Towards adaptivity via a new discrepancy principle for Poisson inverse problems[*]

Grzegorz Mika and Zbigniew Szkutnik

AGH University of Science and Technology, Faculty of Applied Mathematics,
al. A. Mickiewicza 30, 30-059 Krakow, Poland
e-mail: grzegorz.mika@agh.edu.pl*;* szkutnik@agh.edu.pl

Abstract: A new form of the discrepancy principle for Poisson inverse problems with compact operators is proposed and discussed in relation to various other proposals. It is shown that filter-induced spectral regularization with a priori chosen smoothing parameter produces estimators that are rate-minimax under source conditions on the estimated function. With the discrepancy principle used for a posteriori choice of the smoothing parameter, the filter-induced solutions are consistent (in probability), but the convergence rates under source conditions are suboptimal, at least in the finitely smoothing case, which often happens when discrepancy principles are used in stochastic inverse problems. Finite sample performance of the proposed procedure applied to a stereological problem of Spektor, Lord and Willis is illustrated with a simulation experiment.

MSC2020 subject classifications: Primary 62G05, 93E10, 65J20, 45Q05.
Keywords and phrases: Poisson inverse problems, Morozov discrepancy principle, regularization, adaptive estimation.

Received July 2020.

1. Introduction

Let Π_{ng} be a Poisson process on a measure space $(\mathcal{Y}, \mathcal{B}_Y, \mu_Y)$ with an integrable intensity function ng w.r.t. to μ_Y. The parameter $n \in \mathbb{N}$ is known. It quantifies the "experiment size", and will asymptotically tend to infinity. The function g is unknown. Additionally, we assume that

$$g = \mathcal{K}f, \tag{1}$$

with some operator $\mathcal{K}$. The Poisson inverse problem consists in estimating f, given observed the Poisson process Π_{ng}. In applications, nf is usually itself an intensity function of an unobservable Poisson process on possibly another measure space, say $(\mathcal{X}, \mathcal{B}_X, \mu_X)$, and $\mathcal{K}$ is related to a data distortion mechanism that maps the points of the unobservable process on $\mathcal{X}$ to the points of the observable process Π_{ng} on $\mathcal{Y}$. Plethora of practical problems fit into this general scheme including, e.g., restoration of medical images, image deblurring in microscopy and astronomy, and some stereological problems.

In this article, we assume that $\mathcal{K}$ is a linear and compact operator from $L^2(\mathcal{X}, \mathcal{B}_X, \mu_X)$ to $L^2(\mathcal{Y}, \mathcal{B}_Y, \mu_Y)$ of the form

$$(\mathcal{K}f)(y) = \int K(x,y)f(x)\,\mathrm{d}\mu_X(x). \tag{2}$$

Compactness of $\mathcal{K}$ is ensured, e.g., if the kernel K is square-integrable, or if $\mathcal{X} = \mathcal{Y}$ is a compact subset of $\mathbb{R}^k$, μ_X is the Lebesgue measure and K is weakly singular. For simplicity, we also assume that $\mathcal{K}$ is injective. The goal is to estimate f under the L^2 loss. It is well-known that, in this setup, $\mathcal{K}^{-1}$ is unbounded as an operator on $\mathrm{Im}\,\mathcal{K}$ and, in effect, some sort of regularization is necessary. For a general introduction to Poisson processes and the related inverse problems, we refer to [25, 18, 20].

A good introduction to regularization techniques and related problems in a non-stochastic setup is given, e.g., in [14]. A general treatment of minimax convergence rates of so-called filter-induced estimators under various noise models can be found in [4]. Typically, optimal values of regularization parameters depend on the, generally a priori unknown, regularity of the estimated function. Therefore, in all cases, data-driven procedures of choosing the value of the regularization parameter, preferably leading to adaptive estimators, are of central interest. Adaptivity is meant here as the ability to attain the same (preferably minimax) convergence rates as those achieved in cases with a priori known smoothness. Adaptivity issues for inverse problems were studied mainly under Gaussian white noise assumptions. For a recent review see, e.g., [5] and [6], where, in particular, limits for the performance of stopping criteria based on residuals are obtained for a discretized Gaussian white noise model and various spectral filter estimators. In [15, 16], a discrepancy principle is studied in a general stochastic inverse problem setup with unknown error distribution, under the assumption that repeated, independent, noisy measurements of the left-hand side of equation (1) are available. In the present article, we focus on the Poissonian noise, not satisfactorily covered in [15], but we borrow several ideas from that article.

Poisson inverse problems were studied in some generality in, e.g., [29, 30, 1, 17]. Rate minimax solutions of such problems with L^2 risk were found mainly on case-by-case basis (e.g., [8, 13, 32]). Adaptive thresholded wavelet solutions under the L^2 risk, rate-minimax over some Besov balls (up to a log factor) were found for a stereological Spektor-Lord-Willis problem in [10, 11]. In [12], a general Goldenshluger-Lepski oracle-inequality approach was adopted to the same problem and adaptivity of a specific kernel-based estimator was proved over a scale of restricted Sobolev spaces. More general adaptive thresholded wavelet solution was constructed in [1], but only for the Kullback-Leibler (KL) risk. A variant of the Lepski balancing principle suitable for Poisson inverse problems was also studied in [17] in connection with penalized maximum likelihood estimators. In a recent article [19], a circular deconvolution was studied for Poisson data with weighted L^2 risk, an adaptive solution based on model selection was constructed, and many additional references on Poisson inverse problems were given.

An alternative approach to the problem of data-driven choice of regularization parameters was proposed in [34] (see also [2]). A version of the Morozov discrepancy principle suitable for discretized Poisson inverse problems was constructed there, based on matching the KL divergence to its expectation. Existence and uniqueness of the resulting penalized maximum likelihood estimators were proved in [3]. This was further refined in [26] by proving consistency and convergence rates of the KL risk under some source conditions. However, as the KL distance primarily measures the closeness of the distributions and our main interest is in function rather than distribution estimation, we prefer to work with the L^2 loss.

In this article, we propose a new version of the discrepancy principle for Poisson inverse problems that, when compared to [3, 26], is applicable to a broader class of estimators (i.e., not just penalized maximum likelihood estimators), leads to consistent (in probability) filter-induced estimators with explicit and close to optimal rates of convergence under L^2 distance, and does not require any discretization, which may be of importance, because it is known that discretization effects are not always negligible and may affect the convergence rates ([29, 30]).

2. The new discrepancy principle

A nice feature of Poisson inverse problems is that the noise level depends on the signal and may thus be estimated from data. Efficient, $\sqrt{n}$-consistent estimation of the noise level is possible in the pre-conditioned model

$$q := \mathcal{K}^* g = \mathcal{K}^* \mathcal{K} f \tag{3}$$

and our version of the Morozov discrepancy principle will be formulated in terms of q, its estimator $\hat{q}_n$ and the noise level $\delta_n^2 := \mathbb{E}\|\hat{q}_n - q\|^2$. The adjoint operator has the form $(\mathcal{K}^* g)(x) = \int K(x, y) g(y) \, d\mu_Y(y)$ and the observable process can be represented as so-called mixed empirical process (cf., [25, p. 17])

$$\Pi_{ng} = \sum_{j=1}^{N_n} \delta_{Y_j}, \tag{4}$$

where N_n is Poisson distributed with mean $n \int g \, d\mu_Y$, δ_a is the Dirac measure concentrated at a, and $Y_j, j = 1, 2, \ldots$ form an i.i.d. sample from a density proportional to g, independent of N_n. Assume that K is non-negative, which is satisfied in all typical applications. Then, by Campbell's theorem (cf., e.g., [18, Sec. 3.2]),

$$\mathbb{E} \int K(x, y) \, d\Pi_{ng}(y) = \mathbb{E} \sum_{j=1}^{N_n} K(x, Y_j) = n \int K(x, y) g(y) \, d\mu_Y(y) = n(\mathcal{K}^* g)(x),$$

which means that

$$\hat{q}_n(x) := n^{-1} \sum_{j=1}^{N_n} K(x, Y_j) \tag{5}$$

is an unbiased estimator of $q(x)$. Moreover, if $\int K(x,y)g(y)\,\mathrm{d}\mu_Y(y)$ is finite, then, again by Campbell's theorem,

$$\operatorname{Var}\hat{q}_n(x) = n^{-2}\operatorname{Var}\int K(x,y)\,\mathrm{d}\Pi_{ng}(y) = n^{-1}\int K^2(x,y)g(y)\,\mathrm{d}\mu_Y(y)$$

and

$$\mathbb{E}\|\hat{q}_n - q\|^2 = \int \operatorname{Var}\hat{q}_n(x)\,\mathrm{d}\mu_X(x) = n^{-1}\int w(y)g(y)\,\mathrm{d}\mu_Y(y) = n^{-2}\,\mathbb{E}\sum_{j=1}^{N_n} w(Y_j),$$

with

$$w(y) := \int K^2(x,y)\,\mathrm{d}\mu_X(x), \tag{6}$$

which means that the noise level

$$\delta_n^2 := \mathbb{E}\|\hat{q}_n - q\|^2 = n^{-1}\int K^2(x,y)g(y)\,\mathrm{d}\mu_X(x)\,\mathrm{d}\mu_Y(y) \tag{7}$$

may be unbiasedly estimated with

$$\hat{\delta}_n^2 := n^{-2}\sum_{j=1}^{N_n} w(Y_j). \tag{8}$$

Let $\hat{f}_\alpha$ be an estimator of f with a regularization parameter α. We propose to select α as an approximate solution of the following discrepancy principle (DP) equation

$$\|\mathcal{K}^*\mathcal{K}\hat{f}_\alpha - \hat{q}_n\|^2 = \tau\hat{\delta}_n^2 \tag{9}$$

with $\hat{q}_n$ and $\hat{\delta}_n^2$ defined in (5)-(8) and with some fixed positive τ. This is a natural reformulation for the pre-conditioned model of the general Morozov discrepancy principle, as described, e.g., in [14].

In general, the DP equation (9) does not have to have a solution and, even if the solution exists, it does not have to be unique. If the left-hand side of (9) equals zero for $\alpha = 0$, as it will be the case (in the limit) for filter-induced regularizations studied later in this article, we set, for definiteness,

$$\alpha_n := \sup\left\{\alpha : \|\mathcal{K}^*\mathcal{K}\hat{f}_\alpha - \hat{q}_n\|^2 < \tau\hat{\delta}_n^2\right\}, \tag{10}$$

and refer to $(\alpha_n)_{n\in\mathbb{N}}$ as the DP sequence of regularization parameters.

A similar idea of building the DP equation for the pre-conditioned model was studied in [7] for inverse problems with Gaussian white noise error. In that case, it proved necessary to use a weighted version of the discrepancy in order to achieve optimal rates of convergence of DP-based solutions with a priori known smoothness. A posteriori rates, both in [7] and in our Poissonian case are, however, suboptimal (and, in fact, identical, as will be seen later).

The idea proposed in [7] was further developed (including some lower bounds) in [28], to allow for using residuals less smoothed than in [7], which proved

essential for adaptation over some range of Sobolev-type ellipsoids. A version of DP based on smoothed residuals was also recently developed in [9] for non-parametric estimation of a regression function within the framework of learning algorithms embedded in a reproducing kernel Hilbert space. Again, smoothing the residuals proved essential for adaptation, especially over ranges of faster rates.

3. Filter-induced regularizations and estimators with a priori smoothing

Theoretical properties of the discrepancy principle proposed in the previous section will be studied for the family of estimators induced by spectral filters. In this section, we begin, however, with studying the properties of those estimators with a priori chosen smoothing level. Recall that, for an operator $\mathcal{K} \in \mathcal{L}(\mathcal{F}, \mathcal{G})$ (space of linear, bounded maps from $\mathcal{F}$ to $\mathcal{G}$), the family of operators $\{R_\alpha \in \mathcal{L}(\mathcal{G}, \mathcal{F}), \alpha > 0\}$ is called regularization of $\mathcal{K}^{-1}$ with $\alpha \to 0$, if $R_\alpha g \to \mathcal{K}^{-1}g$ for all $g \in \operatorname{Im}\mathcal{K}$, when $\alpha \to 0$. For Hilbert spaces $\mathcal{F}$ and $\mathcal{G}$, and for compact operator $\mathcal{K}$, filter-induced regularizations can be conveniently defined in terms of the singular value decomposition SVD: there exist a sequence $\sigma_1 \geq \sigma_2 \geq \cdots > 0$ of singular values, finite or converging to zero, and corresponding sequences $(v_i)_i$ and $(u_i)_i$ of orthonormal right and left singular vectors such that $\overline{\operatorname{Span}}(v_i)_i = \operatorname{Ker}^\perp \mathcal{K}$, $\overline{\operatorname{Span}}(u_i)_i = \overline{\operatorname{Im}\mathcal{K}}$ and, for all i, $\mathcal{K}v_i = \sigma_i u_i$ and $\mathcal{K}^* u_i = \sigma_i v_i$. Then,

$$\mathcal{K}^{-1}g = \sum_i \sigma_i^{-1} \langle g, u_i \rangle v_i.$$

Here and in what follows $\langle \cdot, \cdot \rangle$ denotes the inner product in respective Hilbert spaces.

