ALEA, Lat. Am. J. Probab. Math. Stat. **21**, 1101–1122 (2024)
DOI: 10.30757/ALEA.v21-43

Weighted discrepancy principle and optimal adaptivity in Poisson inverse problems

Zbigniew Szkutnik

AGH University of Krakow, Faculty of Applied Mathematics, al. A. Mickiewicza 30, 30-059 Krakow, Poland
E-mail address: `szkutnik@agh.edu.pl`

Abstract. A new weighted form of the Morozov discrepancy principle is proposed for Poisson inverse problems with compact but not necessarily Hilbert-Schmidt operators. For a broad class of spectral filters it is shown that the resulting estimators are not only consistent (in probability) but also essentially rate-minimax under source conditions on the estimated signal. This is in contrast with recently published results for discrepancy principle in Poisson inverse problems pre-conditioned in a standard way, where not only the rates were sub-optimal but also the operator was assumed Hilbert-Schmidt. The results are illustrated with a numerical study of two stereological problems: the Wicksell's problem (with a non-Hilbert-Schmidt Abel operator) and the Spektor, Lord and Willis problem (with the Hilbert-Schmidt integration operator). Theoretical results are also discussed in a broader context of white noise inverse problems.

1. Introduction

Let $\mathcal{K} : H_1 \to H_2$ be a bounded, linear operator between some Hilbert spaces H_1, H_2. The respective inner products and norms will be denoted with $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$. The general stochastic inverse problem consists in estimating $f \in H_1$, given some stochastic data with distribution depending on $g = \mathcal{K} f \in H_2$. In this article, $\mathcal{K}$ is assumed injective (for simplicity of exposition) and compact, which makes the problem ill-posed, if $\dim H_1 = \infty$, because $\mathcal{K}^{-1}$ is then unbounded, and some sort of regularization becomes indispensable. Let $\hat{f}_\alpha$ be an estimator of f obtained with some regularization parameter α. Optimal values of α depend on the, generally a priori unknown, regularity of the estimated object f. Therefore, data-driven procedures of choosing a value of the regularization parameter, preferably leading to estimators optimally adaptive over smoothness classes, are of central interest.

Any compact operator $\mathcal{K}$ can be represented via its singular value decomposition SVD: there exist a sequence $\sigma_1 \geq \sigma_2 \geq \ldots > 0$ of singular values, finite or converging to zero, and corresponding sequences $(v_i)_i$ and $(u_i)_i$ of orthonormal right and left singular elements 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 $\mathcal{K}^* u_i = \sigma_i v_i$, where $\mathcal{K}^*$ denotes the adjoint

Received by the editors January 10th, 2024; accepted May 6th, 2024.
2010 *Mathematics Subject Classification.* 62G05, 93E10, 65J20, 45Q05.
Key words and phrases. Adaptive estimation, discrepancy principle, Poisson inverse problems, regularization.

of $\mathcal{K}$. The "signal" f can then be identified with the sequence of its Fourier coefficients

$$\mu_i := \langle f, v_i \rangle.$$

In the best studied special case, g is observed as $\tilde{g}$ in Gaussian white noise W in H_2, i.e.

$$\tilde{g} = \mathcal{K}f + \delta W, \tag{1.1}$$

with some noise level δ, which does not depend on f and tends to zero in the asymptotic regime, see e.g. Bissantz et al. (2007); Blanchard et al. (2018b) and the references therein. W and $\tilde{g}$ are understood here as Hilbert-space processes (i.e. bounded, linear operators from H_2 to the space $L^2(\Omega, \mathcal{F}, \mathrm{P})$ of square-integrable random variables) and $\langle W, \phi \rangle$ and $\langle \tilde{g}, \phi \rangle$ denote their respective actions on $\phi \in H_2$. With $X_i := \langle \tilde{g}, u_i \rangle$, model (1.1) is equivalent to the Gaussian sequence model

$$X_i = \sigma_i \mu_i + \delta \zeta_i, \quad i = 1, 2, \dots,$$

where $(\zeta_i)_i$ is a sequence of independent standard Gaussian variables.

In another important special case, referred to as Poisson inverse problem, the observed object is a Poisson process, say Π_{ng}, on some measure space $(\mathcal{Y}, \mathcal{B}_Y, \mu_Y)$, with an intensity function ng w.r.t. μ_Y. Here, we take $H_2 = L^2(\mathcal{Y}, \mathcal{B}_Y, \mu_Y)$. The parameter n is assumed known. It quantifies the "size of experiment" and tends to infinity in the asymptotic regime, see Szkutnik (2000, 2005); Antoniadis and Bigot (2006); Hohage and Werner (2016); Mika and Szkutnik (2021) and the references therein. The process Π_{ng} can be represented as so-called mixed empirical process (cf. Reiss, 1993, p. 17)

$$\Pi_{ng} = \sum_{j=1}^{N_n} \delta_{Y_j}, \tag{1.2}$$

where N_n is Poisson distributed with mean $n \int g \, \mathrm{d}\mu_Y$, δ_a is the Dirac measure concentrated at a, and $Y_j, j = 1, 2, \dots$ form an i.i.d. sample from a density proportional to g, independent of N_n. As argued in Mika and Szkutnik (2021), the Poisson observation model can also be written in a form similar to (1.1)

$$\tilde{g} = \mathcal{K}f + n^{-1/2}\varepsilon, \tag{1.3}$$

where $\tilde{g} := n^{-1}\Pi_{ng}$ is the Hilbert-space process defined with

$$H_2 \ni \phi \to \langle \tilde{g}, \phi \rangle := \frac{1}{n} \sum_{j=1}^{N_n} \phi(Y_j) = \frac{1}{n} \int \phi \, \mathrm{d}\Pi_{ng},$$

and $\varepsilon := \sqrt{n}(\tilde{g} - g)$ is the scaled Wiener-Ito integral (cf. Last and Penrose, 2018, Sec. 12.1), also understood as a Hilbert-space process with the action

$$H_2 \ni \phi \to \langle \varepsilon, \phi \rangle := \sqrt{n}\Big(\langle \tilde{g}, \phi \rangle - \langle g, \phi \rangle\Big)$$

and the "multiplication by g" covariance operator (cf. Mika and Szkutnik, 2021). With $\varphi = u_i$, define

$$Z_i := \langle \tilde{g}, u_i \rangle = \frac{1}{n} \sum_{j=1}^{N_n} u_i(Y_j)$$

and $\varepsilon_i := \langle \epsilon, u_i \rangle$. Model (1.3) can then be written as a sequence model

$$Z_i = \sigma_i \mu_i + n^{-1/2}\varepsilon_i, \quad i = 1, 2, \dots \tag{1.4}$$

and it follows from the properties of the Wiener-Ito integral (cf. Last and Penrose, 2018, Sec. 12.1) that $E(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \langle g, u_i^2 \rangle$ and $\mathrm{Cov}(\varepsilon_i, \varepsilon_k) = \langle g, u_i u_k \rangle$, for $i, k = 1, 2, \dots$. Hence, contrary to the white noise case, the covariance structure of the noise depends on the signal.

