F_n(W_n))$ yields a standard Gaussian variable. We show that Z has the desired approximation by considering in turn large, moderate and small deviations, defined respectively by sets $A_1 = \{w : |w| > \sqrt{a \log n}\}$, $A_2 = \{w : 1 \leq |w| \leq \sqrt{a \log n}\}$ and $A_3 = \{w : |w| \leq 1\}$. Write

$$\begin{aligned} \mathbb{E}(W_n - Z)^2 &= \mathbb{E}\left((W_n - Z)^2, |W_n| > \sqrt{a \log n}\right) + \int_{A_2 \cup A_3} \left[w - \Phi^{-1}(F_n(w))\right]^2 F_n(dw) \\ &= I_1 + I_2 + I_3. \end{aligned}$$

(a) For the small deviation I_3 , observe that $\sigma_n^2 = \int h^2 d\nu^n = n$ and $\rho_n = \int |h|^3 d\nu^n \leq Hn$. By Lemma 7.6, we have $r_n(x) = |F_n(x) - \Phi(x)| \leq 3\rho_n/\sigma_n^3 \leq 3Hn^{-1/2}$. According to the mean value theorem, $I_3 = \int_{-1}^1 r_n^2(w)/\phi^2(u^*(w)) F_n(dw) \leq CH^2 n^{-1}$ where $u^*(w)$ lies between w and $\Phi^{-1}(F_n(w))$.

(b) Large deviations are easily handled by the Hölder inequality, Bennett's inequality [see Lemma 7.2] and the moment bound. Specifically, observe that

$$I_1 \leq C\left(\mathbb{E}|W_n|^3 + \mathbb{E}|Z|^3\right)^{2/3} \mathbb{P}^{1/3}\left(|W_n| > \sqrt{a \log n}\right).$$

Using Lemma 7.2,

$$\mathbb{P}\left(|W_n| > \sqrt{a \log n}\right) \leq 2 \exp\left(-(1/2)(an \log n)B(H\sqrt{an \log n})\right) \leq 2n^{-3}$$

for large a . It follows from Lemma 7.1 and the Hölder's inequality that

$$\mathbb{E}|W_n|^3 \leq n^{-3/2} \left(\int h^2 d\nu^n\right)^{1/2} \left(\int h^4 d\nu^n + 3\left(\int h^2 d\nu^n\right)^2\right)^{1/2} \leq (H^2 n^{-1} + 3)^{1/2} \leq C.$$

Hence, $I_1 \leq Cn^{-1}$.

(c) For moderate deviations, write $I_2^+ = \int_1^{\sqrt{a \log n}} [w - \Phi^{-1}(\bar{F}_n(w))]^2 F_n(dw)$. From Lemma 7.1,

$$\mathbb{E}W_n^4 = n^{-2} \left(\int h^4 d\nu^n + 3\left(\int h^2 d\nu^n\right)^2\right) \leq C(M^2 n^{-1} + 1) \leq C. \quad (14)$$

By Lemma 7.7, Lemma 5 of [11] and (14), one can show that

$$I_2^+ \leq e^3 \int_1^{\sqrt{a \log n}} x^{-2} \left|\bar{F}_n(x)/\bar{\Phi}(x) - 1\right| F_n(dx) \leq C\mathbb{E}|W_n|^4 M^2 n^{-1} \leq CM^2 n^{-1}. \quad \square$$

5.2.2 Thresholding wavelet estimator

Among nonlinear estimators we study a truncated threshold wavelet estimator. Define the empirical vaguelette coefficients

$$\hat{\alpha}_I = \frac{1}{n} \int \gamma_I dG_g^n \quad \text{for } I \in \mathcal{I}_0^{(j)}, \quad (15)$$

and

$$\hat{\beta}_I = \frac{1}{n} \int \gamma_I dG_g^n = [\gamma_I, G_g], \quad I \in \mathcal{I}^{(j)}. \quad (16)$$

It can be seen that $\hat{\beta}_I$ is an unbiased estimator of the vaguelette coefficient β_I .

Consider estimators of the form

$$\hat{f}_n(\tau, j_1, j_2) = \sum_{I \in \mathcal{I}_{(0)}^{(j)}} \hat{\alpha}_I \Psi_I + \sum_{j_1}^{j_2} \sum_{I \in \mathcal{I}^{(j)}} \delta_S(\hat{\beta}_I, \tau_j) \Psi_I, \quad (17)$$

where $\hat{\alpha}_I$ and $\hat{\beta}_I$ are the empirical wavelet coefficients defined in (15) and (16), δ_S denotes a soft-threshold rule, $\tau = (\tau_j)$ is a threshold, j_1 is a fixed constant and $j_2 = j_2(n)$.