DEFINITION 1. *A family $\{F_\alpha, \alpha > 0\}$ of bounded functions $(0, \sigma_1^2] \to \mathbb{R}$ is called a regularizing filter, if $F_\alpha(\lambda) \to 1/\lambda$ for all λ, as $\alpha \to 0$ and $F_\alpha(\lambda) \leq C_R/\lambda$ for all α and λ, with some constant C_R.*

It is easily seen that

$$R_\alpha g = F_\alpha(\mathcal{K}^*\mathcal{K})\mathcal{K}^* g := \sum_i F_\alpha(\sigma_i^2)\sigma_i \langle g, u_i \rangle v_i$$

is then, indeed, a regularization of $\mathcal{K}^{-1}$. The family of filter-induced regularizations includes, e.g., the k-times iterated Tikhonov regularization ($F_\alpha(\lambda) = ((\lambda + \alpha)^k - \alpha^k)/(\lambda(\lambda + \alpha)^k))$, the truncated SVD ($F_\alpha(\lambda) = 1/\lambda \cdot I(\lambda \geq \alpha)$) and the Landweber iteration ($F_{1/k}(\lambda) = \beta \sum_{j=0}^{k-1}(1 - \beta\lambda)^j$, with $\beta \in (0, \sigma_1^{-2})$) (cf., e.g., [4]).

In order to control the variance of the estimators, we will usually consider so-called bounded filters.

DEFINITION 2. *The filter $\{F_\alpha, \alpha > 0\}$ is called a bounded filter, if there exists a positive constant C_F such that $|F_\alpha(\lambda)| \leq C_F/\alpha$ for all $\lambda \in (0, \|\mathcal{K}\|^2]$ and all $\alpha > 0$.*

Since, in our setup, $\hat{q}_n$ estimates $q = \mathcal{K}^* g$, it seems natural to define $\hat{f}_\alpha := F_\alpha(\mathcal{K}^*\mathcal{K})\hat{q}_n$. Notice, however, that $F_\alpha(\mathcal{K}^*\mathcal{K})$ is an operator in $L^2(\mathcal{X}, \mathcal{B}_X, \mu_X)$ but, in general, $\hat{q}_n$ does not have to be in $L^2(\mathcal{X}, \mathcal{B}_X, \mu_X)$. Nevertheless, even in that case $\hat{f}_\alpha$ may be well defined as an element of $L^2(\mathcal{X}, \mathcal{B}_X, \mu_X)$ with

$$\hat{f}_\alpha = F_\alpha(\mathcal{K}^*\mathcal{K})\hat{q}_n := \sum_i F_\alpha(\sigma_i^2) \langle \hat{q}_n, v_i \rangle \, v_i, \tag{11}$$

where

$$\langle \hat{q}_n, v_i \rangle := n^{-1} \sum_{j=1}^{N_n} \langle K(\cdot, Y_j), v_i \rangle = \sigma_i/n \sum_{j=1}^{N_n} u_i(Y_j) = \sigma_i/n \int u_i \, \mathrm{d}\Pi_{ng}, \tag{12}$$

if $\|F_\alpha(\mathcal{K}^*\mathcal{K})\hat{q}_n\|^2 = \sum_i F_\alpha^2(\sigma_i^2) \langle \hat{q}_n, v_i \rangle^2 < \infty$, which holds true with probability one, if

$$\mathbb{E}\|F_\alpha(\mathcal{K}^*\mathcal{K})(\hat{q}_n - q)\|^2 = \sum_i F_\alpha^2(\sigma_i^2)\mathrm{Var}\,\langle \hat{q}_n, v_i \rangle \tag{13}$$

$$= n^{-1} \sum_i \sigma_i^2 \, F_\alpha^2(\sigma_i^2) \, \langle u_i^2, g \rangle < \infty.$$

The second equality in (13) holds true because of (12) and the Campbell's theorem. Under condition (13), we thus take (11) and (12) as the definition of the estimator.

Convergence rates of this estimator will be studied for smoothness classes defined via so-called source condition (cf., e.g., [4])

$$f \in \mathcal{F}_{\nu,\rho} := \left\{ f = (\mathcal{K}^*\mathcal{K})^{\nu/2}s : \|s\| \le \rho \right\}, \tag{14}$$

with some positive ν and ρ. For finitely smoothing $\mathcal{K}$ with $\sigma_i \asymp i^{-b}$, this is equivalent to

$$f \in E_{b\nu,R} := \left\{ f : \sum_i i^{2b\nu} \langle f, v_i \rangle^2 \le R^2 \right\},$$

with some R, and one may expect minimax L^2 convergence rates $n^{-2b\nu/(2b\nu+2b+1)}$ (cf. [23, 4]).

Recall, that the qualification of the filter $\{F_\alpha, \alpha > 0\}$ is the maximal ν_0 such that for any $\nu \in (0, \nu_0]$ there exists a constant $C_\nu > 0$ such that

$$\sup_{\lambda \in (0, \|\mathcal{K}\|^2]} \lambda^{\nu/2} \, |\lambda F_\alpha(\lambda) - 1| \le C_\nu \alpha^{\nu/2}.$$

The following theorem shows that, in the finitely smoothing case, filter-induced estimators with a priori regularization attain the expected convergence rates uniformly over subsets of $\mathcal{F}_{\nu,\rho}$ defined by an additional boundedness condition on $\|\mathcal{K}f\|_\infty$. Typically, the rates are minimax over $\mathcal{F}_{\nu,\rho}$ (cf. [23, 4]). In

many cases, however, the rates remain minimax over subsets of $\mathcal{F}_{\nu,\rho}$ with sufficiently large upper bound for $\|\mathcal{K}f\|_\infty$. This is easily seen, for example, for the stereological Spektor-Lord-Willis problem, studied in [13, 32], where the finite subsets of $\mathcal{F}_{\nu,\rho}$, used in the direct construction of lower bounds for the risk, have an obvious uniform upper bound for $\|\mathcal{K}f\|_\infty$.

THEOREM 1. *Let $\{F_\alpha, \alpha > 0\}$ be a regularizing filter with qualification ν_0 and let the singular values of an injective, compact operator $\mathcal{K}$ satisfy $\sigma_i \asymp i^{-b}$. Let $f \in \mathcal{F}_{\nu,\rho}$ with $\nu \le \nu_0$, and $\|\mathcal{K}f\|_\infty \le M < \infty$. If either $b > 1/2$ and the filter is bounded, or $b \le 1/2$ and $F_\alpha(\lambda) = 0$ for $\lambda \le \alpha$, then*

$$\mathbb{E}\|\hat{f}_{\alpha(n)} - f\|^2 \le Cn^{-\frac{2b\nu}{2b\nu+2b+1}},$$

if $\alpha(n) \asymp n^{-2b/(2b\nu+2b+1)}$, with $C = C(\nu, \rho, C_F, C_\nu, \mathcal{K}, M)$.

Although, as shortly discussed below, filter-induced estimators, at least for the Hilbert-Schmidt case with $b > 1/2$, may also be constructed and analysed by either embedding our problem in the general noise model considered in [4], or using the white noise model results in [7], we nevertheless decided to provide (see, Appendix) a direct proof of Theorem 1, because it also covers the case $b \le 1/2$.

To see the connection with the general noise model, let $(\Omega, \mathcal{F}, \mathrm{P})$ be the basic probability space and write our model of observations in the form analogous to formula (2.11) in [4]

$$Y = g + n^{-1/2}\varepsilon,$$

where $Y := n^{-1}\Pi_{ng}$ is a Hilbert space process, i.e., a bounded, linear operator $L^2(\mathcal{Y}, \mathcal{B}_Y, \mu_Y) \to L^2(\Omega, \mathcal{F}, \mathrm{P})$ defined with

$$\phi \to \langle Y, \phi \rangle := n^{-1}\sum_{j=1}^{N_n} \phi(Y_j) = n^{-1}\int \phi\, \mathrm{d}\Pi_{ng},$$

and $\varepsilon := \sqrt{n}(Y - g)$ is the scaled Wiener-Ito integral (cf. [20]), also understood as a Hilbert space process with the action

$$\phi \to \langle \varepsilon, \phi \rangle := \sqrt{n}\left(\langle Y, \phi \rangle - \langle g, \phi \rangle\right).$$

Then

$$\langle \mathcal{K}^* Y, \phi \rangle = \langle Y, \mathcal{K}\phi \rangle = n^{-1}\sum_{j=1}^{N_n}(\mathcal{K}\phi)(Y_j) = n^{-1}\sum_{j=1}^{N_n}\int K(x, Y_j)\phi(x)\, \mathrm{d}\mu_X(x)$$

$$= \left\langle n^{-1}\sum_{j=1}^{N_n} K(\cdot, Y_j), \phi \right\rangle,$$

which means that

$$(\mathcal{K}^* Y)(\cdot) = n^{-1}\sum_{j=1}^{N_n} K(\cdot, Y_j) = \hat{q}_n(\cdot),$$

and $\mathcal{K}^*Y$ is thus a random element in $L^2(\mathcal{X}, \mathcal{B}_X, \mu_X)$ for square-integrable K. Further,

$$\mathcal{K}^*\varepsilon = \sqrt{n}(\mathcal{K}^*Y - \mathcal{K}^*g) = \sqrt{n}(\hat{q}_n - q)$$

and $\mathbb{E}\|\mathcal{K}^*\varepsilon\|^2 = n\,\mathbb{E}\|\hat{q}_n - q\|^2 < \infty$ for the Hilbert-Schmidt case and under assumptions of Theorem 1 (cf. (7)). It follows directly from, e.g., Lemma 12.2 in [20] that the covariance operator of ε is the "multiplication by g" operator $M_g : \phi \to M_g\phi := g\phi$ and its norm is $\|M_g\| = \|g\|_\infty \leq M$ under the assumptions of Theorem 1. Also,

$$\mathrm{Var}\langle \mathcal{K}^*\varepsilon, v_s \rangle = n\mathrm{Var}\langle \hat{q}_n, v_s \rangle \leq \sigma_s^2 M =: \rho(s),$$

so that condition (2.14) and the whole Assumption 1 in [4] are satisfied and the general theory applies. In particular, $\hat{f}_\alpha = F_\alpha(\mathcal{K}^*\mathcal{K})\mathcal{K}^*Y$,

$$\mathbb{E}\|\hat{f}_\alpha - \mathbb{E}\hat{f}_\alpha\|^2 = \mathbb{E}\|F_\alpha(\mathcal{K}^*\mathcal{K})\mathcal{K}^* n^{-1/2}\varepsilon\|^2 \leq Mn^{-1}\sum_s F_\alpha^2(\sigma_s^2)\sigma_s^2$$

and the last expression is the special case of the upper bound in formula (3.5) in [4], obtained as the integral w.r.t. the counting measure on $\mathbb{N}$. The rates for the finitely smoothing Hilbert-Schmidt case with $b > 1/2$ thus follow from Section 5.3 in [4].

The white noise model results from [7] can also be used in our problem, because our basic bias-variance decomposition (see the proof of Theorem 1)

$$\mathbb{E}\|\hat{f}_\alpha - f\|^2 \leq \|F_\alpha(\mathcal{K}^*\mathcal{K})q - f\|^2 + Mn^{-1}\mathrm{Tr}\left(F_\alpha^2(\mathcal{K}^*\mathcal{K})\mathcal{K}^*\mathcal{K}\right)$$

is formally identical to analogous formula (11) in [7] and the whole rate calculation in [7] applies. In particular, not only the rates given in Theorem 1 follow (for $b > 1/2$), but one also covers the infinitely smoothing case with a weaker source condition traditionally studied with such operators, which is summarized for completeness in the following theorem.

THEOREM 2. *Let $\{F_\alpha, \alpha > 0\}$ be a bounded regularizing filter that satisfies the following qualification condition: there exist positive constants r and γ such that for all $\alpha > 0$ and with $\psi(t) := (-\log t)^{-r}$*

$$\sup_{0 < \lambda \leq \|\mathcal{K}\|^2} \psi(\lambda)\,|\lambda F_\alpha(\lambda) - 1| \leq \gamma\psi(\alpha),$$

and let the singular values of an injective, compact operator $\mathcal{K}$ satisfy $\sigma_i \asymp \exp[-ci]$, with some positive c. If $f = \psi(\mathcal{K}^*\mathcal{K})s$, with some $\|s\| \leq \rho$ is such that $\|\mathcal{K}f\|_\infty \leq M < \infty$ and $\alpha(n)$ are chosen to satisfy $\alpha(n)/(-\log \alpha(n))^{2r+1} \asymp n^{-1}$, then

$$\mathbb{E}\|\hat{f}_{\alpha(n)} - f\|^2 \leq C/(\log n)^r,$$

with C depending only on the filter and operator and on the function class parameters.