Discrepancy principle (DP) is a popular data-driven method of choosing the value of α in such a way that the discrepancy $\|\mathcal{K}\hat{f}_\alpha - \tilde{g}\|$ matches a "natural" level related to the level of noise. It is well studied in the theory of non-stochastic inverse problems, where it is known to produce, in many cases, rate optimal solutions, see, e.g. Engl et al. (1996). In the white noise model (1.1), since W and $\tilde{g}$ can only be understood as Hilbert-space processes, one often works with the symmetrized model. If $\mathcal{K}$ is a Hilbert-Schmidt operator, then $\mathcal{K}^*\tilde{g} : H_1 \to L^2(\Omega, \mathcal{F}, \mathrm{P})$ and $(\mathcal{K}\mathcal{K}^*)^{1/2}\tilde{g} : H_2 \to L^2(\Omega, \mathcal{F}, \mathrm{P})$ are Hilbert-Schmidt as well. Hence, they are induced by (and may be identified with) separably-valued random elements in, respectively, H_1 and H_2 (cf. Vakhania et al., 1987, Prop. 2.5). DP can then be based on the well defined quantity $\|\mathcal{K}^*(\mathcal{K}\hat{f}_\alpha - \tilde{g})\| = \|(\mathcal{K}\mathcal{K}^*)^{1/2}(\mathcal{K}\hat{f}_\alpha - \tilde{g})\|$, which means that one works with smoothed residuals. It turns out, however, that the resulting estimates are, in some typical setups, rate suboptimal, see Blanchard and Mathé (2012), essentially because the residuals are smoothed too much. To cope with this problem, more flexible residuals smoothing strategies have been proposed. In Stankewitz (2020), for DP based on $\|(\mathcal{K}\mathcal{K}^*)^{\gamma/2}(\mathcal{K}\hat{f}_\alpha - \tilde{g})\|$ with some $\gamma \geq 0$, it was proved for truncated SVD estimators that, in a somewhat different, discretized setup proposed earlier in Blanchard et al. (2018a,b), "moderate smoothing of the residuals can be used to adapt over classes of signals with varying smoothness, while oversmoothing yields suboptimal convergence rates". In Jahn (2021), somewhat similar weighted DP was used to select the number of components for truncated SVD estimators in a setup with multiple unbiased measurements of g and unknown error distribution. It was proved that the resulting estimators are optimally adaptive under standard source conditions on f. As written in Jahn (2021), "The general idea is to rescale the operator $\mathcal{K}$ with a weighting operator S such that the smoothness of f relative to the rescaled operator $S\mathcal{K}$ is better than the original one relative to $\mathcal{K}$". In Blanchard and Mathé (2012), another somewhat similar weighted discrepancy, although with a differently weighted norm $\|(\lambda I + \mathcal{K}^*\mathcal{K})^{-1/2}(\mathcal{K}^*(\mathcal{K}\hat{f}_\alpha - \tilde{g}))\|$, with some $\lambda \geq 0$ was used to cope with sub-optimal rates in the Gaussian white noise model. $\lambda = 0$ corresponds here to $\gamma = 0$ (no smoothing of residuals) and $\lambda \to \infty$ corresponds to $\gamma = 1$ (standard symmetrization with $\mathcal{K}^*$). Finally, note that in Blanchard et al. (2018a,b) and Jahn (2022) modifications of DP based on discretization rather than on smoothing residuals are investigated in the white noise scenario with resulting optimally adaptive estimators.

A special version of DP, suitable for Poisson inverse problems, was recently proposed in Mika and Szkutnik (2021). As in the white noise case, it proved convenient to work with the symmetrized model. The value of α was selected there as an approximate solution of the following DP equation

$$\|\mathcal{K}^*\mathcal{K}\hat{f}_\alpha - \hat{q}\|^2 = \tau\hat{\delta}^2, \tag{1.5}$$

with some fixed, positive τ and with unbiased estimators $\hat{q} = \mathcal{K}^*\tilde{g}$ and $\hat{\delta}^2$ of, respectively, $q := \mathcal{K}^*g$ and the "noise level" $\delta^2 := E(\|\hat{q} - q\|^2)$. Although spectral filter estimators combined with this version of DP adapt to the unknown smoothness, defined via source conditions or Sobolev-type ellipsoids, the resulting convergence rates are, as in Blanchard and Mathé (2012), suboptimal.

The goal of the present study is to investigate whether intermediate smoothing of residuals can also help to obtain optimally adaptive estimators in the Poisson inverse problem setup. Following Mika and Szkutnik (2021), it is natural to express the weighted discrepancy in terms of the symmetrized model. It follows easily from the spectral calculus that, with any $\bar{g} \in H_2$ and $\bar{q} := \mathcal{K}^*\bar{g}$,

$$\|(\mathcal{K}\mathcal{K}^*)^{\gamma/2}(\mathcal{K}f - \bar{g})\| = \|(\mathcal{K}^*\mathcal{K})^{(\gamma-1)/2}(\mathcal{K}^*\mathcal{K}f - \bar{q})\|,$$

so that it is natural to replace equation (1.5) with

$$\|(\mathcal{K}^*\mathcal{K})^{(\gamma-1)/2}(\mathcal{K}^*\mathcal{K}\hat{f}_\alpha - \hat{q})\|^2 = \tau\hat{\delta}_\gamma^2,$$

with some unbiased estimator $\hat{\delta}_\gamma^2$ of the "weighted noise level"

$$\delta_\gamma^2 = E\Big(\|(\mathcal{K}^*\mathcal{K})^{(\gamma-1)/2}(\hat{q} - q)\|^2\Big).$$

Theoretical properties of spectral filter estimators combined with this version of DP are studied in Section 2, with proofs deferred to Section 5. It is shown, in particular, that, with properly chosen value of the residual smoothing parameter γ in DP, the (truncated) spectral filter estimators are essentially optimally adaptive under source conditions. This optimal value of γ depends only on the known folding operator $\mathcal{K}$ and not on the estimated signal. Numerical experiments that illustrate the behavior of the proposed procedures in two stereological inverse problems are described in Section 3. Section 4 contains some final conclusions.

Earlier results on adaptivity in Poisson inverse problems under L^2 risk are scarce and often restricted to specific problems. For the tomography problem, an adaptive solution based on the wavelet-vaguelet decomposition of the Radon transform was constructed in Cavalier and Koo (2002). Adaptive wavelet solutions, rate-minimax over some Besov balls (up to a log factor) were found for a stereological Spektor-Lord-Willis problem in Ćmiel (2010, 2012). In Ćmiel et al. (2018), 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 Sobolev spaces. A variant of the Lepski balancing principle, suitable for general Poisson inverse problems handled with penalized likelihood approach, was also studied in Werner and Hohage (2012); Hohage and Werner (2016). In Kroll (2019), a circular deconvolution was studied for Poisson data with weighted L^2 risk, an adaptive solution based on model selection ideas was constructed and many additional references on Poisson inverse problems were given.

Another version of DP, suitable for discretized Poisson inverse problems and based on matching the Kullback-Leibler divergence to its expectation, was proposed in Zanella et al. (2009), see also Bardsley and Goldes (2009); Bertero et al. (2010). This was further refined in Sixou et al. (2018) by proving consistency and convergence rates of the Kullback-Leibler risk under some source conditions. Thresholded wavelet solution, adaptive under Kullback-Leibler risk, was also constructed in Antoniadis and Bigot (2006). We prefer, however, to work with the L^2 loss, which measures the distance between the objects of interest rather than between the corresponding distributions, as Kullback-Leibler divergence does.