Since $s > 2/p$, we can choose

$$\frac{s}{2s+3} \frac{1}{s'} \log_2 n \ll j_2 \ll \frac{s+3}{2(2s+3)} \log_2 n - (3/4) \log_2 \log_2 n, \quad (18)$$

where $a_n \ll b_n$ means that $b_n - a_n \rightarrow \infty$.

Theorem 3 Suppose that $p \geq 1$ and $s > 2/p$. Let f_n^* is an estimator of the form (17) with the choice of τ, j_1 and j_2 checking (18). Then there exist $C_3 = C_3(s, p, q, M)$, $C_4 = C_4(s, p)$, $C_5 = C_5(s, p, q, M)$ and $C_6 = C_6(s, p, q, M)$ such that

$$\inf_{f_n^*} \sup_{f \in \mathcal{F}_{spq}(M)} \mathbb{E}_f \|f_n^* - f\|_2^2 \leq C_5 R_z \left(C_3 n^{-1/2}, \Theta_{spq}(C_4 M) \right) (1 + o(1)) \leq C_6 n^{-2\alpha},$$

as $n \rightarrow \infty$, where $\alpha = s/(2s+3)$.

PROOF. Note that $\mathbb{E}\hat{\beta}_I = \beta_I$ and $\text{var}\hat{\beta}_I = n^{-1}\sigma_I^2$, $\sigma_I^2 = \sigma_I^2(f) = \int \gamma_I^2 d\nu_g$. Let $h_I = \gamma_I/\sigma_I$ and note that $\int h_I^2 d\nu_g = 1$ and $\|h_I\|_{L_\infty(T)} \leq C_3 2^j/\sigma_I = H_I$, say. We construct $\hat{\eta}_I = \beta_I + n^{-1/2}\sigma_I Z_I$, where Z_I is given below.

(a) If $\sigma_I^2 \geq C_7^2 2^{2j} \log^3 n / C_1 n$, then use Lemma 5.1 to construct Z_I and note that

$$D \equiv \mathbb{E}(\hat{\beta}_I - \hat{\eta}_I)^2 = n^{-1}\sigma_I^2 \mathbb{E} \left(n^{-1/2} S_n - Z_I \right)^2 \leq C_7^2 C_2 2^{2j} n^{-2},$$

where $S_n = \int h_I d(G_g^n - \nu_g^n)$.

(b) If $\sigma_I^2 \leq C_7^2 2^{2j} \log^3 n / C_1 n$, choose an independent $Z \sim \mathcal{N}(0, 1)$ to obtain $D \leq 2\text{var}(\hat{\alpha}_I) + 2n^{-1}\sigma_I^2 \leq 4C_7^2 C_1^{-1} 2^{2j} n^{-2} \log^3 n$. In either case, we have, therefore, for all I, n ,

$$D \leq C_8 2^{2j} n^{-2} \log^3 n. \quad (19)$$

By (19), Lemma 3 in [11] and the fact that $|\delta_S(w_1, \tau) - \delta_S(w_2, \tau)| \leq |w_1 - w_2|$,

$$\mathbb{E}[\tilde{\beta}_I - \beta_I]^2 \leq 2C_8 2^{2j} n^{-2} \log^3 n + 4r(\delta_{\tau_I}, \beta_I; \bar{a}n^{-1/2}),$$

where $r(\delta_\tau, \beta; a)$ denotes the Gaussian mean squared error $\mathbb{E}[\delta_S(\beta + a2^{j/2}Z, \lambda) - \beta]^2$ for the estimation of β from a single Gaussian observation with mean β and variance $(a2^{j/2})^2$. Here \bar{a} is any common upper bound on σ_I^2 . For example, all intensities $f \in \mathcal{F}_{spq}(M)$ are uniformly bounded and so $\sigma_I^2 = \int \gamma_I^2 d\nu_g \leq C$.