Finally, the polynomial case with $b > 1/2$ can also be viewed in relation to recent results in [15], because $\hat{q}_n$ can be thought of as an arithmetic mean of n independent copies of $\int K(\cdot, y) \, d\Pi_g(y)$. If $\|g\|_\infty \leq M$, one has $\mathbb{E} \|\hat{q}_n - q\|^2 = \mathcal{O}(n^{-1})$, and one could try to directly use Theorem 2 in [15] to our preconditioned model (3), or $\mathcal{A}f = q$, with $\mathcal{A} := \mathcal{K}^*\mathcal{K}$. This would, however, mean that one uses regularization $R_\alpha^0 := F_\alpha(\mathcal{A}^2)\mathcal{A}$ and the estimator

$$\hat{f}_\alpha^0 = R_\alpha^0 \hat{q}_n = \sum_i F_\alpha(\sigma_i^4)\sigma_i^2 \langle \hat{q}_n, v_i \rangle v_i,$$

with the source condition $f = (\mathcal{A}^2)^{r/2}s = (\mathcal{K}^*\mathcal{K})^r s$. Theorem 2 in [15] would then give for $\hat{f}_\alpha^0$ the convergence rate $n^{-r/(r+1)}$, which would correspond to $n^{-\nu/(\nu+2)}$ under our source condition $f = (\mathcal{K}^*\mathcal{K})^{\nu/2}s$. Our Theorem 1 gives for our estimator

$$\hat{f}_\alpha = R_\alpha \hat{q}_n = \sum_i F_\alpha(\sigma_i^2)\langle \hat{q}_n, v_i \rangle v_i$$

the rate $n^{-2b\nu/(2b\nu+2b+1)}$, which is strictly better for $b > 1/2$. Note that $\mathbb{E}\|\hat{q}_n - q\|^2 < \infty$ is necessary for the applicability of Theorem 2 in [15], and that this condition is roughly the same as $b > 1/2$.

4. Consistency and rates of convergence with discrepancy principle

Analysis of the properties of filter-induced estimators with regularization parameters selected with our version of the discrepancy principle will use both some classic techniques, described, e.g., in [14], and some new ideas proposed recently in [15]. As discussed in the latter article, consistency with L^2 risk cannot generally be expected, if plain discrepancy principle is used. Intuitively, this is because it can happen, although with diminishing probability, that the true noise level gets significantly underestimated, which may lead to huge L^2 loss and, in effect, to L^2 risk not converging to zero. We illustrate that with a counterexample that is an adaptation of a similar example from [15] to our Poissonian setup. On the other hand, L^2-consistency can usually be guaranteed, although possibly with very slow rates, by adding to the DP-procedure a so-called "emergency stop" (cf., e.g., [15]). We, however, do not follow the latter idea and only prove convergence to zero in probability of the L^2 loss, i.e., weak consistency, and obtain corresponding convergence rates, both in close parallel to the development in [15].

4.1. Counterexample for convergence of L^2 risk

Let $\mathcal{X} = \mathcal{Y} = \mathbb{N}$ and let $\mu_X = \mu_Y$ be the counting measure on $2^\mathbb{N}$. With the canonical basis $(v_i)_{i \in \mathbb{N}}$ in $\ell^2(\mathbb{N})$, define a diagonal operator $\mathcal{K} : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ with $\mathcal{K}v_i = 10^{-i}v_i$, $i \in \mathbb{N}$. Equivalently, in terms of equation (2) and with $f = (f_1, f_2, ...) \in \ell^2(\mathbb{N})$ and $g = (g_1, g_2, \dots) = \mathcal{K}f$, we have $g_j = \sum_{i=1}^\infty K(i, j)f_i = 10^{-j}f_j$, $j \in \mathbb{N}$, with $K(i, j) = 10^{-j}\delta_{i,j}$, where $\delta_{i,j}$ stands for the Kronecker

delta. We observe a Poisson process Π_{ng} on $\mathbb{N}$ or, equivalently, the independent Poissonian counts, say $(Z_i)_{i \in \mathbb{N}}$, with means ng_i, where Z_i is the number of points Y_j in the representation (4) of Π_{ng} that are equal to i. Then, obviously, the quantities defined in Section 2 take the form $q = (10^{-i}g_i)_{i \in \mathbb{N}}$, $\hat{q}_n = n^{-1}(10^{-i}Z_i)_{i \in \mathbb{N}}$, $w = (10^{-2i})_{i \in \mathbb{N}}$ and $\hat{\delta}_n^2 = n^{-2}\sum_{i=1}^{\infty} 10^{-2i}Z_i$. Consider the truncated SVD estimator

$$\hat{f}_\alpha = F_\alpha(\mathcal{K}^*\mathcal{K})\hat{q}_n = n^{-1}\left(10^i Z_i I(10^{-2i} \geq \alpha)\right)_{i \in \mathbb{N}}.$$

It is more natural to parametrize this estimator directly with the number, say k, of retained components and write

$$\hat{f}_k = n^{-1}\left(10^i Z_i I(i \leq k)\right)_{i \in \mathbb{N}}.$$

Denote with $(k_n)_{n \in \mathbb{N}}$ the corresponding DP sequence, i.e.

$$k_n := \min\left\{k : \|\mathcal{K}^*\mathcal{K}\hat{f}_k - \hat{q}_n\| < \tau\hat{\delta}_n^2\right\}. \tag{15}$$

Consequently, with $\tau = 1$, the inequality in (15) can be written as

$$\sum_{i=1}^{\infty}\left(10^{-2i}(\hat{f}_k)_i - n^{-1}10^{-i}Z_i\right)^2 < n^{-2}\sum_{i=1}^{\infty} 10^{-2i}Z_i,$$

or, equivalently, as

$$\sum_{i>k} 10^{-2i}Z_i^2 < \sum_{i=1}^{\infty} 10^{-2i}Z_i. \tag{16}$$

Define $\Omega_n := \{Z_1 = \cdots = Z_n = 0, Z_{n+1} > 0\}$ and note that on Ω_n one necessarily has $k_n \geq n+1$. (Otherwise, the effective ranges of summation would be $i \geq n+1$ on both sides of (16) and the inequality could not be fulfilled.)

Set $f := (c/i)_{i \in \mathbb{N}}$, with $c = 9\log 5$. Then, $g = (c/(i10^i))_{i \in \mathbb{N}}$ and, for n large enough,

$$\mathrm{P}(\Omega_n) = \left(1 - \exp\left[-nf_{n+1}10^{-(n+1)}\right]\right)\prod_{i=1}^{n}\exp\left[-nf_i 10^{-i}\right]$$

$$\geq \frac{nc}{2(n+1)10^{n+1}}\exp[-nc/9] \geq \frac{nc}{2(n+1)10^{n+1}5^n}.$$

On Ω_n one has

$$\|\hat{f}_{k_n} - f\|^2 \geq \left(\left(\hat{f}_{k_n}\right)_{n+1} - f_{n+1}\right)^2 = \left(\frac{1}{n}10^{n+1}Z_{n+1} - \frac{c}{n+1}\right)^2 \geq \frac{10^{2(n+1)}}{2n^2},$$

for n large enough and, finally,

$$\mathbb{E}\|\hat{f}_{k_n} - f\|^2 \geq \mathbb{E}\|\hat{f}_{k_n} - f\|^2 I(\Omega_n) \geq \frac{10^{2(n+1)}}{2n^2}\frac{nc}{2(n+1)10^{n+1}5^n} \to \infty,$$

as $n \to \infty$. Large error on Ω_n is not compensated by the small probability of Ω_n, and $\hat{f}_{k_n}$ is not L^2-consistent.

It should be noted that, due to exponentially decreasing $\mathrm{P}(\Omega_n)$, this construction only applies to exponentially smoothing operators $\mathcal{K}$.

4.2. Convergence in probability of L^2 loss

The discrepancy principle defined in Section 2 is meaningful, if the noise level is finite, which is guaranteed by the following assumption (cf. formula (7)):

Assumption 1: $\int K^2(x, y)g(y)\, \mathrm{d}\mu_X(x)\, \mathrm{d}\mu_Y(y) < \infty.$

For the results in this section, we also need the filter to be left-continuous. This condition is clearly satisfied, e.g., for k-times iterated Tikhonov regularization, truncated SVD and Landweber iteration.

Assumption 2: For any fixed $\lambda \in (0, \|\mathcal{K}\|^2]$ and for $\alpha > 0$, the function $\alpha \to F_\alpha(\lambda)$ is left-continuous.

One immediate consequence of this assumption is that the left-hand side of the DP equation (9), which can be written as

$$\|\mathcal{K}^*\mathcal{K}\hat{f}_\alpha - \hat{q}_n\|^2 = \sum_i \left(F_\alpha(\sigma_i^2)\sigma_i^2 - 1\right)^2 \langle \hat{q}_n, v_i \rangle^2,$$

is left-continuous in α. As it also equals zero for $\alpha = 0$ (in the limit), a DP sequence of regularization parameters is well defined with formula (10) and it satisfies the inequality $\|\mathcal{K}^*\mathcal{K}\hat{f}_{\alpha_n} - \hat{q}_n\|^2 \leq \tau\hat{\delta}_n^2$. The following theorem applies, e.g., to truncated SVD and Landweber regularization, which are of infinite qualification, and, with $k > 1$, to the k-times iterated Tikhonov regularization, which is of qualification $2k$.

THEOREM 3. *Let $\{F_\alpha, \alpha > 0\}$ be a bounded regularizing filter with qualification $\nu_0 > 2$. Let $(\alpha_n)_{n\in\mathbb{N}}$ be a DP sequence, as defined in (10), and let $\hat{f}_{\alpha_n} = F_{\alpha_n}(\mathcal{K}^*\mathcal{K})\hat{q}_n$ be the estimator of f in the model (1) with an injective, Hilbert-Schmidt operator $\mathcal{K}$. If Assumptions 1 and 2 hold true, then $\hat{f}_{\alpha_n} \to f$ in probability, as $n \to \infty$. If additionally $f \in \mathcal{F}_{\nu,\rho}$ with $\nu + 2 \leq \nu_0$, then $\|\hat{f}_{\alpha_n} - f\|^2 = \mathcal{O}_P(n^{-\nu/(\nu+2)})$.*

In the finitely smoothing case, the obtained rates of convergence are suboptimal, when compared to estimators with a priori regularization in Theorem 1. However, in the limit, as $b \searrow 1/2$ in Theorem 1, i.e., on the boundary of the set of cases handled by both theorems, the two rates coincide. On the other extreme, as $b \to \infty$, the rates in Theorem 1 approach $n^{-\nu/(\nu+1)}$, while those in Theorem 3 remain constant.

It should be noted that the rate $n^{-\nu/(\nu+2)}$ is the same (up to a log factor) as that obtained in [7] for DP in the pre-conditioned model with Gaussian white noise (cf. Remark 10 and Corollary 2, together with Example 3, with somewhat different notation). This suboptimal rate also matches the lower bound obtained in [28] for discretized Gaussian white noise model and truncated SVD, with DP based on smoothed residuals (Proposition 4.2, with $2\beta = 2b\nu$, $p = b > 1/2$ and $\alpha = 1$, which corresponds to our version of DP for a pre-conditioned model). Although the results in [28] do not directly apply to our Poissonian setup, Proposition 4.2 (case $\alpha p < 1/2$) may suggest that adaptation could possibly

be achieved also in our Poissonian setup, with DP based on less smoothed residuals. This may be a subject of further research.

Although stopping criteria based on cross validation, Lepski balancing principle or penalized risk estimation may allow for better adaptation (cf., e.g., [12] for the Poissonian case and the discussion of the idealized Gaussian white noise case in [5]), their computational cost may be prohibitive (cf., e.g., [12]). On the other hand, using DP is usually computationally much cheaper, because not all possible models have to be used in the optimization process, but its adaptation ability is often limited.

5. Simulation results

Finite sample behaviour of the proposed parameter selection rule was verified in the Spektor-Lord-Willis (SLW) stereological problem of unfolding the distribution of the radii of balls, randomly placed in an opaque medium, from an observed linear section through the medium. Early versions of the problem were described in [27] and [22] as models of some measurements in material sciences. The following formulation of the SLW problem as a Poisson inverse problem was taken from [31] and [13]. Assume that the balls' radii distribution is supported in $[0, 1]$ and the balls' centers form a homogeneous Poisson process in $\mathbb{R}^3$ with the expected number of c points per unit volume. With Lebesgue measures μ_X and μ_Y, if the balls' radii distribution has a density ρ, then the observed radii of the line segments (intersections of the balls with the linear probe) form a Poisson process on $[0, 1]$ with intensity function $ng(u)$, where n is the "size of the experiment" related to the total length of the observed linear probe and

$$g(y) = 2y \int_y^1 f(x) d\mu_X(x) =: \left(\mathcal{K}f\right)(y),$$

with $f(x) = c\rho(x)$. The goal is to unfold f from the observed linear sections. With $\mathcal{K}\colon L^2([0, 1], \mu_X) \to L^2([0, 1], \mu_Y)$, $\mathcal{K}$ is a compact Hilbert-Schmidt operator, its inverse is unbounded and the SLW problem is ill-posed in the Hadamard sense. Clearly, $\mathcal{K}$ is of the form (2), with the kernel

$$K(x, y) = 2y I_{\{x \geq y\}}(x, y).$$

In order to work with a simpler form of the SVD of $\mathcal{K}$, the following modified formulation of the SLW problem was used in the simulation study. Instead of considering the operator $\mathcal{K}$ as an operator from $L^2([0, 1], \mu_X)$ to $L^2([0, 1], \mu_Y)$, we change the dominating measures, both in the data space and in the solution space, and consider the folding operator $\mathcal{K}$ as an operator from $L^2([0, 1], \tilde{\mu}_X)$ to $L^2([0, 1], \tilde{\mu}_Y)$ with $d\tilde{\mu}_X(x) = x d\mu_X(x)$ and $d\tilde{\mu}_Y(y) = y d\mu_Y(y)$. The functions g and f are then replaced with $l(y) = g(y)/y$ and $h(x) = f(x)/x$ and the operator becomes

$$l(y) = \left(\mathcal{K}h\right)(y) = 2 \int_y^1 h(x) d\tilde{\mu}_X(x). \tag{17}$$

The SVD of this operator was found in [13].