2. Spectral filter estimators

The Poisson inverse problem (1.3) will be considered in its sequence form (1.4), with the signal f described via its Fourier coefficients $(\mu_i)_i$ w.r.t. the basis of right singular elements $(v_i)_i$ of the operator $\mathcal{K}$. As in Mika and Szkutnik (2021), we define a general spectral filter estimator $\hat{f}_\alpha$ of f as

$$\hat{f}_\alpha = \sum_i F_\alpha(\sigma_i^2)\sigma_i Z_i v_i, \tag{2.1}$$

where $\{F_\alpha(\cdot), \alpha > 0\}$ is a *regularizing filter*. This means that $F_\alpha : (0, \sigma_1^2] \to \mathbb{R}$ are bounded functions, $F_\alpha(\lambda) \to 1/\lambda$ for all λ in the domain, as $\alpha \to 0$ and $F_\alpha(\lambda) \leq C_R/\lambda$ for all α and λ, with some constant C_R. The filter is called a *bounded filter*, if there exists a positive constant C_F such that $|F_\alpha(\lambda)| \leq C_F/\alpha$ for all λ in the domain and all $\alpha > 0$. As special cases, one obtains 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: $F_{1/k}(\lambda) = [1-(1-\beta\lambda)^k]/\lambda$, with $\beta \in (0, \sigma_1^{-2})$.

In order to control the stochastic part of the estimation error with DP based on smoothed residuals, we propose a modified filter boundedness condition.

Definition 2.1. The filter $\{F_\alpha, \alpha > 0\}$ is called γ-bounded, if there exists a positive constant $C_{F,\gamma}$ such that

$$\lambda^{\frac{1-\gamma}{2}} |F_\alpha(\lambda)| \leq C_{F,\gamma}/\alpha^{\frac{1+\gamma}{2}}$$

for all $\lambda \in (0, \|\mathcal{K}\|^2]$ and all $\alpha > 0$.

Note that, with $\gamma = 1$, the γ-boundedness of the filter reduces to the standard boundedness used in the inverse problems literature (e.g., Harrach et al., 2020; Mika and Szkutnik, 2021). It is also easily seen that truncated SVD (and, more generally, any filter with $F_\alpha(\lambda) = 0$ for $\lambda < C\alpha$, with some $C > 0$) is γ-bounded with any positive γ, and that the Landweber and the k-times iterated Tikhonov filters are γ-bounded for $\gamma \in [0,1]$. In particular, for the Landweber filter with the regularization parameter $\alpha = 1/k$, one has, after consecutive substitutions $x := \beta\lambda$ and $z := (1-x)^k$,

$$\lambda^{\frac{1-\gamma}{2}}\Big[1-(1-\beta\lambda)^k\Big]/\lambda = \beta^{\frac{1+\gamma}{2}}(1-z)/(1-z^\alpha)^{\frac{1+\gamma}{2}}$$
$$\leq \beta^{\frac{1+\gamma}{2}} \sup_{z\in(0,1)} (1-z)^{\frac{1-\gamma}{2}}/\alpha^{\frac{1+\gamma}{2}} = \beta^{\frac{1+\gamma}{2}}/\alpha^{\frac{1+\gamma}{2}},$$

because $1-z^\alpha \geq \alpha(1-z)$ for $z \in (0,1)$.

In order to properly handle the bias part of the error, we need filters with sufficiently high qualification.

Definition 2.2. 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 $C_\nu > 0$ for which

$$\sup_{\lambda\in(0,\|\mathcal{K}\|^2]} \lambda^{\nu/2}\Big|\lambda F_\alpha(\lambda) - 1\Big| \leq C_\nu \alpha^{\nu/2} \tag{2.2}$$

is satisfied for all α in some interval $(0, \alpha_0]$.

It is easy to see that truncated SVD and Landweber regularization have infinite qualification and the qualification of the k-times iterated Tikhonov regularization equals $2k$. Note, however, that the qualification is sometimes (e.g. in Engl et al., 1996) defined as $\nu_0/2$.

It will be convenient to express the norms in the DP condition in terms of Fourier coefficients w.r.t. the basis of right singular elements $(v_i)_i$ of the operator $\mathcal{K}$. Note for future reference that

$$\hat{q}_i := \langle \mathcal{K}^*\tilde{g}, v_i\rangle = \langle \tilde{g}, \mathcal{K}v_i\rangle = \sigma_i\langle \tilde{g}, u_i\rangle = \sigma_i Z_i$$

is an unbiased estimator of $q_i := \langle \mathcal{K}^* g, v_i\rangle$ and write $\hat{q} := \sum_i \sigma_i Z_i v_i$ and $q = \sum_i q_i v_i$. Standard calculation gives

$$\|(\mathcal{K}^*\mathcal{K})^{(\gamma-1)/2}(\mathcal{K}^*\mathcal{K}\hat{f}_\alpha - \hat{q})\|^2 = \sum_i \sigma_i^{2\gamma}(\sigma_i^2 F_\alpha(\sigma_i^2) - 1)^2 Z_i^2$$

and

$$E\Big(\|(\mathcal{K}^*\mathcal{K})^{(\gamma-1)/2}(\hat{q} - q)\|^2\Big) = \sum_i \sigma_i^{2\gamma-2}\mathrm{Var}(\sigma_i Z_i) = \frac{1}{n}\sum_i \sigma_i^{2\gamma}\langle g, u_i^2\rangle$$
$$= E\Big(\frac{1}{n^2}\sum_{j=1}^{N_n} w(Y_j)\Big),$$

with

$$w(y) := \sum_i \sigma_i^{2\gamma} u_i^2(y).$$

This means that

$$\hat{\delta}^2_{n,\gamma} := \frac{1}{n^2}\sum_{j=1}^{N_n} w(Y_j)$$

is an unbiased estimator of the "weighted noise level" δ^2_γ.

The DP sequence $(\alpha_n)_{n\in\mathbb{N}}$ of regularization parameters will be defined with

$$\alpha_n = \sup\Big\{\alpha : \sum_i \sigma_i^{2\gamma}(\sigma_i^2 F_\alpha(\sigma_i^2) - 1)^2 Z_i^2 < \tau\hat{\delta}^2_{n,\gamma}\Big\}. \tag{2.3}$$

For $\alpha \to 0$, the left-hand side of the defining inequality tends to zero. If the functions $\alpha \to F_\alpha(\lambda)$ are left-continuous for all fixed λ, the DP sequence is well defined with (2.3) and satisfies the inequality

$$\|(\mathcal{K}^*\mathcal{K})^{(\gamma-1)/2}(\mathcal{K}^*\mathcal{K}\hat{f}_{\alpha_n} - \hat{q})\|^2 \leq \tau\hat{\delta}^2_{n,\gamma}. \tag{2.4}$$

Convergence rates will be studied for smoothness classes defined via source condition (cf., e.g. Bissantz et al., 2007)

$$f \in \mathcal{F}_{\nu,\rho} := \Big\{ f = (\mathcal{K}^*\mathcal{K})^{\nu/2} s : \|s\| \le \rho \Big\}, \tag{2.5}$$

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

$$f \in E_{b\nu,R} := \Big\{ f : \sum_i i^{2b\nu} \langle f, v_i \rangle^2 \le R^2 \Big\}, \tag{2.6}$$

with some R, and one may expect minimax L^2 convergence rates $n^{-2b\nu/(2b\nu+2b+1)}$ (cf. Mair and Ruymgaart, 1996; Bissantz et al., 2007; Mika and Szkutnik, 2021).