(c) Let $S_j = \{I : 2^{-j}|k_x| < S + A \text{ and } 2^{-j}|k_y| < S + A\}$. Then

$$\mathbb{E}\|f_n^* - f\|_2^2 \leq Cn^{-1} + 4 \sum_j \sum_{I \in S_j} r(\delta_{\tau_j}, \beta_I; \bar{a}n^{-1/2}) + C2^{2j_2} n^{-2} \log^3 n + M^2 n^{-2j_2 s}. \quad (20)$$

The sum can be bounded by

$$A_{spq}(\bar{a}n^{-1/2}, \Theta_{spq}(2^{j_0 s} M))(1 + o(1))$$

for appropriate choice of τ_j , where $2^{j_0-1} \leq S + A < 2^{j_0}$. Using (18) we have that the final two right-side terms in (20) are of smaller order than $n^{-2s/(2s+3)}$ by the cutoff $j_2 = j_2(n)$. Using (13) we get the result. \square

6 Adaptation results

In Section 5 we have defined an estimator which attains the optimal rate of convergence on some Besov ball $\mathcal{F}_{spq}(M)$. However, the main drawback is that the estimator (18), depends on the values of s, p, q of the Besov ball. Clearly, prior knowledge of the exact smoothness parameter s and of p and q is not possible in applications. For this reason, adaptive methods have appeared (cf., [10] and [12]). The aim therefore, is to construct estimators which achieve the optimal rate of convergence, or almost the optimal rate, without the knowledge of the parameters of the Besov ball, i.e. estimators which adapt to the unknown smoothness.

Fix an integer r_0 and define a class

$$\mathcal{J} = \left\{ (s, p, q) : \frac{2}{p} < s < r_0, 1 \leq p, q \leq \infty \right\}. \quad (21)$$

The estimator f_{TW}^* , is obtained from compactly supported and $(r_0 + 1)$ -regular functions ϕ, ψ .

Define the following wavelet thresholding estimator associated with T, j_0 and j_1 :

$$f_{TW}^* = \sum_{I \in \mathcal{I}_0^{(j)}} \hat{\alpha}_I \Psi_I + \sum_{j_1}^{j_0} \sum_{I \in \mathcal{I}^{(j)}} \delta_H(\hat{\beta}_I, Tc_j) \Psi_I, \quad (22)$$

where $\hat{\alpha}_I$ and $\hat{\beta}_I$ are the empirical wavelet coefficients defined in (15) and (16), $c_j = 2^{j/2} \sqrt{j/n}$, $j_0 = j_0(n), j_1 = j_1(n)$, T is a positive constant, δ_H denotes a hard-threshold rule,

$$\delta_H(\hat{\beta}_I, Tc_j) = \begin{cases} \hat{\beta}_I & \text{if } |\hat{\beta}_I| > Tc_j, \\ 0 & \text{if } |\hat{\beta}_I| \leq Tc_j. \end{cases}$$

Furthermore we choose j_0 and j_1 such that

$$2^{j_1} \asymp n^{1/(3+2r_0)}, \quad 2^{j_0} \asymp n/\log_2 n. \quad (23)$$

The estimator f_{TW}^* is called adaptive in the sense that it does not depend on s, p, q and achieve the optimal rate of convergence, up to a logarithmic term, on each $\mathcal{F}_{spq}(M)$. Indeed, this means that for a given function f the estimator will automatically obtain almost the optimal rate of convergence, i.e. adapt to the unknown smoothness. This result would be of particular interest for applications in that the cost of a logarithmic factor in the rate of convergence, is more than offset by the advantage of no prior restrictions on s, p, q .

Theorem 4 *Let $f \in \mathcal{F}_{spq}(M)$ defined in (1), where $(s, p, q) \in \mathcal{J}$ defined in (21). Let f_{TW}^* be the estimator defined in (22) with j_0 and j_1 checking (23), then there exists $C_9 = C_9(s, p, q, M)$ such that as $n \rightarrow \infty$,*

$$\sup_{f \in \mathcal{F}_{spq}(M)} \left(\mathbb{E}_f \|f_{TW}^* - f\|_2^2 \right)^{1/2} \leq C_9 (\log n/n)^\alpha,$$

where $\alpha = s/(2s+3)$.