In this case, the estimators $\hat{\delta}_n$ and $\hat{q}_n$ take the following simple form:

$$\hat{\delta}_n^2 = \frac{2}{n^2} \sum_{i=1}^{N} \left(1 - Y_i^2\right) \quad \text{and} \quad \hat{q}_n(x) = \frac{2}{n} \# \left\{i : Y_i < x\right\},$$

where $\#A$ stands for the cardinality of the set A.

The simulations were conducted in the Python environment, version 3.7 and the code is available in [24]. Selected results for the following four functions (taken from [13, 11, 12]):

- *Beta*$(4, 2)$:
  $f(x) = 20x^3(1 - x)I_{[0,1]}(x)$,
- Swapped Minerbo-Levy A ($SMLA$):
  $f(x) = 4x^2 I_{[0,0.5]}(x) + [2 - 4(1 - x)^2]I_{[0.5,1]}(x)$,
- Swapped Minerbo-Levy B ($SMLB$):
  $f(x) = 1.241(2x - x^2)^{-3/2} \exp\left(1.21[1 - (2x - x^2)^{-1}]\right) I_{[0,1]}(x)$,
- *Bimodal*:
  $f(x) = \frac{28125}{512}x^2(0.8 - x)^2 I_{[0,0.8]}(x) + \frac{9375}{8}(0.6 - x)^2(1 - x)^2 I_{[0.6,1]}(x)$

will be presented. We set $c = 1$, so that all functions are probability densities. For a given density f of the spheres' radii and the experiment size n, artificial data samples were generated with the following algorithm. First, a number $\tilde{n}$ was generated from the Poisson distribution with expectation n, using the method *poisson* from the package *numpy.random*. Then, $\tilde{n}$ pairs (R, D) of independent random variables were generated: the ball radius R from the density f, with the acceptance-rejection method and the distance D from the ball center to the linear probe from the density $d(x) = 2xI_{[0,1]}$, using the method *beta* from the package *numpy.random*. Pairs with $D > R$ (the probe then misses the ball) were dropped, otherwise $Y = \left(R^2 - D^2\right)^{1/2}$ was being taken as the observed value. The mean numbers of the observed Y data points were 47.7%, 54.2%, 41.4%, and 35.4% of n, respectively for the functions $Beta(4, 2)$, $SMLA$, $SMLB$ and *Bimodal*.

For each function and each experiment size, 10 artificial data samples were generated and the estimators $\hat{f}_n$ were constructed as $\hat{f}_n(x) = x\hat{h}_n(x)$, where $\hat{h}_n$ was the estimator of the function h in the problem (17). Three main regularization methods were used:

- Landweber iterative procedure,
- 2-times iterated Tikhonov procedure,
- truncated singular value decomposition (TSVD).

For each estimation procedure, values of τ between 1 and 3 with step 0.1 were tested. The best results were obtained with τ set to 1.1, 1.4 and 1.0 for, respectively, Landweber, TSVD and 2-times iterated Tikhonov procedure and the relaxation parameter (β) for Landweber procedure set to 1.97, and those values were used in all results presented below. Large values of τ, above 1.5 for Landweber and Tikhonov procedure and above 2 for TSVD procedure, led to visible deterioration of the results. Varying the values of τ in the range of

smaller values had only a minor impact on the obtained results. For the relaxation parameter of the Landweber procedure, values 2-, 4- and 8-times smaller than that finally selected were also tested. With all other parameters fixed, the results did not change significantly, but smaller values led to a materially longer computation time. A maximal number of iterations of the Landweber algorithm has been restricted to 1000 as a safety measure against occasional "pathological" data.

For each data sample and for each estimation procedure, the best possible regularization parameter was found, i.e. that minimizing the (numerically computed) L^2 distance between $\hat{f}_n$ and the true f. The best and worst cases (out of 10 data samples) are presented for the first two regularization methods in the left panels of Figures 1–8. The right panels of those figures illustrate the performance of the discrepancy principle versus the oracle, able to produce for each given data the best possible solutions. Estimators based on TSVD usually tend to select too few components and perform somewhat worse.

The $SMLB$ and $Bimodal$ functions are much more difficult to estimate in the SLW problem than "easy" functions $Beta(4, 2)$ and $SMLA$, which can adequately be approximated with only a few low-rank singular functions of the SLW operator. The $SMLB$ function is characterized by rapid local changes and needs essentially larger sample sizes than the "easy" functions. The same is true for $Bimodal$, because of relatively close local maxima separated by a shallow minimum. Clear improvement of the performance of estimators is seen, when n increases from 2 000 to 10 000 and to 1 000 000. The DP-based procedure seems to work reasonably well, the errors of the obtained estimators being close to those of the best possible solutions. The simulation results are also comparable with those presented in [13, 11, 12].

The simulation study was also performed for the standard 1-time iterated Tikhonov procedure, the qualification ν_0 of which equals 2. This procedure does not fulfill the assumptions of Theorem 3 and was included in simulations to check the effect of the violated condition. In almost all cases, the results indicate a substantial increase in estimation errors in comparison to the other procedures, which may suggest that the assumption on the qualification of the filter to be strictly larger than 2 may indeed be necessary in Theorem 3. Some exemplary outcomes are shown in Figure 9 for $\tau = 1$ and sample size 1 000 000.

In all simulations, the performance of various DP-based procedures was assessed from the point of view of a potential user, who wants to apply them to his/her specific and fixed data set. Because of that, errors (rather than risks) are presented in the right-hand panels of Figures 1–8 and compared to the minimal errors that can be achieved with the regularization parameter chosen optimally *for the given sample* by an oracle that knows and uses both the true signal and the sample. This also allows for comparisons with other similar results for Poisson inverse problems, published earlier in [13, 11, 12]. Our approach is in contrast with some simulation results for Gaussian problems, published, e.g., in [6, 5, 28, 33], and aimed more at illustrating theoretical adaptivity properties of the procedures. In that case, it is more natural to compare with oracle risks
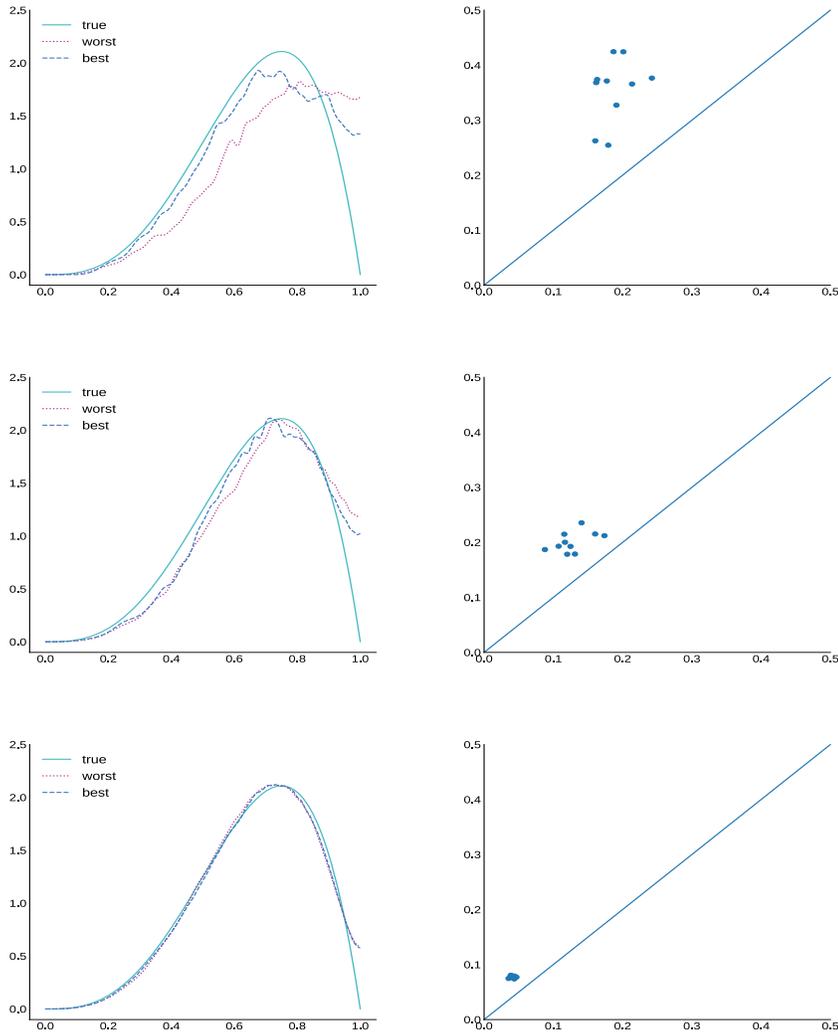
FIG 1. *Worst (dotted) and best (dashed) reconstruction (out of 10 data samples) of Beta(4, 2) intensity function (solid line) based on the Landweber procedure for the experiment size, from top: $n = 2000$, $n = 10^4$, $n = 10^6$. On the right-hand side, the scatter plots of L^2 errors of the DP solutions versus those of the best possible solutions.*

(rather than sample-wise minimal errors) and to show the dependence of the risk on the sample size.

6. Summary and conclusions

In this paper, a novel approach to data-driven choice of regularization parameters in Poisson inverse problems is proposed. This specialized version of the
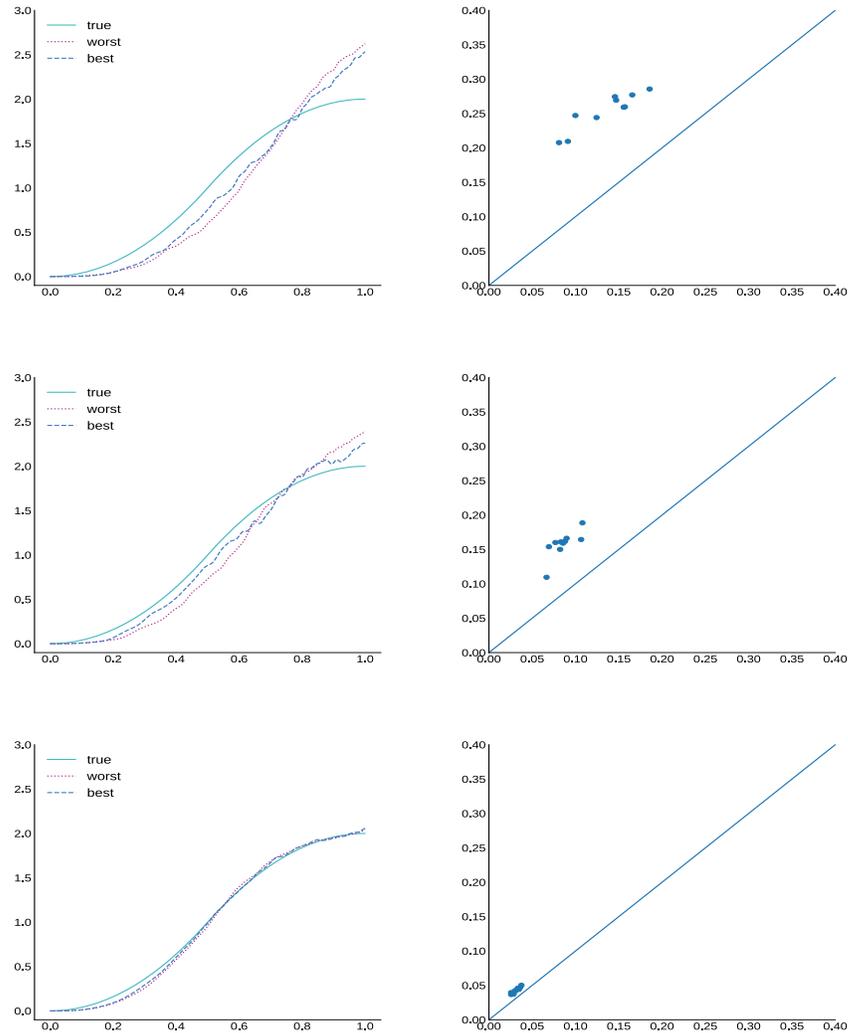
FIG 2. *Same as in Fig. 1, but for the SMLA function.*

Morozov discrepancy principle leads to weakly consistent filter-induced estimators with explicit rates of convergence of the L^2 loss under source conditions on the estimated function.