To ensure finiteness of the weighted noise level δ_γ^2 and existence of $\hat{f}_\alpha$ as a well defined element of H_1 for filters that are bounded and γ-bounded, we make

Assumption 1: $\sum_i \sigma_i^{2\gamma} \langle g, u_i^2 \rangle < \infty$.

With $\gamma = 1$, this is exactly the same as Assumption 1 in Mika and Szkutnik (2021). For finitely smoothing $\mathcal{K}$ with $\sigma_i \asymp i^{-b}$, our Assumption 1 is satisfied, if, e.g., $\|g\|_\infty < \infty$, and $\gamma > 1/(2b)$.

With $\hat{\mu}_i := \sigma_i F_\alpha(\sigma_i^2) Z_i$, one has

$$\hat{f}_\alpha = \sum_i \hat{\mu}_i v_i, \;\; E(\hat{\mu}_i) = \sigma_i^2 F_\alpha(\sigma_i^2)\mu_i \text{ and } \mathrm{Var}(\hat{\mu}_i) = \sigma_i^2 F_\alpha^2(\sigma_i^2)\langle g, u_i^2 \rangle / n.$$

Define $f_\alpha = \sum_i \sigma_i^2 F_\alpha(\sigma_i^2)\mu_i v_i$. Then, for γ-bounded filters,

$$E(\|\hat{f}_\alpha - f_\alpha\|^2) = \sum_i \mathrm{Var}(\hat{\mu}_i) = n^{-1} \sum_i \sigma_i^2 F_\alpha^2(\sigma_i^2)\langle g, u_i^2 \rangle \le C \sum_i \sigma_i^{2\gamma} \langle g, u_i^2 \rangle < \infty,$$

with some C, under Assumption 1. This implies that, almost surely, $\|\hat{f}_\alpha - f_\alpha\| < \infty$ and, consequently, $\|\hat{f}_\alpha\| < \infty$ because, for bounded filters,

$$\|f_\alpha\|^2 = \sum_i \sigma_i^4 F_\alpha^2(\sigma_i^2)\mu_i^2 \le C_R \|f\|^2 < \infty.$$

As in Mika and Szkutnik (2021), we also make

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.

As discussed above, this ensures that the DP sequence is well defined.

The following theorem generalizes Theorem 3 in Mika and Szkutnik (2021) and shows that DP-based spectral filter estimators may converge at essentially minimax rates, provided the residuals are not smoothed too much. Theorem 3 in Mika and Szkutnik (2021) is obtained as special case with $\gamma = 1$ and a Hilbert-Schmidt operator $\mathcal{K}$.

Theorem 2.3. *Let $\{F_\alpha, \alpha > 0\}$ be a bounded regularizing filter that is also γ-bounded and has qualification $\nu_0 > \max\{2, \gamma + 1\}$. Let $(\alpha_n)_{n\in\mathbb{N}}$ be a DP sequence, as defined in (2.3), and let $\hat{f}_{\alpha_n}$ be the estimator (2.1) of f in the model (1.3) with an injective and compact 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 $\nu + \gamma + 1 \le \nu_0$, then $\|\hat{f}_{\alpha_n} - f\|^2 = \mathcal{O}_P(n^{-\nu/(\nu+\gamma+1)})$.*

Under conditions of Theorem 2.3 and for finitely smoothing $\mathcal{K}$ with $\sigma_i \asymp i^{-b}$, if $\gamma \searrow 1/(2b)$, the convergence rates become arbitrarily close to $n^{-2b\nu/(2b\nu+2b+1)}$, known to be minimax in several problems. The related constants may, however, increase when γ decreases. In effect, the behavior of the error for finite n is hard to predict from Theorem 2.3 alone. Some specific examples are given in a simulation study described in the next section.

At this point, we need no specific assumptions on the value of τ. Any positive τ leads to the same asymptotic conclusions in Theorem 2.3. In finite samples, however, the quality of estimators may heavily depend on the value of τ. Larger values typically lead to more regular solutions. The very idea of DP - to make the size of residuals comparable with the level of noise in the observation model - suggests $\tau \approx 1$ and many previous simulations with various forms of DP (including our previous experience with non-weighted DP for Poisson inverse problems in Mika and Szkutnik, 2021) confirm that, indeed, $\tau = 1$ or somewhat greater than 1 give best finite sample results.

Recall that in non-stochastic inverse problems it is usually assumed that $\tau > 1$, which is used in the convergence analysis to get control over the size of α (cf., e.g., Engl et al., 1996, Thm. 4.17). In our case, any positive τ is allowed, and this is related to a different type of convergence considered in those two problems: worst case convergence in norm in non-stochastic inverse problems, and stochastic convergence in norm in our setup. We get control over the size of α_n by excluding a set of probability diminishing to zero as $n \to \infty$, on which the level of noise may be highly underestimated, which may lead to much too small values of α_n and, consequently, to huge L^2 errors (see the proof of Theorem 2.3 in Section 5 below). This is also the reason why convergence of the standard L^2 risk (mean L^2 error) cannot generally be expected, if standard DP is used in a stochastic inverse problem, cf., e.g., Harrach et al. (2020); Mika and Szkutnik (2021). A counterexample described in Sec. 4.1 in Mika and Szkutnik (2021) can be trivially adapted to our weighted DP. The coefficients in formula (16) in Mika and Szkutnik (2021) change then from 10^{-2i} to $10^{-2i\gamma}$ and the rest of the analysis carries over and shows that, also with our weighted DP, convergence in mean L^2 norm cannot be generally expected. Versions of DP augmented with various forms of so-called "emergency stop" may ensure the convergence of the L^2 risk, but only at non-optimal, slow rates (cf., e.g., Harrach et al., 2020). A referee suggested that, perhaps, stronger convergence of L^2 error might be obtained outside of exponentially small events by using concentration inequalities rather than laws of large numbers and central limit theorems in Hilbert spaces. This interesting idea is left for further studies.

An application of Theorem 2.3 to problems with non-Hilbert-Schmidt operators and standard Landweber or Tikhonov filters may be problematic, because γ should then be greater than 1 and those filters are γ-bounded only for $\gamma \in [0, 1]$. As a remedy, one can then consider modified versions of those filters.

Definition 2.4. Let $\{F_\alpha, \alpha > 0\}$ be a filter and $\eta : (0, \infty) \to (0, \infty)$ be a function for which $\eta(\alpha) \to 0$, as $\alpha \to 0$. Then $\{F_\alpha(\lambda) \cdot I(\lambda \geq \eta(\alpha)), \alpha > 0\}$ is called *a filter truncated by* $\eta(\alpha)$.

The numbers of components in (2.1) with a filter truncated by $\eta(\alpha)$ is finite. Obviously, if $\eta(\alpha) = A\alpha$ with some $A > 0$, the filter truncated by $\eta(\alpha)$ is γ-bounded with any positive γ. Also, if the original filter has qualification ν_0, its version truncated by $\eta(\alpha)$ has qualification at least ν_0, if $\eta(\alpha) \leq A\alpha$. (Nothing changes for $\lambda \geq \eta(\alpha)$, and for $\lambda < \eta(\alpha)$ the inequality in (2.2) holds true, possibly with suitably modified C_ν.)