PROOF. By Lemma 3.3,

$$\int_T \left| \gamma_I(r, \omega) \right|^m \pi^{-1} dr d\omega = 2^{mj} \int_T \left| \gamma_{(0,0,\epsilon)}(2^j r - \zeta_k(\omega), \omega) \right|^m \pi^{-1} dr d\omega \leq C 2^{(m-1)j}. \quad (24)$$

By Lemma 7.1 and (24),

$$\mathbb{E}_f \left| \widehat{\beta}_I - \beta_I \right|^2 \leq C \frac{1}{n^2} \int \gamma_I^2 d\nu_g^n \leq C \frac{2^j}{n} \quad (25)$$

and

$$\mathbb{E}_f \left| \widehat{\beta}_I - \beta_I \right|^4 = \frac{1}{n^4} \left[\int \gamma_I^4 d\nu_g^n + 3 \left(\int \gamma_I^2 d\nu_g^n \right)^2 \right] \leq C \left(\frac{2^{3j}}{n^3} + \frac{2^{2j}}{n^2} \right). \quad (26)$$

There exists $C_{10} = C_{10}(M, \Psi_I)$ such that

$$\int \gamma_I^2 d\nu_g^n \leq C_{10} 2^j \text{ and } \|\gamma_I\|_{L_\infty(T)} \leq C_{10} 2^j.$$

If $j2^j \leq n$, then there exists a constant $T = C_{11}(M, \Psi_I)\eta$ such that, for all $\eta \geq 1$,

$$\mathbb{P} \left(|\widehat{\beta}_I - \beta_I| > (T/2)c_j \right) \leq 2^{-\eta j} \quad (27)$$

by Lemma 7.2.

Write $f = E_{j_1} f + D_{j_1, j_0} f + f - E_{j_0} f$, where $D_{j_1, j_0} f = \sum_{j=j_1}^{j_0} D_j f$. Note that by the definition of f_{TW}^* ,

$$\mathbb{E}_f \|f_{TW}^* - f\|_2^2 \leq 3 \left(\mathbb{E}_f \|f_{j_1}^* - E_{j_1} f\|_2^2 + \mathbb{E}_f \|\widehat{D}_{j_1, j_0} - D_{j_1, j_0} f\|_2^2 + \|f - E_{j_0} f\|_2^2 \right),$$

where $\widehat{D}_{j_1, j_0} = D_{j_1, j_0} f_{TW}^*$.

Consider the bias and linear terms. Using the fact that $B_{spq} \subset B_{s'2q}$, we have that, since $s'j_0 \geq 1$,

$$\|f - E_{j_0}f\|_2^2 \leq Cn^{-2\alpha}.$$

By Lemma 7.1, $\mathbb{E}_f \hat{\alpha}_I = \alpha_I$. From Lemma 3.3 we obtain

$$\mathbb{E}_f \|f_{j_1}^* - E_{j_1}f\|_2^2 = \sum_{I \in \mathcal{I}_0^{(j)}} \mathbb{E}_f |\hat{\alpha}_I - \alpha_I|^2 \leq Cn^{-2\alpha}.$$

For the detail terms, define

$$\begin{aligned} \hat{B}_j &= \left\{ I \in \mathcal{I}^{(j)} : |\hat{\beta}_I| > Tc_j \right\}, & \hat{S}_j &= \hat{B}_j^c \\ B_j &= \left\{ I \in \mathcal{I}^{(j)} : |\beta_I| > (T/2)c_j \right\}, & S_j &= B_j^c \\ B'_j &= \left\{ I \in \mathcal{I}^{(j)} : |\beta_I| > 2Tc_j \right\}, & S'_j &= B'_j{}^c. \end{aligned}$$

We may then write

$$\begin{aligned} \hat{D}_{j_1, j_0} - D_{j_1, j_0}f &= \sum_{j_1}^{j_0} \sum_{I \in \mathcal{I}^{(j)}} \left(\hat{\beta}_I - \beta_I \right) \Psi_I \left[I\{I \in \hat{B}_j S_j\} + I\{I \in \hat{B}_j B_j\} \right] \\ &\quad - \sum_{j_1}^{j_0} \sum_{I \in \mathcal{I}^{(j)}} \beta_I \Psi_I \left[I\{I \in \hat{S}_j B'_j\} + I\{I \in \hat{S}_j S'_j\} \right] \\ &= (e_{bs} + e_{bb}) + (e_{sb} + e_{ss}). \end{aligned}$$

For the term e_{bs} , we set $\hat{f}_I = (\hat{\beta}_I - \beta_I)I\{I \in \hat{B}_j S_j\}$ and

$$G_j = \left\{ I \in \mathcal{I}^{(j)} : \left| \hat{\beta}_I - \beta_I \right| \geq (T/2)c_j \right\},$$

the large-deviation event. Clearly, $\hat{B}_j S_j \subset G_j$. Using this, (26), (27) and the Cauchy-Schwarz inequality, we have

$$\mathbb{E}_f \|e_{bs}\|_2^2 = \sum_{j_1}^{j_0} \sum_I \mathbb{E}_f |\hat{f}_I|^2 \leq Cn^{-2\alpha}.$$