As the reference case, properties of filter-induced estimators with a priori regularization in Poisson inverse problems are summarized. In the finitely smoothing case, minimax rates of convergence are attained over subsets of smoothness classes defined via source conditions and an additional boundedness condition on the image function. In order to also cover the non-Hilbert-Schmidt case, we provide a direct proof of Theorem 1, even if the case of operators with poly-
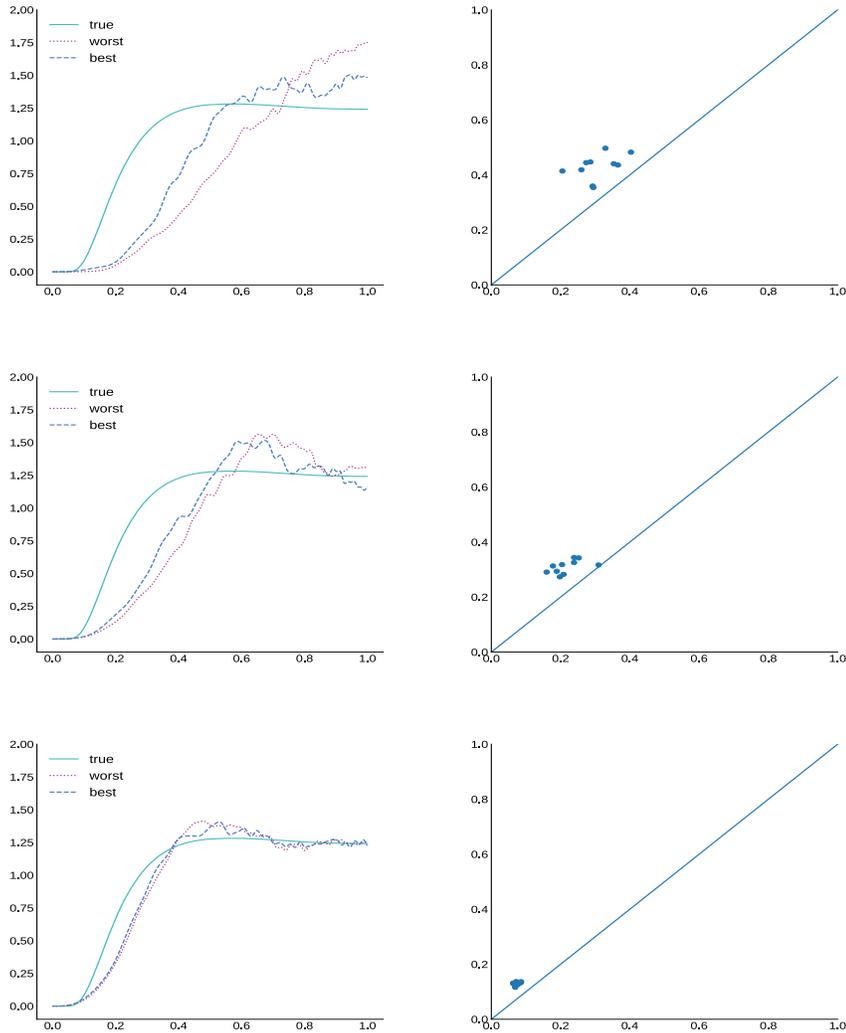
FIG 3. *Same as in Fig. 1, but for the SMLB function.*

nomially decaying singular values with exponent $b > 1/2$ is covered by Section 5.3 in [4]. The case of trace class operators may also be handled similarly to how the white noise model is treated in [7], which not only reproduces the rates in the finitely smoothing case with $b > 1/2$, but also covers in Theorem 2 the infinitely smoothing case with a weaker source condition traditionally studied with such operators.

The Morozov principle is used after pre-conditioning the initial model (1), which allows for construction of an unbiased and $\sqrt{n}$-consistent estimator of the noise level. As it often happens for discrepancy principle in stochastic inverse
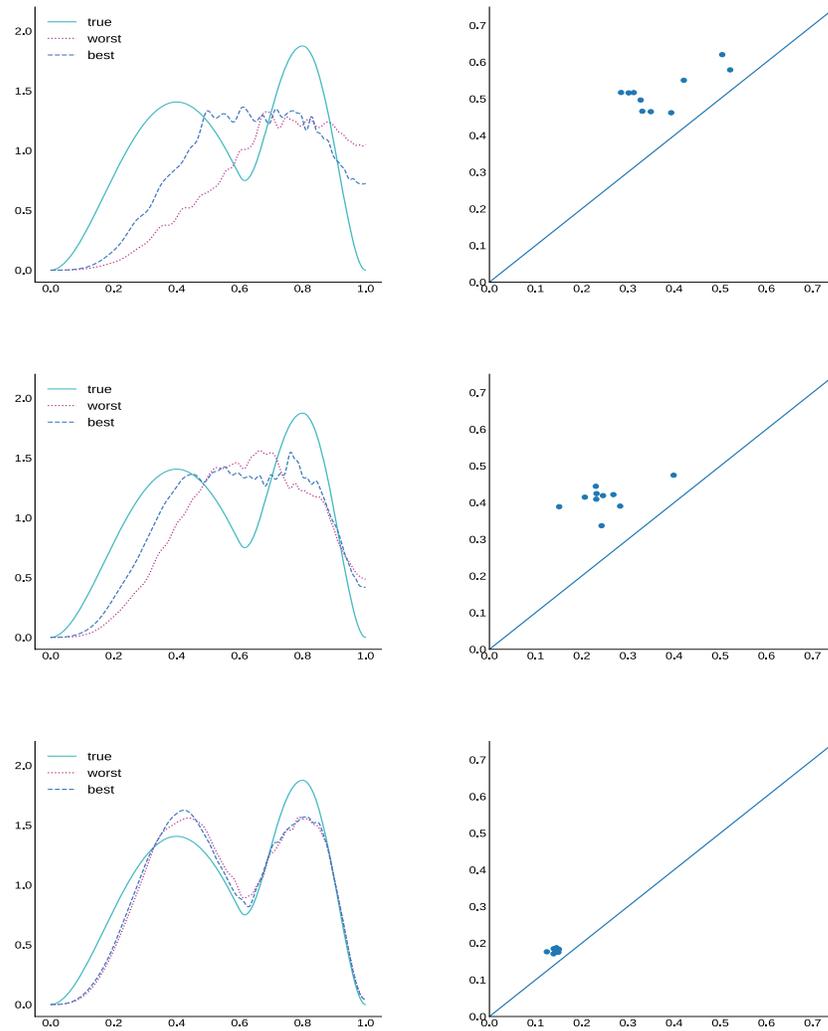
FIG 4. *Same as in Fig. 1, but for the Bimodal function.*

problems, the resulting estimators are not strictly adaptative in the sense defined in Section 1. In the finitely smoothing case, the rate of convergence with a posteriori regularization coincides with the respective rate of convergence with a priori regularization only on the boundary of the cases handled by both methods and they differ in other cases. The results for a posteriori regularization are valid only for the filter-induced estimators with qualification strictly larger than 2. This covers, in particular, truncated SVD, Landweber and k-times iterated Tikhonov regularization with $k > 1$, but excludes the standard (1-time iterated) Tikhonov regularization. Poor behaviour of the estimation procedure in the
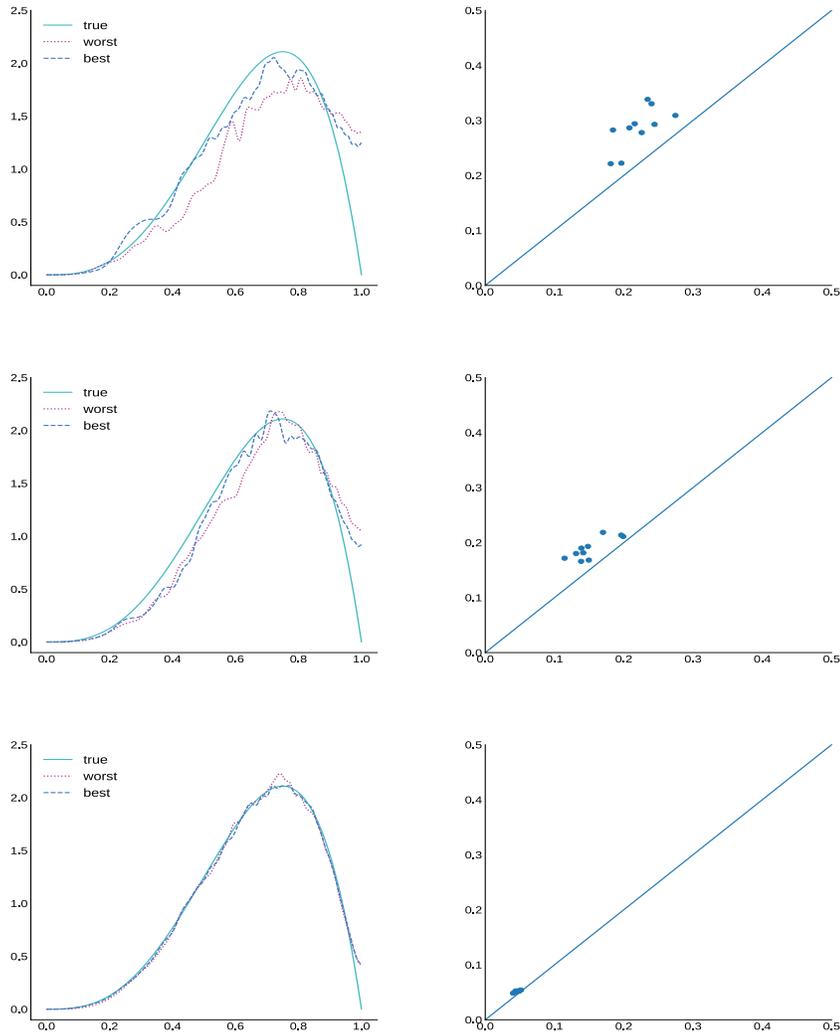
FIG 5. *Worst (dotted) and best (dashed) reconstruction (out of 10 data samples) of Beta(4, 2) intensity function (solid line) based on the 2-times iterated Tikhonov procedure for the experiment size, from top: $n = 2000$, $n = 10^4$, $n = 10^6$. On the right-hand side the scatter plots of L^2 error of the discrepancy principle solutions versus those of the best possible solutions.*

latter case is clearly seen in simulations.

Our theory covers DP-estimators only in problems with Hilbert-Schmidt operators, which excludes, for example, a well-known and important Wicksell's problem of stereology. Whether the presented methodology can be modified to also cover non-Hilbert-Schmidt cases, needs further investigation. It should be noted here that a version of DP applicable to compact but not necessarily Hilbert-Schmidt operators has been studied recently in [16] for discretized in-
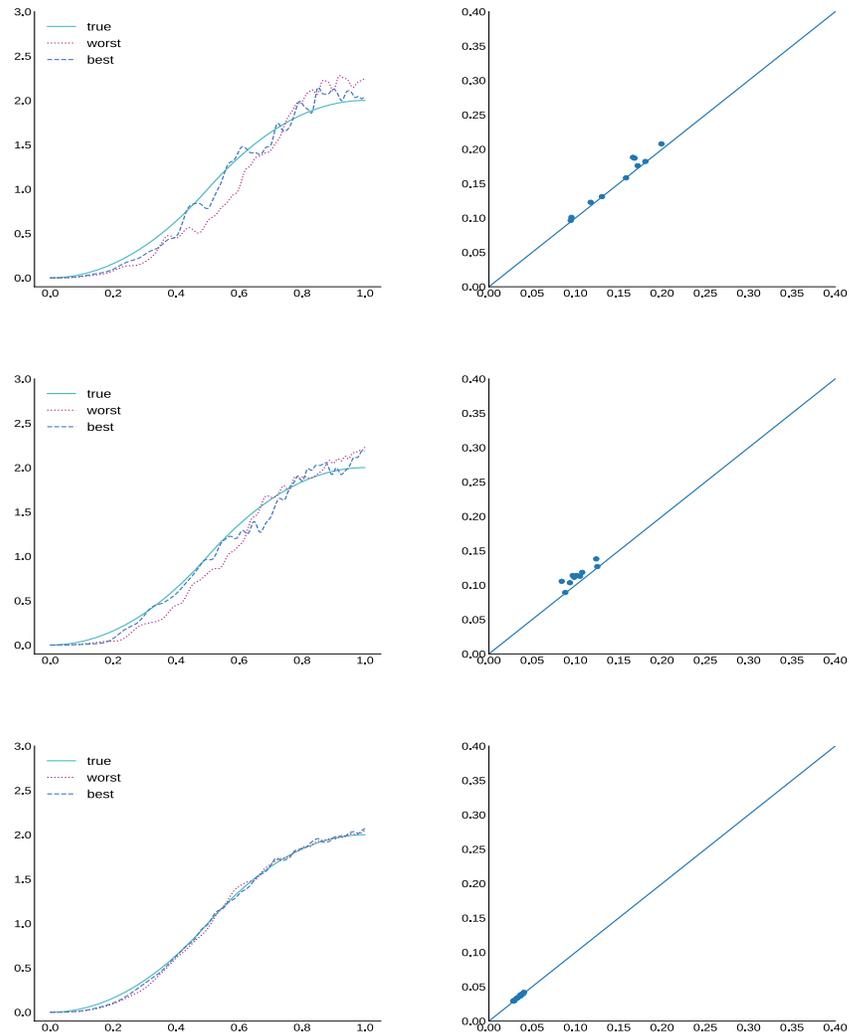
Fig 6. *Same as in Fig. 5, but for the SMLA function.*

verse problems with independently repeated measurenements in each channel and with unknown error distribution. However, as measurements in white noise are assumed in [16], the results do not seem to satisfactorily cover our Poissonian setup (cf. a related discussion in the last paragraph of our Sec. 3).