3. Examples and simulations

The behavior of the procedures introduced and studied in the previous section will be illustrated on two stereological problems of unfolding the distribution of the radii of balls randomly placed in an opaque medium and only observable via cross-sections. The main goal will be an empirical verification of the predictions of Theorem 2.3 about the impact of smoothing residuals on decay rates of the L^2 errors.

We assume that the balls' radii are i.i.d. random variables in $[0, 1]$ with density ρ w.r.t. Lebesgue measure and that the balls' centers form a homogeneous Poisson process in $\mathbb{R}^3$ with the expected number of c points per unit volume, stochastically independent of the radii.

In the Spektor-Lord-Willis (SLW) problem, one observes a random, linear section through the medium and the line segments formed as intersections of the balls with the linear probe. It is known (cf. Szkutnik, 2007) that the observed line segments' radii Y_j form a Poisson process on $[0,1]$ with the intensity function (w.r.t. Lebesgue measure) $ng(u)$, where n is known and related to the total length of the observed linear probe and

$$g(y) = 2y \int_y^1 f(x)\,\mathrm{d}x,$$

with $f(x) = c\rho(x)$. Given Y_j's, the goal is to unfold f. The SVD of this operator acting in $L^2([0,1], \mathrm{d}x)$, although explicitly known (see Szkutnik, 2007), is, however, rather complicated. A simpler form of SVD is obtained, if one defines new dominating measures $\mathrm{d}\mu_X = x\,\mathrm{d}x$ and $\mathrm{d}\mu_Y = y\,\mathrm{d}y$ in the solution and data space, respectively. Then, g and f are replaced with $l(y) = g(y)/y$ and $h(x) = f(x)/x$ and the operator becomes

$$l(y) = (\mathcal{K}h)(y) = 2\int_y^1 h(x)\,\mathrm{d}\mu_X(x) = \int_0^1 K(x,y)h(x)\,\mathrm{d}\mu_X(x),$$

with $K(x,y) = 2yI(x \geq y)$. For this Hilbert-Schmidt operator acting from $L^2([0,1],\mu_X)$ to $L^2([0,1],\mu_Y)$, one obtains the following SVD (see Dudek and Szkutnik, 2008)

$$\sigma_i = \frac{2}{\pi(2i-1)}, \quad v_i(x) = 2\sin\left[(2i-1)\pi x^2/2\right],$$
$$u_i(y) = 2\cos\left[(2i-1)\pi y^2/2\right], \quad i = 1,2,\ldots.$$

Formula (2.1) with (α_n) given in (2.3) then gives an estimator of h, which further gives an estimator of f after multiplication by x. For $\gamma = 1$, the function $w(y)$ can be written in an explicit form

$$w(y) = \int \sum_{i=1}^{\infty} \sigma_i^{2\gamma} u_i^2(y) v_i^2(x)\,\mathrm{d}\mu_X(x) = \int K^2(x,y)\,\mathrm{d}\mu_X(x) = 2(1-y^2).$$

Otherwise, it is only given as an infinite sum with guaranteed pointwise convergence for $\gamma > 1/2$, because u_i's are uniformly bounded.

In the Wicksell's problem, one observes a random planar section through the medium and the radii Y_j of the circles formed as intersections of the balls with the cutting plane. With $\mathrm{d}\mu_X = (4x)^{-1}\,\mathrm{d}x$, $\mathrm{d}\mu_Y = 4\pi^{-1}(1-y^2)^{1/2}\,\mathrm{d}y$ and $l(y) = (\pi/4)(1-y^2)^{-1/2}g(y)$, $h(x) = 4xf(x)$, one obtains

$$l(y) = (\mathcal{K}h)(y) = \frac{\pi y}{4(1-y^2)^{1/2}} \int_y^1 \frac{h(x)}{(x^2-y^2)^{1/2}}\,\mathrm{d}\mu_X(x)$$

and the SVD of the form

$$\sigma_i = \frac{\pi}{8i^{1/2}}, \quad v_i(x) = 4i^{1/2}x^2 P_{i-1}^{0,1}(2x^2-1)$$
$$u_i(y) = 2yU_{i-1}(2y^2-1) = U_{2i-1}(y), \quad i = 1,2,\ldots,$$

where $P_i^{0,1}$ is the Jacobi polynomial of type $(\alpha,\beta) = (0,1)$ and order i and U_i is the second kind Chebyshev polynomial of order i (see Johnstone and Silverman, 1991, but note that the factor 16 has to be corrected to 8 in the denominator of σ_i there). Note that the operator $\mathcal{K}$ is not Hilbert-Schmidt in this case. It is an Abel-type operator, which is closely related to the half-integration operator, cf. Anderssen and de Hoog (1990). Again, formula (2.1) then gives an estimator of h, which further gives an estimator of f after division by $4x$.

As in Mika and Szkutnik (2021), we used in simulations the following test functions:

- *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 ($SMLB$):
  $f(x) = 1.241(2x - x^2)^{-3/2} \exp\Big(1.21[1 - (2x - x^2)^{-1}]\Big) I_{[0,1]}(x)$,
- $Bimodal$:
  $f(x) = \frac{28125}{512} x^2(0.8 - x)^2 I_{[0,0.8]}(x) + \frac{9375}{8}(0.6 - x)^2(1 - x)^2 I_{[0.6,1]}(x)$

All those functions are probability densities, which means that we set $c = 1$.

The quality of spectral filter estimators depends on how well the function h can be approximated in the space $L^2(\mu_X)$ with the initial right singular elements v_i, which is quantified with the decay rate of the corresponding Fourier coefficients. Assume that $\langle h, v_i\rangle^2 \asymp i^{-r}$ and $\sigma_i \asymp i^{-b}$. Then, in order to have $h \in E_{b\nu,R}$ with some R, one needs $2b\nu - r < -1$, cf. (2.6), or $2b\nu < r - 1$. With $\gamma \searrow 1/(2b)$, Theorem 2.3 ensures then (in the limit) the convergence rate

$$n^{-2b\nu/(2b\nu+2b+1)} = n^{-(r-1)/(r+2b)},$$

which is minimax over $E_{b\nu,R}$, and the actual convergence to h should be at least that fast.

In order to get at least some approximate idea of what the values of r for our test functions are, the Fourier coefficients $\langle h, v_i\rangle$ were computed numerically and the slopes of the linear regression of $\log\langle h, v_i\rangle^2$ on $\log(1/i)$ were taken as the approximate values of r. Numerical integration was performed with the R function `integrate()`. The integrated functions are well behaved in the SLW problem, for which the values of r were computed from the first 100 Fourier coefficients. In the Wicksell's problem, however, the functions v_i are polynomials of degree $2i$ and they are not uniformly bounded. This leads to serious problems with numerical integration and, more generally, with evaluating the values of v_i for large i, especially in the vicinity of 1. To circumvent numerical instabilities, we thus restricted the numerical integration to the interval [0, 0.96] and based the computing of r on the first 20 Fourier coefficients only. (An attempt to go beyond 20 coefficients or beyond the interval [0, 0.96] resulted in the sums of squares of the Fourier coefficients significantly greater than the squared L^2 norms of the test functions, which indicated that the computed values of the Fourier coefficients were unreliable.) Consequently, the expected rates for the Wicksell's problem should be treated with respective caution. The initial Fourier coefficients are illustrated in Figure 3.1 and the expected minimax rates are summarized in Table 3.1.