To give a bound for e_{sb} , we note that $\hat{S}_j B'_j \subset G_j$. Hence, if $\eta \geq 2r_0$ we have

$$\mathbb{E}_f \|e_{sb}\|_2^2 \leq C \leq C2^{-j_1(\eta+2s')} \leq n^{-2\alpha}.$$

For a bound on e_{bb} , it can be shown that

$$\begin{aligned} e_{bb} &= \left(\sum_{j_1(n)}^{j_1(s,p,q)} + \sum_{j_1(s,p,q)}^{j_0(n)} \right) \sum_I (\hat{\beta}_I - \beta_I) I\{I \in \hat{B}_j B_j\} \\ &\leq Cn^{-2\alpha} + \leq Cn^{-2\alpha}(\log n)^{-\omega} \end{aligned}$$

since $\rho = ps' - (2-p)/2 > 0$ and $j^{-p/2} \leq 1$. Finally, we consider the case e_{ss} . Let $\tau_I = 2^{-j/2}\beta_I$, and write $\tau = (\tau_I)_{I \in \mathcal{I}^{(j)}, j \geq j_1}$ and $\beta = (\beta_I)_{I \in \mathcal{I}^{(j)}, j \geq j_1}$. By the definition of Besov norm, if $\beta \in B_{spq}(M)$,

then $\tau \in B_{\bar{s}pq}(M)$, where $\bar{s} = s + 1/2$. Using the orthonormality of $\{\psi_I\}$ and the structure of sequence norms, we have

$$\|e_{ss}\|_2 \leq \left\| \left(\{\beta_I : j_1 \leq j \leq j_0, I \in S'_j\} \right) \right\|_{022} = \left\| \left(\{\tau_I : j_1 \leq j \leq j_0, I \in S'_j\} \right) \right\|_{\frac{1}{2}22}.$$

The condition $I \in S'_j$ implies $|\tau_I| \leq 2T\sqrt{j_0/n} \equiv \Lambda_n$. Let

$$\Omega^0(\Lambda; \|\cdot\|, B) = \sup \left\{ \|\tau\| : \tau \in B, |\tau_I| \leq \Lambda \right\}.$$

Clearly we have

$$(\mathbb{E}_f \|e_{ss}\|_2^2)^{1/2} \leq \Omega_n \equiv \Omega^0(\Lambda_n; \|\cdot\|_{\frac{1}{2}22}, B_{\bar{s}pq}(M)).$$

From Theorem 3 of [12], we read off that $\Omega_n \leq M^{1-2\bar{\alpha}}(2T\sqrt{j_0/n})^{2\bar{\alpha}}$, where $\bar{\alpha} = (\bar{s}-1/2)/(2\bar{s}+2) = \alpha$. Since $j_0 \asymp \log n$, we conclude from this argument that

$$\Omega_n \leq C \left(\frac{\log n}{n} \right)^\alpha.$$

7 Appendix: Basic Properties of Poisson Processes

This Appendix summarizes properties of spatial Poisson processes. We summarize results concerning moment and large-deviation bounds, Kullback-Leibler divergence between Poisson processes and an information inequality. Most results are straightforward and therefore are stated without proofs.

Let S be a state space which is a subset of \mathbb{R}^2 . Let N , N_1 and N_2 denote Poisson processes on (S, \mathcal{B}) with finite intensity measures, ν , ν_1 and ν_2 , respectively. For a Poisson process N with intensity measure ν , let $\bar{N} = N - \nu$.