Another idea worth investigating is using our version of the Morozov principle with a weighted norm, pretty much in the same spirit as in [7]. Introductory simulation experiments suggest that this may significantly improve the quality of estimation, at least in finite samples.
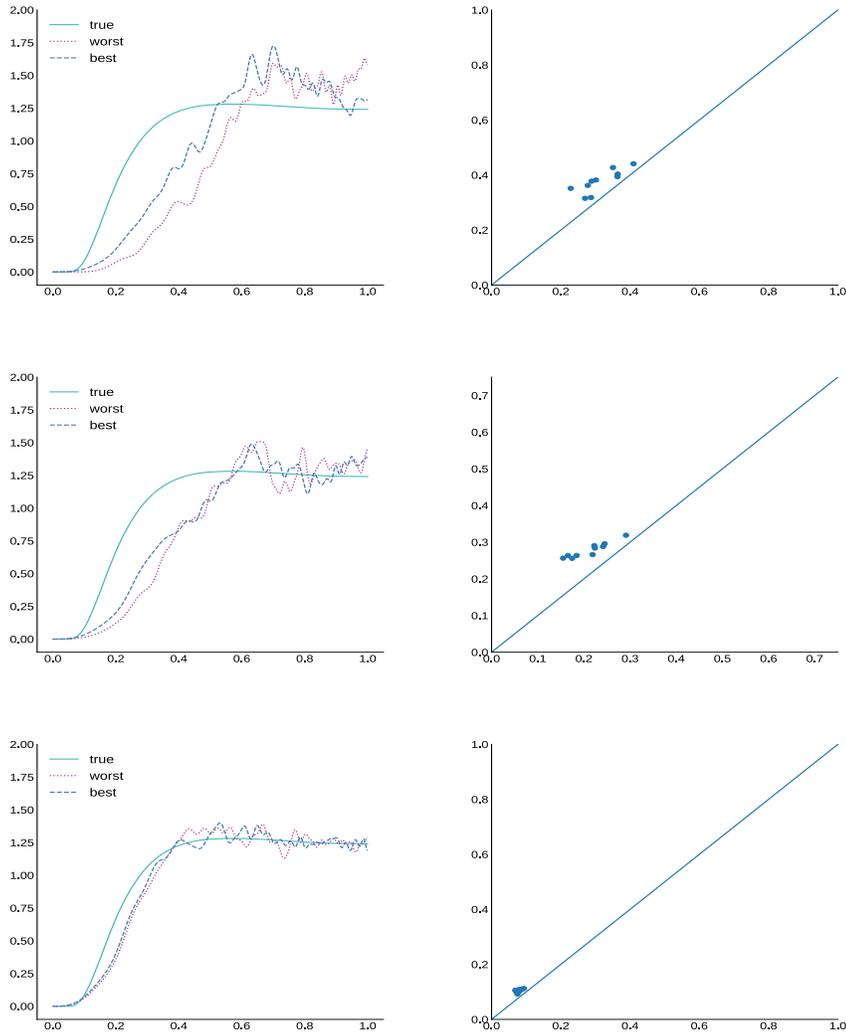
FIG 7. *Same as in Fig. 5, but for the SMLB function.*

7. Proofs

In all proofs, *Id* stands for the identity operator. For brevity, we also define
$S_\alpha := \mathcal{K}^*\mathcal{K}F_\alpha(\mathcal{K}^*\mathcal{K}) - Id$.

The following lemma collects various technical results on regularizing filters. In the proof, we adopt to the current setup standard techniques used, e.g., in many places in [14]. Recall that $q = \mathcal{K}^*\mathcal{K}f$.
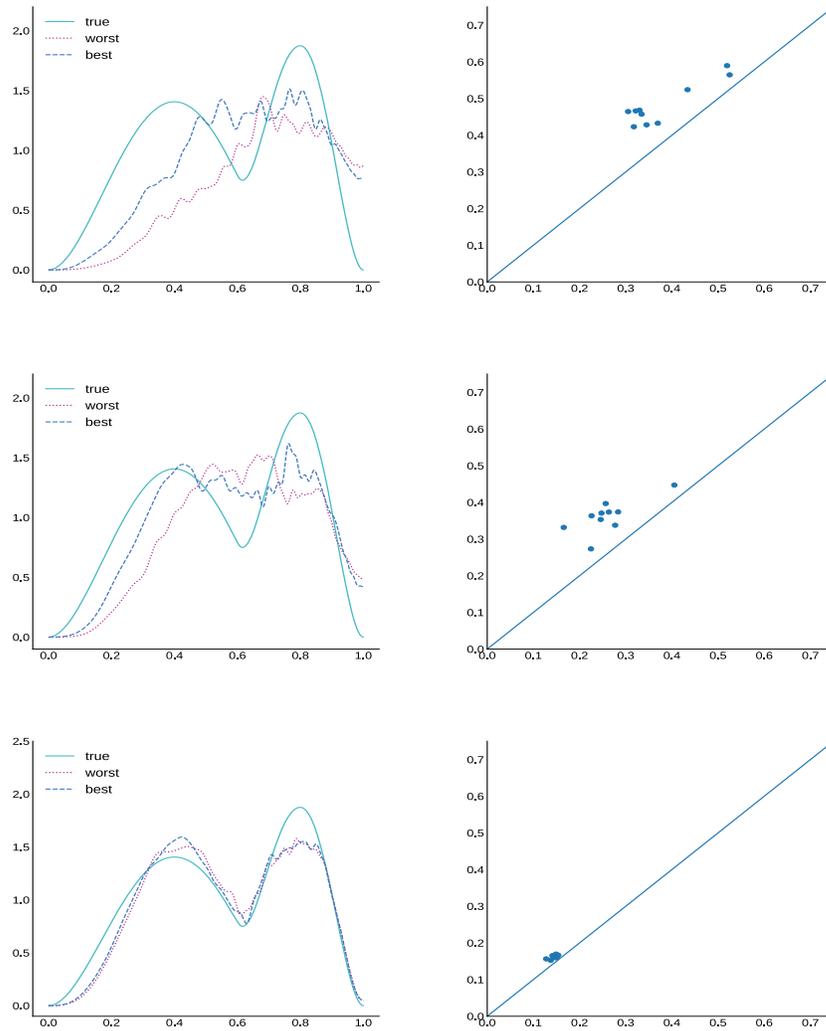
FIG 8. *Same as in Fig. 5, but for the Bimodal function.*

LEMMA 1. *Let $\{F_\alpha, \alpha > 0\}$ be a regularizing filter with qualification ν_0 and let $f \in \mathcal{F}_{\nu,\rho}$ with $\nu \leq \nu_0$. Then, with some constant C,*

1. $\|F_\alpha(\mathcal{K}^*\mathcal{K})q - f\| \leq C_\nu \rho \, \alpha^{\nu/2}$,
2. $\|F_\alpha(\mathcal{K}^*\mathcal{K})q - f\| \leq C\|S_\alpha q\|^{\frac{\nu}{\nu+2}} \rho^{\frac{2}{\nu+2}}$.

If additionally $\nu + 2 \leq \nu_0$, then

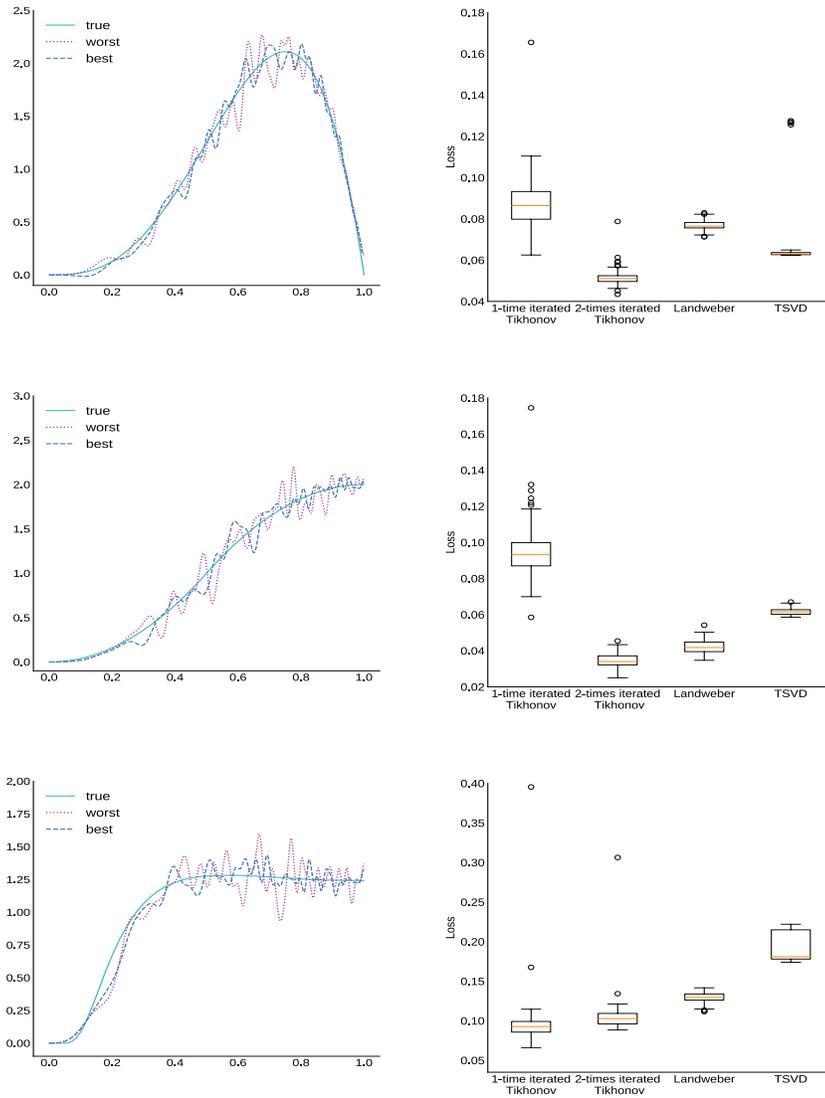3. $\|S_\alpha q\| \leq C_{\nu+2}\rho\alpha^{\frac{\nu+2}{2}}$.

FIG 9. *Worst (dotted) and best (dashed) reconstruction (out of 10 data samples) of Beta*(4, 2) *(upper), SMLA (middle) and SMLB (lower) intensity function (solid line) for sample size* $n = 1\,000\,000$ *obtained with* 1*-time iterated Tikhonov procedure. On the right-hand side the comparison of empirical distributions (based on 100 data samples) of* L^2 *errors for all considered methods (Tukey's style: median, quartiles, minimal and maximal non-outliers and, as individual points, outliers more than* $1.5 * interquartile$ *range away from the box).*

Proof. For the first statement, write

$$\|F_\alpha(\mathcal{K}^*\mathcal{K})q - f\| = \|(F_\alpha(\mathcal{K}^*\mathcal{K})\mathcal{K}^*\mathcal{K} - Id)(\mathcal{K}^*\mathcal{K})^{\nu/2}s\|$$

$$= \left[\sum_i \left(F_\alpha(\sigma_i^2)\sigma_i^2 - 1 \right)^2 \sigma_i^\nu \langle s, v_i \rangle^2 \right]^{1/2}$$

$$\leq \sup_{\lambda \in (0, \sigma_1^2]} \left| (\lambda F_\alpha(\lambda) - 1)\lambda^{\nu/2} \right| \cdot \|s\| \leq C_\nu \alpha^{\nu/2} \rho.$$

For the second statement, set $\xi = \sum_i \left(F_\alpha(\sigma_i^2)\sigma_i^2 - 1 \right) \sigma_i^{-\nu} \langle f, v_i \rangle v_i$. Then,

$$(\mathcal{K}^*\mathcal{K})^{\nu/2}\xi = F_\alpha(\mathcal{K}^*\mathcal{K})q - f.$$

The interpolation inequality ((2.49) in [14]) gives

$$\|F_\alpha(\mathcal{K}^*\mathcal{K})q - f\| = \| (\mathcal{K}^*\mathcal{K})^{\nu/2}\xi \| \leq \| (\mathcal{K}^*\mathcal{K})^{\frac{\nu+2}{2}} \xi \|^{\frac{\nu}{\nu+2}} \|\xi\|^{\frac{2}{\nu+2}}$$

$$= \|\mathcal{K}^*\mathcal{K}F_\alpha(\mathcal{K}^*\mathcal{K})q - q\|^{\frac{\nu}{\nu+2}} \|\xi\|^{\frac{2}{\nu+2}}.$$

Moreover,

$$\|\xi\| = \left(\sum_i \left(F_\alpha(\sigma_i^2)\sigma_i^2 - 1 \right)^2 \sigma_i^{-2\nu} \langle f, v_i \rangle^2 \right)^{1/2} \leq C \left(\sum_i \sigma_i^{-2\nu} \langle f, v_i \rangle^2 \right)^{1/2} \leq C\rho,$$

with some $C > 0$, because $F_\alpha(\sigma_i^2)\sigma_i^2 \leq C_R$ for all i. Combining both inequalities proves the second statement.