TABLE 3.1. Approximate decay rates of squared Fourier coefficients and corresponding guaranteed rates for $L^2(\mu_X)$ squared errors.

	SLW		Wicksell	
	squared Fourier coeff.	squared error	squared Fourier coeff.	squared error
$Beta(4,2)$	$i^{-4.1}$	$n^{-0.51}$	$i^{-3.3}$	$n^{-0.53}$
$SMLA$	$i^{-3.2}$	$n^{-0.42}$	$i^{-2.7}$	$n^{-0.46}$
$SMLB$	$i^{-3.1}$	$n^{-0.41}$	$i^{-3.0}$	$n^{-0.50}$
$Bimodal$	$i^{-2.8}$	$n^{-0.38}$	$i^{-4.4}$	$n^{-0.63}$

Two spectral filters were used in the simulation: the Landweber filter and the k-times iterated Tikhonov filter, both with the filter truncating function $\eta(\alpha) = \sigma_1^2\alpha^2$. In order to predict convergence rates from Theorem 2.3, one needs to fulfill the conditions $\nu_0 > \max\{2, \nu + \gamma + 1\}$ and $\nu < (r - 1)/(2b)$. It is thus sufficient to have

$$\nu_0 > \max\{2, (r - 1)/(2b) + \gamma + 1\}. \tag{3.1}$$

Recall that the qualification ν_0 is infinite for the Landweber filter, and equals $2k$ for the k-times iterated Tikhonov filter. For the SLW problem, results will be presented for $\gamma = 0.5, 1.0$ and 1.5 and, with the numerically obtained values of r's, one may reasonably expect that condition (3.1) is

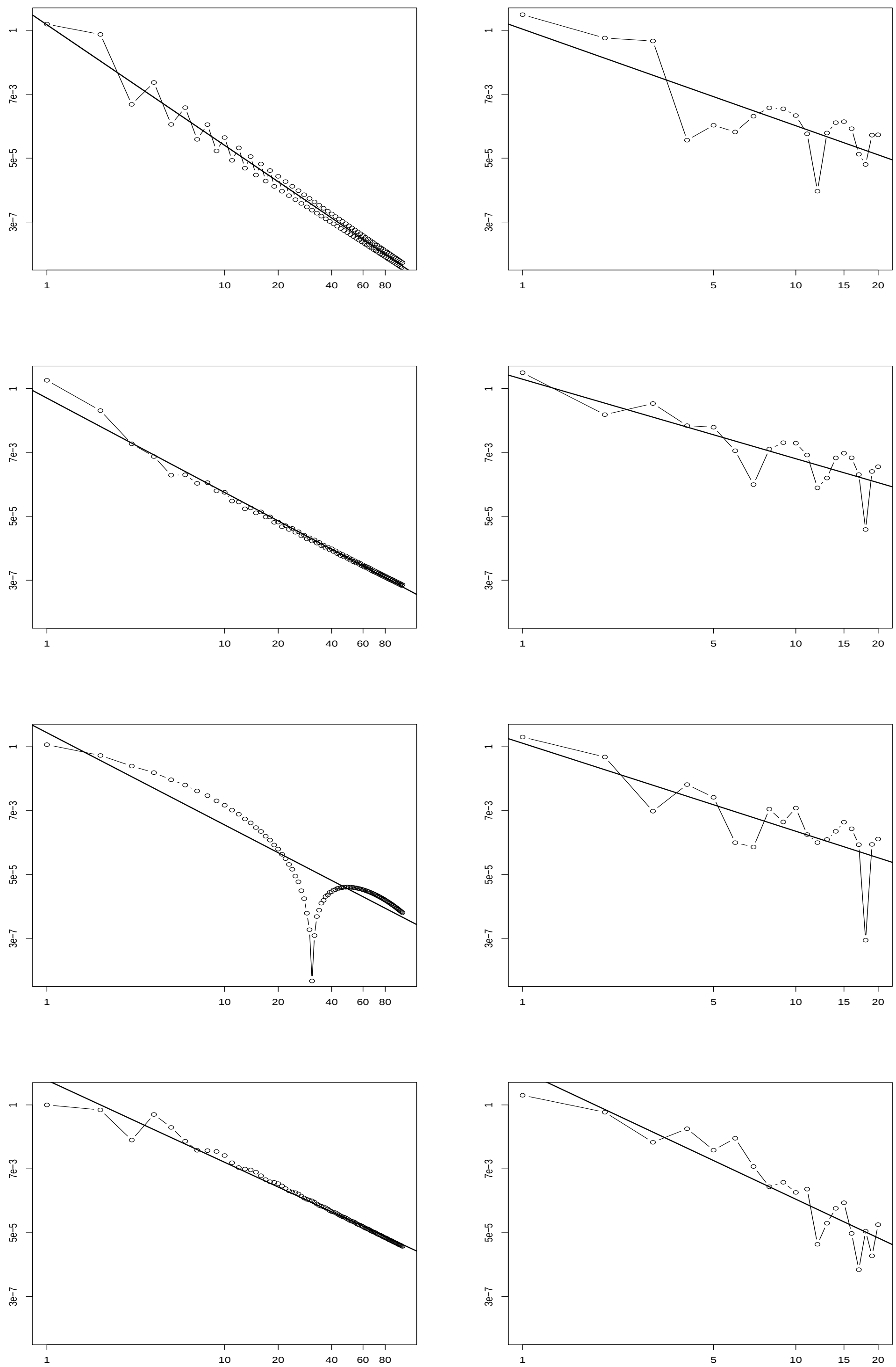

Figure 3.1. First nc squared Fourier coefficients in the SLW problem (left, $nc = 100$) and in the Wicksell's problem (right, $nc = 20$) for $Beta(4, 2)$, $SMLA$, $SMLB$ and $Bimodal$ functions (top to bottom). Points represent squared coefficients vs their indices, both on logarithmic scales, along with least squares straight lines fitted to assess decay rates.

satisfied for 2-times iterated Tikhonov filter. For the Wicksell's problem, results will be presented for $\gamma = 1, 2$ and 5, and 4-times iterated Tikhonov filter should satisfy (3.1).

Four values of the experiment size n were used: 2000, 10^4, 10^5 and 10^6. For each n and each test function, 10 artificial data samples were generated and the estimators of h and their squared $L^2(\mu_X)$ errors were computed for each sample. The errors were computed via numerical integration using the R function `integrate()`. Data samples were generated by mimicking the physical mechanism of data generation. First, for a given experiment size n, a number $\tilde{n}$ of balls was generated from a Poisson distribution with expectation n. Then, $\tilde{n}$ pairs (R, D) of random variables were independently generated: ball radii R from density f and the distances D of the ball center from the probe. For the SLW problem, D was generated from the density $2xI_{[0,1]}$ and for the Wicksell's problem, from the uniform density on $[0, 1]$. If $D > R$, which means that the probe misses the ball, the pair (R, D) was dropped. Otherwise, the observed radius Y was computed as $Y = (R^2 - D^2)^{1/2}$. The mean sample sizes as percentages of n are given in Table 3.2.

TABLE 3.2. Mean sample sizes in percentages of experiment size.