Lemma 7.1 [23]

$$\mathbb{E} \exp \left(\xi \int h dN \right) = \exp \left(\int (e^{\xi h} - 1) d\nu \right) \quad \text{for } \xi \in \mathbb{C} \quad (28)$$

$$\mathbb{E} \int g dN = \int g d\nu \quad (29)$$

$$\text{var} \int g dN = \int g^2 d\nu \quad (30)$$

$$\text{cov} \left(\int g dN, \int h dN \right) = \int gh d\nu. \quad (31)$$

$$\mathbb{E} \left(\int g d\bar{N} \right)^4 = \int g^4 d\nu + 3 \left(\int g^2 d\nu \right)^2 \quad (32)$$

Lemma 7.2 Suppose that $\int h^2 d\nu \leq V$ and $\|h\|_{L^\infty(S)} \leq H$. Then for each $\eta > 0$,

$$\mathbb{P} \left(\left| \int h d\bar{N} \right| \geq \eta \right) \leq 2 \exp \left(-\frac{1}{2} \eta^2 B(H\eta V^{-1}) \right) \quad (33)$$

$$\leq 2 \exp \left(-\frac{1}{2} \eta^2 / (V + H\eta/3) \right), \quad (34)$$

where $B(x) = 2x^{-2}[(1+x)\log(1+x) - x]$ for $x > 0$.

We define the Kullback-Leibler distance Δ between N_1 and N_2 by

$$\Delta(N_1, N_2) = \int \log \frac{g_1}{g_2} d\mathcal{L}(N_1),$$

where g_i is a density of N_i with respect to some N . By Theorem 3.1.1 in [34],

$$g_j(\mu) = \left(\prod_{i=1}^{\mu(S)} h_j(x_i) \right) \exp(\nu(S) - \nu_j(S)) \quad (35)$$

if $\mu = \sum_{i=1}^{\mu(S)} \varepsilon_{x_i}$, where h_i is a density of ν_i with respect to ν . In correspondence to (35), we define the Kullback-Leibler distance Δ between ν_1 and ν_2 by

$$\Delta(\nu_1, \nu_2) = \int \left(h_1 \log \frac{h_1}{h_2} - h_1 + h_2 \right) d\nu.$$

Lemma 7.3 $\Delta(N_1, N_2) = \Delta(\nu_1, \nu_2)$.

Lemma 7.4 If $\int h_1 d\nu = \int h_2 d\nu = H$, then

$$\Delta(\nu_1, \nu_2) \leq \int \frac{(h_1 - h_2)^2}{h_2} d\nu.$$

Assume that the intensity function of the Poisson process N_θ , $\theta \in \Theta$, has intensity function f_θ and intensity measure ν_θ , where the parameter space Θ is an open subset of \mathbb{R} . We make standard regularity conditions where

$$I(\theta) = \text{var}_\theta \int \frac{\partial}{\partial \theta} \log f_\theta dN_\theta = \int \left(\frac{\partial}{\partial \theta} \log f_\theta \right)^2 d\nu_\theta,$$

is well defined.

Lemma 7.5 Suppose that (i), (ii) and (iii) hold and $0 < I(\theta) < \infty$. Then for all θ ,

$$\text{var}_\theta \int h dN_\theta \geq \frac{[\partial/\partial \theta \mathbb{E}_\theta \int h dN_\theta]^2}{I(\theta)}.$$

The following lemma is of importance and will be used in Section 5. It shows that the difference between the Poisson process and its intensity measure is ‘close’ to a Gaussian distribution.

Lemma 7.6 Suppose that N^n is a Poisson process with intensity measure ν^n such that $\nu^n = n\nu$ with $\nu(S) = 1$. Let $S_n = \int f d\bar{N}^n$, $\sigma_n^2 = \int f^2 d\nu^n$ and $\rho_n = \int |f|^3 d\nu^n$. Let F_n and Φ respectively denote the distribution functions of S_n and of a $\mathcal{N}(0, 1)$ random variable. Then

$$|F_n(x) - \Phi(x/\sigma_n)| \leq \frac{3\rho_n}{\sigma_n^3}.$$

PROOF. Let $\chi_Z(t) = \exp(-\sigma_n^2 t^2/2)$ and $\chi_{S_n}(t) = \exp\left(\int (\exp(itf) - 1 - itf) d\nu^n\right)$. Observe that

$$|D(t)| \equiv \left| \int \left(e^{itf} - 1 - itf + t^2 f^2/2 \right) d\nu^n \right| \leq \frac{\rho_n}{6} |t|^3.$$

Since $|e^z - 1| \leq |z| \exp(|z|)$ for all complex z ,

$$\left| \chi_{S_n}(t) - \chi_Z(t) \right| \leq \chi_Z(t) |D(t)| \exp(|D(t)|) \leq \frac{\rho_n}{6} |t|^3 \exp(\rho_n |t|^3/6 - \sigma_n^2 t^2/2).$$