For the third statement, write

$$\|S_\alpha q\| = \left(\sum_i \sigma_i^4 \left(F_\alpha(\sigma_i^2)\sigma_i^2 - 1 \right)^2 \langle f, v_i \rangle^2 \right)^{1/2}$$

$$= \left(\sum_i \sigma_i^{2(\nu+2)} \left(F_\alpha(\sigma_i^2)\sigma_i^2 - 1 \right)^2 \sigma_i^{-2\nu} \langle f, v_i \rangle^2 \right)^{1/2}$$

$$\leq \sup_{\lambda \in (0, \|K\|^2]} \lambda^{\frac{\nu+2}{2}} |\lambda F_\alpha(\lambda) - 1| \left(\sum_i \sigma_i^{-2\nu} \langle f, v_i \rangle^2 \right)^{1/2} \leq C_{\nu+2} \alpha^{\frac{\nu+2}{2}} \rho.$$

$\square$

Without regularity assumption on f, we have

LEMMA 2. *Let $\{F_\alpha, \alpha > 0\}$ be a regularizing filter with qualification $\nu_0 > 2$ that satisfies Assumption 2. Then, for any fixed f, there exists a function $g : \mathbb{R}^+ \to \mathbb{R}^+$ such that $g(\alpha) \to +\infty$ as $\alpha \to 0$ and $\lim_{\alpha \to 0} \|S_{\alpha g(\alpha)}q\|\alpha^{-1} = 0$.*

Proof. We proceed as in the proof of Proposition 1 in [15] and first prove that $\lim_{\alpha \to 0} \|S_\alpha q\|\alpha^{-1} = 0$. Let $\epsilon > 0$ and fix L such that $C_2^2 \sum_{i=L+1}^\infty \langle f, v_i \rangle^2 < \epsilon$. Then

$$\|S_\alpha q\|^2 \alpha^{-2} = \alpha^{-2} \sum_i \left(F_\alpha(\sigma_i^2)\sigma_i^2 - 1 \right)^2 \sigma_i^4 \langle f, v_i \rangle^2$$

$$\leq \alpha^{-2} \sum_{i=1}^{L} \left(F_\alpha(\sigma_i^2)\sigma_i^2 - 1 \right)^2 \sigma_i^{2\nu_0} \sigma_i^{4-2\nu_0} \langle f, v_i \rangle^2$$

$$+ \alpha^{-2} \sum_{i=L+1}^{\infty} \left(F_\alpha(\sigma_i^2)\sigma_i^2 - 1 \right)^2 \sigma_i^4 \langle f, v_i \rangle^2$$

$$\leq \alpha^{-2} \sup_{\lambda \in (0, \|\mathcal{K}\|^2]} (\lambda^{\nu_0/2}|F_\alpha(\lambda)\lambda - 1|)^2 \|f\|^2 \sigma_L^{2(2-\nu_0)}$$

$$+ \alpha^{-2} \sup_{\lambda \in (0, \|K\|^2]} (\lambda|F_\alpha(\lambda)\lambda - 1|)^2 \sum_{i=L+1}^{\infty} \langle f, v_i \rangle^2$$

$$\leq C_{\nu_0}^2 \|f\|^2 \sigma_L^{2(2-\nu_0)} \alpha^{\nu_0-2} + C_2^2 \sum_{i=L+1}^{\infty} \langle f, v_i \rangle^2 < 2\epsilon$$

for α sufficiently small, which proves that $\lim_{\alpha \to 0} \|S_\alpha q\| \alpha^{-1} = 0$. Further, define

$$g(\alpha) := \sup \left\{ t > 0 : \|S_{\alpha t} q\|/\alpha \leq t^{-1} \right\}$$

and note that, for any positive t and for α small enough, $\|S_{\alpha t} q\|/\alpha \leq t^{-1}$, because $\|S_{\alpha t} q\|/\alpha \to 0$, as $\alpha \to 0$. This implies that $g(\alpha) \to \infty$, as $\alpha \to 0$ and completes the proof, because $\|S_{\alpha t} q\|$ is left-continuous in t due to Assumption 2, and we thus have $\|S_{\alpha g(\alpha)} q\|/\alpha \leq 1/g(\alpha) \to 0$, as $\alpha \to 0$. $\qquad \square$

7.1. Proof of Theorem 1

The proof goes along standard lines of writing the risk as the sum of variance and squared bias components, and separating low and high frequency terms in the variance component. Since $\mathbb{E} F_\alpha(\mathcal{K}^* \mathcal{K})(\hat{q}_n - q) = 0$, one has

$$\mathbb{E}\|\hat{f}_\alpha - f\|^2 = \mathbb{E}\|F_\alpha(\mathcal{K}^* \mathcal{K})(\hat{q}_n - q)\|^2 + \|F_\alpha(\mathcal{K}^* \mathcal{K})q - f\|^2$$
$$\leq \mathbb{E}\|F_\alpha(\mathcal{K}^* \mathcal{K})(\hat{q}_n - q)\|^2 + C_\nu^2 \rho^2 \alpha^\nu,$$

because of Lemma 1. For the variance term, we obtain

$$\mathbb{E}\|F_\alpha(\mathcal{K}^* \mathcal{K})(\hat{q}_n - q)\|^2 = \sum_{i \geq 1} F_\alpha^2(\sigma_i^2)\mathbb{E} \langle \hat{q}_n - q, v_i \rangle^2 = \sum_{i \geq 1} F_\alpha^2(\sigma_i^2)\mathrm{Var} \langle \hat{q}_n, v_i \rangle$$
$$= n^{-1} \sum_{i \geq 1} \sigma_i^2 F_\alpha^2(\sigma_i^2) \langle u_i^2, g \rangle \leq M n^{-1} \sum_{i \geq 1} \sigma_i^2 F_\alpha^2(\sigma_i^2),$$

because of (13), and because $\|g\|_\infty \leq M$. Consider first the case $b > 1/2$. Separating high ($\sigma_i^2 \leq \alpha$) and low ($\sigma_i^2 > \alpha$) frequency terms, and using the definition of the bounded filter, we obtain

$$\mathbb{E}\|F_\alpha(\mathcal{K}^* \mathcal{K})(\hat{q}_n - q)\|^2 \leq \frac{M}{n} \left[\frac{C_F^2}{\alpha^2} \sum_{\sigma_i^2 \leq \alpha} \sigma_i^2 + C_R^2 \sum_{\sigma_i^2 > \alpha} \frac{1}{\sigma_i^2} \right].$$

The condition $\sigma_i^2 \leq \alpha$ implies that $i \geq L := B\alpha^{-1/(2b)}$, with some constant B, and $\sigma_i^2 > \alpha$ implies that $i < U := D\alpha^{-1/(2b)}$ with some constant D. Consequently,

$$\mathbb{E}\|F_\alpha(\mathcal{K}^*\mathcal{K})(\hat{q}_n - q)\|^2 \leq \frac{C}{n}\left(\alpha^{-2}\int_L^\infty (x+1)^{-2b}\,\mathrm{d}x + \int_0^U x^{2b}\,\mathrm{d}x\right)$$

$$= Cn^{-1}\alpha^{-\frac{2b+1}{2b}},$$

with a constant $C = C(C_R, C_F, \mathcal{K}, M)$. Summing that with the bias term and substituting $\alpha = \alpha(n) \asymp n^{-2b/(2b\nu+2b+1)}$, which balances the orders of magnitude of the two terms, gives the conclusion. For the case $b \leq 1/2$, the high-frequency terms are equal to zero for low-pass filters satisfying $F_\alpha(\lambda) = 0$ for $\lambda \leq \alpha$, the upper bound for the variance term has the same order of magnitude, and the conclusion follows in the same way.

7.2. Proof of Theorem 3

We use notation introduced in Section 2 and adapt the ideas from the proofs of Theorems 3 and 4 in [15] to the Poissonian case. Let $\Pi_g^{(i)}$, $i = 1, \ldots, n$, be independent copies of a Poisson process Π_g with intensity function g. Then $\Pi_{ng} \stackrel{d}{=} \sum_{i=1}^n \Pi_g^{(i)}$, because of the superposition theorem for Poisson processes (Th. 3.3 in [20]). Under Assumption 1, we thus have by the law of large numbers

$$n\hat{\delta}_n^2 = \frac{1}{n}\sum_{j=1}^{N_n} w(Y_j) = \frac{1}{n}\int w(y)\,\mathrm{d}\Pi_{ng} \stackrel{d}{=} \frac{1}{n}\sum_{i=1}^n \int w(y)\,\mathrm{d}\Pi_g^{(i)} \stackrel{P}{\longrightarrow} \mathbb{E}\int w(y)\,\mathrm{d}\Pi_g$$

$$\tag{18}$$

$$= \int K^2(x,y)g(y)\,\mathrm{d}\mu_X(x)\,\mathrm{d}\mu_Y(y) =: \gamma_g^2 < \infty,$$

which implies that $n\hat{\delta}_n^2 \stackrel{P}{\longrightarrow} \gamma_g^2$ because convergence in distribution to a constant is equivalent to convergence in probability to that constant.

For further use, define $\Delta_n^2 := \|\hat{q}_n - q\|^2$ and note that, by the CLT for random elements in Hilbert spaces (Th. 13.25 in [21]),

$$\sqrt{n}\,(\hat{q}_n - q) = \sqrt{n}\left(\frac{1}{n}\int K(\cdot, y)\,\mathrm{d}\Pi_{ng}(y) - \mathbb{E}\hat{q}_n\right)$$

$$\stackrel{d}{=} \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \int K(\cdot, y)\,\mathrm{d}\Pi_g^{(i)} - q\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n (Z_i - \mathbb{E}Z_i)\right) \stackrel{d}{\longrightarrow} W,$$

where Z_i are independent copies of $Z := \int K(\cdot, y)\,\mathrm{d}\Pi_g(y)$ – a random element in $L^2(\mathcal{X}, \mathcal{B}_X, \mu_X)$, because $K(\cdot, \cdot)$ is square integrable as the kernel of a Hilbert-Schmidt operator. Obviously, $\mathbb{E}Z = \mathcal{K}^*g = q$, by the Campbell's theorem. Also,

$$\mathbb{E}\|Z\|^2 = \int \left[\mathrm{Var}Z(x) + \mathbb{E}^2 Z(x)\right]\,\mathrm{d}\mu_X(x) = \gamma_g^2 + \|q\|^2 < \infty$$

and W is a Gaussian random element in $L^2(\mathcal{X}, \mathcal{B}_X, \mu_X)$. This implies that

$$n\Delta_n^2 \xrightarrow{d} \|W\|^2 \qquad (19)$$

by continuous mapping theorem and, by Slutsky lemma,

$$\Delta_n^2/\hat{\delta}_n^2 \xrightarrow{d} \|W\|^2/\gamma_g^2. \qquad (20)$$

We start the proof from the second assertion of Theorem 3. Define

$$\Omega_n = \left\{ |\sqrt{n}\hat{\delta}_n - \gamma_g| < \frac{\gamma_g}{2}, \ \|S_{c_n}(\hat{q}_n - q)\| < \frac{\sqrt{\tau}\gamma_g}{4\sqrt{n}} \right\},$$

with some $c_n \to 0$. Set $\epsilon := \sqrt{\tau}\gamma_g/4$ and observe that, by Chebyshev's inequality,

$$\mathrm{P}\left(\|S_{c_n}(\hat{q}_n - q)\| \geq \frac{\epsilon}{\sqrt{n}} \right) \leq \frac{n\mathbb{E}\|S_{c_n}(\hat{q}_n - q)\|^2}{\epsilon^2}$$

$$= \frac{1}{\epsilon^2} \sum_i \left(F_{c_n}(\sigma_i^2)\sigma_i^2 - 1 \right)^2 \sigma_i^2 \langle u_i^2, g \rangle \to 0,$$

as $n \to \infty$ because, using the expansion $K(x, y) = \sum_i \sigma_i u_i(y) v_i(x)$, one easily obtains $\sum_i \sigma_i^2 \langle u_i^2, g \rangle = \int K^2(x, y)g(y) \, d\mu_X(x) \, d\mu_Y(y) < \infty$ by Assumption 1 and, obviously, $F_{c_n}(\sigma_i^2)\sigma_i^2 \to 1$ for all i. This and (18) imply that $\mathrm{P}(\Omega_n) \to 1$, as $n \to \infty$. Using Lemma 1 and setting $c_n = [\sqrt{\tau}\gamma_g/(4\rho C_{\nu+2}\sqrt{n})]^{\frac{2}{\nu+2}}$, one obtains

$$\|S_{c_n}\hat{q}_n\|I(\Omega_n) \leq \|S_{c_n}q\|I(\Omega_n) + \|S_{c_n}(\hat{q}_n - q)\|I(\Omega_n)$$

$$\leq C_{\nu+2} \, \rho \, c_n^{\frac{\nu+2}{2}} I(\Omega_n) + \frac{\sqrt{\tau}\gamma_g}{4\sqrt{n}} I(\Omega_n) < \sqrt{\tau}\hat{\delta}_n I(\Omega_n),$$

so $\alpha_n I(\Omega_n) \geq c_n I(\Omega_n) \geq \left(\frac{\sqrt{\tau}\hat{\delta}_n}{6\rho C_{\nu+2}} \right)^{\frac{2}{\nu+2}} I(\Omega_n)$. Further,

$$\|F_{\alpha_n}(\mathcal{K}^*\mathcal{K})\hat{q}_n - f\| \leq \|F_{\alpha_n}(\mathcal{K}^*\mathcal{K})q - f\| + \|F_{\alpha_n}(\mathcal{K}^*\mathcal{K})(\hat{q}_n - q)\|$$

and, by Lemma 1 and equation (10) and because $\|S_{\alpha_n}\| \leq 1 + C_R$,

$$\|F_{\alpha_n}(\mathcal{K}^*\mathcal{K})q - f\| \leq C\|S_{\alpha_n}q\|^{\frac{\nu}{\nu+2}}\rho^{\frac{2}{\nu+2}} \leq C\rho^{\frac{2}{\nu+2}} (\|S_{\alpha_n}\hat{q}_n\| + \|S_{\alpha_n}(\hat{q}_n - q)\|)^{\frac{\nu}{\nu+2}}$$

$$\leq C\rho^{\frac{2}{\nu+2}} \left(\sqrt{\tau}\hat{\delta}_n + (1 + C_R)\Delta_n \right)^{\frac{\nu}{\nu+2}}.$$