	SLW	Wicksell
$Beta(4, 2)$	48%	66%
$SMLA$	54%	71%
$SMLB$	41%	60%
$Bimodal$	35%	56%

The number of Fourier components used for computing the estimators was 50 and 20, respectively, in the SLW and the Wicksell's problem. The relaxation parameter for the Landweber filter was set to $\beta = 2$ and the threshold parameter τ in the DP inequality in (2.3) was set to 1.1 for the Landweber and to 2.0 for the Tikhonov filter, which proved to give reasonable results in all cases under study. The choice was partially based on our previous experience described in Mika and Szkutnik (2021) and no attempt was made to further optimize those values. The values α_n of the regularization parameter were computed by studying the inequality in (2.3) on the grid $q, q^2, q^3, \ldots$ of values of α, with $q = 0.99$. The computed squared $L^2(\mu_X)$ errors, normalized by squared $L^2(\mu_X)$-norms of the respective test function and averaged over 10 data samples generated for each combination of n and test function, are presented in Figures 3.2 and 3.3 on logarithmic scales as functions of n along with least squares straight lines fitted in order to asses the squared errors convergence rates. The quality of the estimators of the original test functions for the experiment size $n = 10^5$ is illustrated in Figures 3.4 and 3.5.

Computations were performed in the R environment v. 3.6.2, with packages `orthopolynom`, `polynom` and `poly` to handle the Jacobi and Chebyshev polynomials. The Poisson and $Beta(4, 2)$ random numbers were generated with `rpois()` and `rbeta()`. The acceptance-rejection method was used to generate random numbers from the remaining test functions.

4. Discussion and conclusions

The simulation results are in good agreement with Theorem 2.3 and show that the proper choice of the residual smoothing parameter γ is crucial for the quality of estimators, both in finite samples and in terms of the convergence rates. The optimal (in the limit) value of γ, as predicted by Theorem 2.3, is 0.5 for the SLW problem and 1 for the Wicksell's problem. In almost all cases handled in the simulation, the errors obtained with those optimal values of γ not only converge at rates better than those obtained with larger γ's, but are also smaller, in many cases significantly so, than the errors obtained with larger γ's. As expected, for optimal γ's, the error decay rates are in all cases faster than the minimax rates over Sobolev ellipsoids $E_{b\nu,R}$ that contain the respective

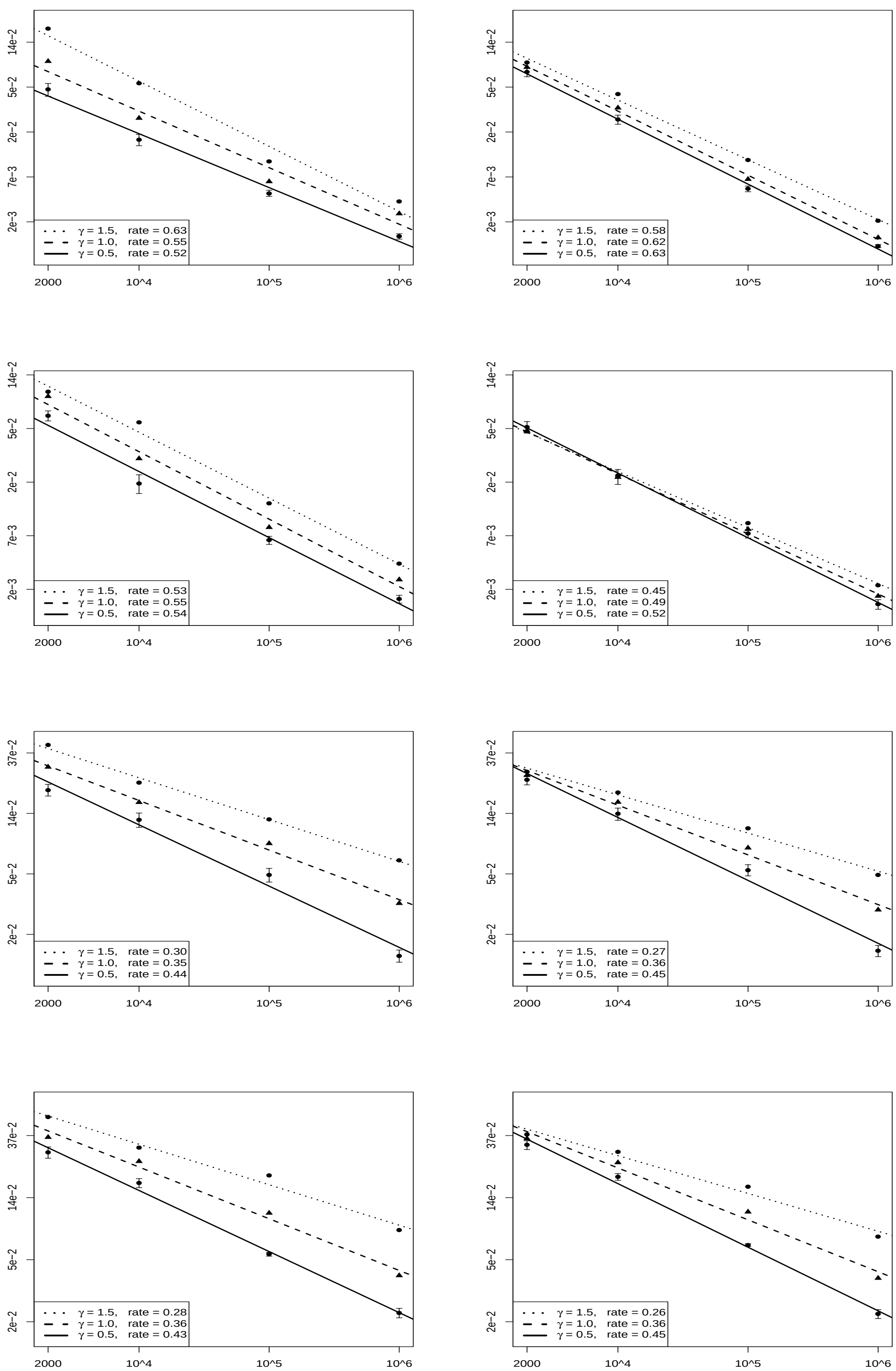

FIGURE 3.2. Convergence of squared errors in SLW problem for Landweber (left) and 2-times iterated Tikhonov (right) for $Beta(4, 2)$, $SMLA$, $SMLB$ and $Bimodal$ test functions (top to bottom). Points represent normalized squared errors vs experiment size on logarithmic scales, with one-standard-error bars added to points corresponding to optimal $\gamma = 0.5$.