Choose $T_n = 4\sigma_n^2/3\rho_n$. For $|t| \leq T_n$, $\rho_n |t|^3/6\sigma_n^3 - t^2/2 \leq -5t^2/18$. The desired result now follows from this and (3.13) in XVI.4 of [15]. \square

Lemma 7.7 *Let N^n be a Poisson process with intensity measure ν^n such that $\nu^n = n\nu$ with $\nu(S) = 1$. Set $L(h) = L(h, f) = \int \max(e^{hf}, 1) |f|^3 d\nu^n$, $S_n = \int f d\bar{N}^n$, $B^2 = \text{var} S_n$, $\bar{F}_n(x) = \mathbb{P}(S_n \geq x)$, $\bar{\Phi} = 1 - \Phi$ and $\phi(x) = \Phi'(x)$. Suppose that $0 < B < \infty$ and $x \geq 0$. If*

$$4xL(2x/B^2) \leq B^4,$$

then Cramér-Petrov series $\eta(x)$ is defined and satisfies

$$|\eta(x)| \leq x^3 B^{-6} L(2x/B^2) \tag{36}$$

$$|e^{-\eta(x)} \bar{F}_n(x) - \Phi(x/B)| \leq 29B^{-3} L(2x) \phi(x/B). \tag{37}$$

A SKETCH OF PROOF. Define $\Lambda(h) = \Lambda(h, S_n) = \log \mathbb{E} e^{hS_n} = \exp\left(\int (e^{hf} - 1 - hf) d\nu^n\right)$. Corresponding to S_n , introduce an attending random variable $S_n(h)$ which has the attending distribution: $\mathbb{E} \exp(itS_n(h)) = \mathbb{E} \exp((h + it)S_n) / \mathbb{E} \exp(hS_n)$. Obviously, $\mathbb{E} S_n(h) = \Lambda'_n(h)$, $B^2(h) \equiv \text{var} S_n(h) = \Lambda''(h)$, $\Lambda(0) = \Lambda'(0) = 0$ and $\Lambda''(0) = B^2(0) = B^2$. Set $L_0(h) = \int e^{hf} |f|^3 d\nu^n$ and $L^*(h) = L_0(h)/B^3(h)$. Then Theorem 1 and Corollary 1 of [35] are true for S_n , where Lemma 2 of [35] can be proved by the argument used to prove Lemma 7.6. Observe that $|\Lambda^{(3)}(h)| \leq L_0(h) \leq L(h)$ and $\max_{0 \leq t \leq 1} |\Lambda^{(3)}(th)| \leq L(h)$. Now by using the results in Sections 2 and 3 of [35], one can obtain the desired result. \square

Acknowledgment

The authors wish to thank Peter T. Kim and Eric D. Kolaczyk for various helpful discussions and comments. The authors wish to thank an Associate Editor and the referees who each made several helpful suggestions that improved the presentation and the results of the paper.

References

- [1] Abramovich, F. and Silverman, B. W. (1997). The vaguelette-wavelet decomposition approach to statistical inverse problems. *Biometrika* **85** 115–129.

- [2] Birgé, L. (1983) Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Geiete.* **65** 181–237.
- [3] Cavalier, L. (2000). Asymptotically efficient estimation of a density in tomography. *Ann. Statist.* **28** 330–347.
- [4] Cavalier, L. (2001). On the problem of local adaptive estimation in tomography. *Bernoulli* **7** 63–78.
- [5] Cavalier, L. and Tsybakov, A. B. (2001). Sharp adaptation for inverse problems with random noise. To appear in *Proba. Theory Relat. Fields*.
- [6] Choi, Y., Koo, J. -Y. and Lee, N. -Y. (2001). Image reconstruction using the wavelet transform for positron emission tomography. *IEEE Trans. Medical Imaging.* **20** 1188–1193.
- [7] Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.
- [8] Deans, S. R. (1983). *The Radon Transform and Some of its Applications*. Wiley, New York.
- [9] Donoho, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmonic Anal.* **2** 101–126.
- [10] Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Ann. Statist.* **81** 425–455.
- [11] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996a). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539.
- [12] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996b). Universal near minimaxity of wavelet shrinkage. In *Festschrift for Lucien Le Cam* (D. Pollard and G. Yang, eds.) Springer, New York.
- [13] Donoho, D. L. and Low, M. G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20** 944–970.
- [14] Fan, J. and Koo, J. -Y. (1999). Deconvoluting wavelet density estimators. Manuscript.
- [15] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*. Vol. 2. Wiley, New York.
- [16] Fessler, J. A. (1994). Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Trans. Medical Imaging* **13** 290–300.
- [17] Geman S. and McClure D. E. (1987). Statistical methods for tomographic image reconstruction. *Proc. 46 Sect. ISI, Bull. ISI* **52** 5–21.
- [18] Härdle, W, Kerkyacharian, G., Picard, D, and Tsybakov, A.B. (1998). *Wavelets, Approximation and Statistical Applications*. Springer.
- [19] Hoffman E. J., Huang S. C., Phelps M. E. and Kuhl D. E. (1981). Quantitation in positron emission computed tomography: 4. Effect of accidental coincidences. *J. Comput. Assisted Tomography* **5** 391–400.
- [20] Johnstone, I. M. (1999). Wavelet shrinkage for correlated data and inverse problems: adaptivity results. *Statistica Sinica* **9**, 51–83.
- [21] Jones, M. C. and Silverman, B. W. (1989). An orthogonal series density estimation approach to reconstructing positron emission tomography images. *J. Appl. Statist.* **16** 177–191.