Moreover, on the set Ω_n,

$$\|F_{\alpha_n}(\mathcal{K}^*\mathcal{K})(\hat{q}_n - q)\| \leq \frac{C_F}{\alpha_n}\Delta_n \leq C_F \left(\frac{6\rho C_{\nu+2}}{\sqrt{\tau}} \right)^{\frac{2}{\nu+2}} \hat{\delta}_n^{-\frac{2}{\nu+2}} \Delta_n$$

and, combining everything,

$$\|F_{\alpha_n}(\mathcal{K}^*\mathcal{K})\hat{q}_n - f\|$$

$$\leq C\rho^{\frac{2}{\nu+2}}\left(\sqrt{\tau}\hat{\delta}_n + (1+C_R)\Delta_n\right)^{\frac{\nu}{\nu+2}} + C_F\left(\frac{6\rho C_{\nu+2}}{\sqrt{\tau}}\right)^{\frac{2}{\nu+2}}\hat{\delta}_n^{-\frac{2}{\nu+2}}\Delta_n$$

$$\leq D\max\left\{\hat{\delta}_n^{\frac{\nu}{\nu+2}}, \Delta_n^{\frac{\nu}{\nu+2}}\left(\frac{\Delta_n}{\hat{\delta}_n}\right)^{\frac{2}{\nu+2}}\right\},$$

with some constant D, which completes the proof of the second assertion of Theorem 3, because of (18), (19) and (20), and because $\mathrm{P}(\Omega_n) \to 1$.

It is elementary to see that if f has a finite expansion in terms of the singular elements v_i, then it satisfies the source condition with any positive ν, hence $\|\hat{f}_{\alpha_n} - f\|^2 = \mathcal{O}_P(n^{-\nu/(\nu+2)})$ and, consequently, $\hat{f}_{\alpha_n}$ is consistent. To complete the proof of the consistency part of Theorem 3, it is thus sufficient to consider f with infinite expansions. Since $\langle q, v_i\rangle = \sigma_i^2\langle f, v_i\rangle$, q then also has an infinite expansion. Let us fix $\epsilon' > 0$ and select L such that $\langle q, v_L\rangle \neq 0$ and $(\sigma_L^2 F_\alpha(\sigma_L^2) - 1)^2 > 1/2$ for all $\alpha \geq \epsilon'$. This is possible because, for such α, $|F_\alpha(\sigma_L^2)| \leq C_F/\alpha \leq C_F/\epsilon'$ and $\sigma_i \to 0$, as $i \to \infty$. Define

$$\Omega_n^1 = \left\{\gamma_g/2 \leq \sqrt{n}\,\hat{\delta}_n \leq 2\gamma_g, \ \langle \hat{q}_n, v_L\rangle^2 \geq \langle q, v_L\rangle^2/2\right\}$$

and note that $\mathrm{P}(\Omega_n^1) \to 1$ because of (18) and because, by the law of large numbers, $\langle \hat{q}_n, v_L\rangle \xrightarrow{P} \langle q, v_L\rangle$. Then, for $n \geq 64\tau^2\gamma_g^2/\langle q, v_L\rangle^2$ and for all $\alpha \geq \epsilon'$, one has

$$\|S_\alpha\hat{q}_n\|I(\Omega_n^1) \geq \left[(\sigma_L^2 F_\alpha(\sigma_L^2) - 1)^2\langle \hat{q}_n, v_L\rangle^2\right]^{1/2}I(\Omega_n^1)$$

$$\geq \sqrt{\frac{\langle q, v_L\rangle^2}{4}} \geq \frac{4\tau\gamma_g}{\sqrt{n}} \geq 2\tau\hat{\delta}_n I(\Omega_n^1).$$

This means that on Ω_n^1 we have $\alpha_n \leq \epsilon'$, because of (10). Consequently, $\alpha_n \xrightarrow{P} 0$.

Consider the set $\Omega_n \cap \Omega_n^1$ with $c_n = \min\{n^{-1/2}g(n^{-1/2}), n^{-1/4}\}$, where $g(u)$ is defined in Lemma 2, note that $c_n \to 0$ and $c_n\sqrt{n} \to \infty$. Recall that $\mathrm{P}(\Omega_n) \to 1$ and $\mathrm{P}(\Omega_n^1) \to 1$, so $\mathrm{P}(\Omega_n \cap \Omega_n^1) \to 1$. On the set $\Omega_n \cap \Omega_n^1$, one has for large n,

$$\|S_{c_n}\hat{q}_n\| \leq \|S_{c_n}q\| + \|S_{c_n}(\hat{q}_n - q)\| \leq \frac{\sqrt{\tau}\gamma_g}{4\sqrt{n}} + \frac{\sqrt{\tau}\gamma_g}{4\sqrt{n}} \leq \sqrt{\tau}\hat{\delta}_n,$$

utilizing Lemma 2, and this means that, on the set $\Omega_n \cap \Omega_n^1$, one has $\alpha_n \geq c_n$. Having control on the size of the DP sequence, one can write

$$\|\hat{f}_{\alpha_n} - f\| \leq \|(F_{\alpha_n}(\mathcal{K}^*\mathcal{K}) - (\mathcal{K}^*\mathcal{K})^{-1})q\| + \|F_{\alpha_n}(\mathcal{K}^*\mathcal{K})(\hat{q}_n - q)\|.$$

Both terms tend to zero in probability: the first one because $\alpha_n \xrightarrow{P} 0$, and the second one because it can, on $\Omega_n \cap \Omega_n^1$, be bounded as follows

$$\|F_{\alpha_n}(\mathcal{K}^*\mathcal{K})(\hat{q}_n - q)\| \leq \frac{C_F}{\alpha_n\sqrt{n}}\|\sqrt{n}(\hat{q}_n - q)\| \leq \frac{C_F}{\alpha_n\sqrt{n}}\mathcal{O}_P(1).$$

This completes the proof of Theorem 3.

Acknowledgments

References

[1] ANTONIADIS, A. and BIGOT, J. (2006). Poisson inverse problems. *The Annals of Statistics* **34** 2132-2158. MR2291495

[2] BARDSLEY, J. M. and GOLDES, J. (2009). Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation. *Inverse Problems* **25** 095005. MR2540885

[3] BERTERO, M., BOCCACCI, P., TALENTI, G., ZANELLA, R. and ZANNI, L. (2010). A discrepancy principle for Poisson data. *Inverse Problems* **26** 105004. MR2679468

[4] BISSANTZ, N., HOHAGE, T., MUNK., A. and RUYMGAART, F. (2007). Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM Journal on Numerical Analysis* **45** 2610-2636. MR2361904

[5] BLANCHARD, G., HOFFMANN, M. and REISS, M. (2018). Optimal adaptation for early stopping in statistical inverse problems. *SIAM/ASA Journal on Uncertainty Quantification* **6** 1043-1075. MR3829522

[6] BLANCHARD, G., HOFFMANN, M. and REISS, M. (2018). Early stopping for statistical inverse problems via truncated SVD estimation. *Electronic Journal of Statistics* **12** 3204-3231.

[7] BLANCHARD, G. and MATHÉ, P. (2012). Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. *Inverse Problems* **28** 115011. MR2992966

[8] CAVALIER, L. and KOO, J. Y. (2002). Poisson intensity estimation for tomographic data using a wavelet shrinkage approach. *IEEE Transactions on Information Theory* **48** 2794-2802. MR1930348

[9] CELISSE, A. and WAHL, M. (2020). Analyzing the discrepancy principle for kernelized spectral filter learning algorithms. arXiv: 2004.08436 [math.ST].

[10] ĆMIEL, B. (2010). An adaptive wavelet shrinkage approach to the Spektor-Lord-Willis problem. *Journal of Multivariate Analysis* **101** 1458-1470. MR2609506

[11] ĆMIEL, B. (2012). Poisson intensity estimation for the Spektor-Lord-Willis problem using a wavelet shrinkage approach. *Journal of Multivariate Analysis* **112** 194-206. MR2957296

[12] ĆMIEL, B., SZKUTNIK, Z. and WOJDYŁA, J. (2018). Asymptotic confidence bands in the Spektor-Lord-Willis problem via kernel estimation of intensity derivative. *Electronic Journal of Statistics* **12** 194-223. MR3760886

[13] DUDEK, A. and SZKUTNIK, Z. (2008). Minimax unfolding of the spheres size distribution from linear sections. *Statistica Sinica* **18** 1063–1080. MR2440404

[14] ENGL, H. W., HANKE, M. and NEUBAUER, M. (1996). *Regularization of Inverse Problems*. Kluwer, Dordrecht. MR1408680

[15] HARRACH, B., JAHN, T. and POTTHAST, R. (2020). Beyond the Bakushinskii veto: Regularising linear inverse problems without knowing the noise distribution. *Numerische Mathematik* **145** 581-603. MR4119042

[16] HARRACH, B., JAHN, T. and POTTHAST, R. (2020). Regularizing linear inverse problems under unknown non-Gaussian white noise. `https://arxiv.org/pdf/2010.04519.pdf`.

[17] HOHAGE, T. and WERNER, F. (2016). Inverse problems with Poisson data: statistical regularization theory, applications and algorithms. *Inverse Problems* **32** 093001. MR3543331

[18] KINGMAN, J. F. C. (1993). *Poisson Processes*. Oxford University Press, Oxford.

[19] KROLL, M. (2019). Nonparametric intensity estimation from noisy observations of a Poisson process under unknown error distribution. *Metrika* **82** 961-990. MR4018433

[20] LAST, G. and PENROSE, M. (2017). *Lectures on the Poisson Process*. Cambridge University Press. MR3791470

[21] LEDOUX, M. and TALAGRAND, M. (2013). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, New York.

[22] LORD, G. W. and WILLIS, T. F. (1951). Calculation of air bubble distribution from results of a Rosiwal traverse of aerated concrete. *A.S.T.M. Bulletin* **56** 177-187.

[23] MAIR, B. A. and RUYMGAART, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM Journal of Applied Mathematics* **56** 1424-1444. MR1409127

[24] MIKA, G. (2020). Towards adaptivity via a new discrepancy principle for Poisson inverse problems. Simulation code. `https://github.com/GrzegorzMikaAGH/Towards-adaptivity-via-a-new-discrepancy-principle-for-Poisson-inverse-problems`. doi: 10.5281/zenodo.3780604

[25] REISS, R. D. (1993). *A Course on Point Processes*. Springer, New York. 10.1007/978-1-4613-9308-5 MR1199815

[26] SIXOU, B., HOHWEILLER, T. and DUCROS, N. (2018). Morozov principle for Kullback-Leibler residual term and Poisson noise. *Inverse Problems and Imaging* **12** 607-634. MR3810172

[27] SPEKTOR, A. G. (1950). Analysis of distribution of spherical particles in non-transparent structures. *Zavodskaja Laboratorija* **16** 173–177.

[28] STANKEWITZ, B. (2020). Smoothed residual stopping for statistical inverse problems via truncated SVD estimation. *Electronic Journal of Statistics* **14** 3396-3428.

[29] SZKUTNIK, Z. (2000). Unfolding intensity function of a Poisson process in models with approximately specified folding operator. *Metrika* **52** 1–26. MR1791921

[30] Szkutnik, Z. (2005). B-splines and discretization in an inverse problem for Poisson processes. *Journal of Multivariate Analysis* **93** 198–221. MR2119771

[31] Szkutnik, Z. (2007). Unfolding spheres size distribution from linear sections with B-splines and EMDS algorithm. *Opuscula Mathematica* **27** 151–165. MR2320844

[32] Szkutnik, Z. (2010). A note on minimax rates of convergence in the Spektor-Lord-Willis problem. *Opuscula Mathematica* **30** 203–207. MR2669132

[33] Werner, F. (2018). Adaptivity and oracle inequalities in linear statistical inverse problems: a (numerical) survey. In *New trends in parameter identification for mathematical models* (B. Hofmann, A. Leitão and J. Zubelli, eds.) 291-316. Birkhäuser, Cham.

[34] Zanella, R., Boccacci, P., Zanni, L. and Bertero, M. (2009). Efficient gradient projection methods for edge-preserving removal of Poisson noise. *Inverse Problems* **25** 045010. MR2482161