- [22] Johnstone, I. M. and Silverman, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.* **18** 251–280.
- [23] Kingman, J. F. C. (1993). *Poisson Processes*. Clarendon Press, London.
- [24] Kolaczyk, E. D. (1996). A wavelet shrinkage approach to tomographic image reconstruction. *J. Amer. Statist. Assoc.* **91**, 1079–1090.
- [25] Koo, J. -Y. (1993). Optimal rates of convergence for nonparametric statistical inverse problems. *Ann. Statist.* **21**, 590–599.
- [26] Koo, J. -Y. (1998). Convergence rates for logspline tomography. *J. Multiv. Anal.* **67** 367–384.
- [27] Koo, J. -Y. and Chung, H. -Y. (1998). Log-density estimation in linear inverse problems. *Ann. Statist.* **26**, 335–362.
- [28] Korostelev, A. P. and Tsybakov, A. B. (1991). Optimal rates of convergence of estimators in a probabilistic setup of tomography problem. *Problems of Info. Trans.* **27**, 73–81.
- [29] Meyer, Y. (1992). *Wavelets and operators*. Cambridge University Press, Cambridge.
- [30] Moulin P. and Liu J. (2000). Statistical imaging and complexity regularization. *IEEE Trans. Information Theory* **46** 1762–1777.
- [31] Natterer, F. (1986). *The Mathematics of Computerized Tomography*. Wiley, Chichester.
- [32] O’Sullivan, F. (1995). A study of least squares and maximum likelihood for image reconstruction in positron emission tomography. *Ann. Statist.* **23** 1267–1300.
- [33] Politte D. G. and Snyder D. L. (1991). Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron emission tomography. *IEEE Trans. Medical Imaging* **10** 82–89.
- [34] Reiss, R. -D. (1993). *A Course on Point Processes*. Springer-Verlag, New York.
- [35] Sakhanenko, A. I. (1991). Berry-Esseen type estimates for large deviation probabilities. *Siberian Math. J.* **32** 647–656.
- [36] Silverman, B. W., Jones, M. C., Nychka, D. W. and Wilson, J. D. (1990). A smoothed EM algorithm to indirect estimation problems, with particular reference to stereology and emission tomography. *J. Roy. Statist. Soc. Ser. B* **52** 271–324.
- [37] Snyder D. L. and Miller M. I. (1985). The use of sieves to stabilize images produced with the EM algorithm for emission tomography. *IEEE Trans. Nuclear Science* **32** 3864–3871.
- [38] Snyder, D. L., O’Sullivan, J. A., Whiting, B. R., Murphy, R. J., Benac, J., Cataldo, J. A., Politte, D. G. and Williamson, J. F. (2001). Deblurring subject to nonnegativity constraints when known functions are present with application to object-constrained computerized tomography. *IEEE Trans. Medical Imaging* **20** 1009–1017.
- [39] Snyder, D. L., Schulz, T. J. and O’Sullivan, J. A. (1992). Deblurring subject to nonnegativity constraints. *IEEE Trans. Signal Processing* **40** 1143–1150.
- [40] Solo V. (2000). Limits to estimation in stochastic ill-conditioned inverse problems. *IEEE Trans. Information Theory* **46** 1872–1880.
- [41] Vardi, Y., Shepp, L. A. and Kaufman, L. (1985). A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.* **80** 8–37.
- [42] Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.* **24**, 466–484.