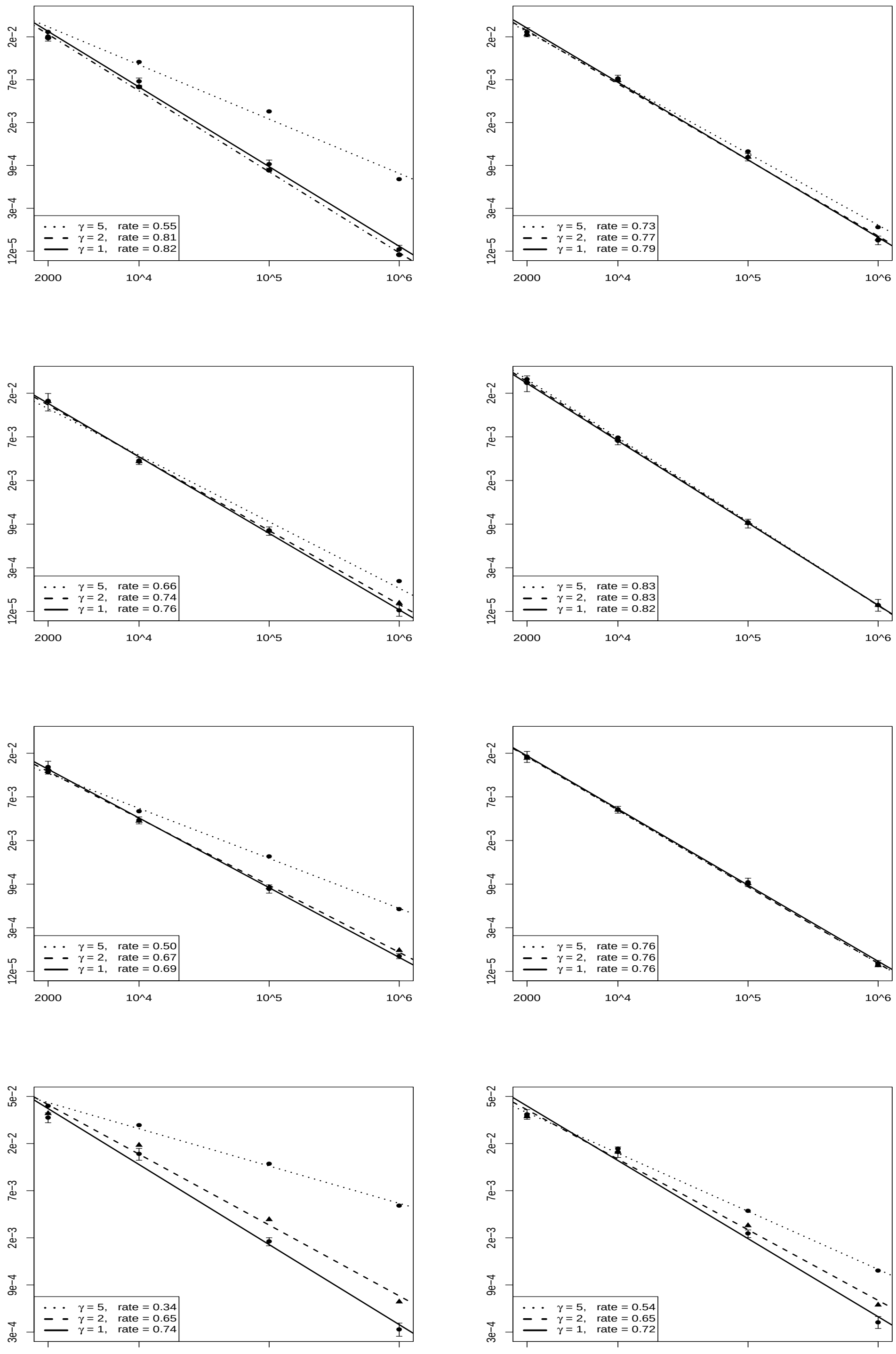

FIGURE 3.3. Convergence of squared errors in Wicksell's problem for Landweber (left) and 4-times iterated Tikhonov (right) for $Beta(4,2)$, $SMLA$, $SMLB$ and $Bimodal$ (top to bottom). Points represent normalized squared errors vs experiment size on logarithmic scales, with one-standard-error bars added to points corresponding to optimal $\gamma = 1$.

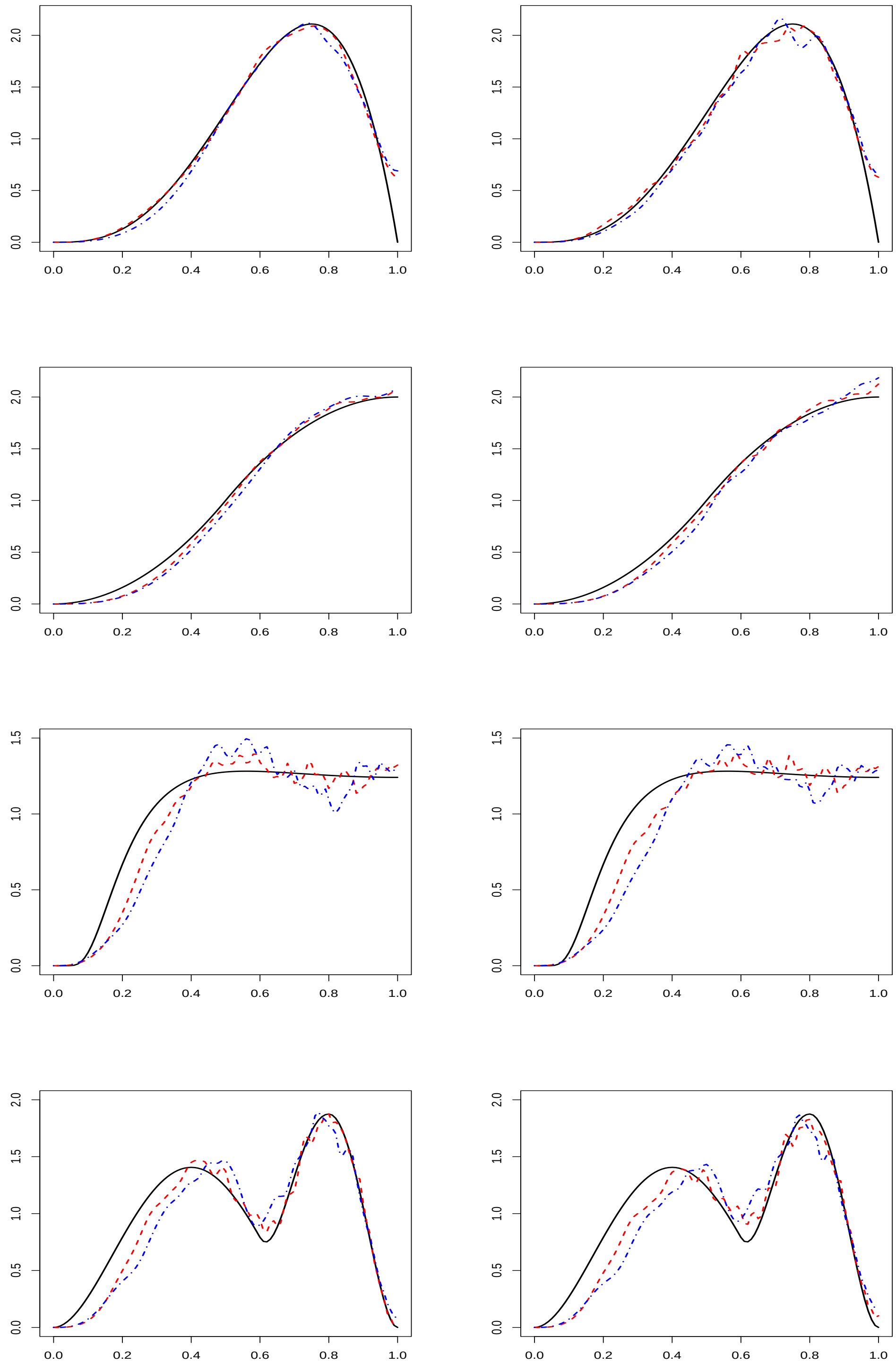

FIGURE 3.4. Best (red, dashed) and worst (blue, dot-dashed) estimators (out of 10 data samples) of *Beta*(4, 2), *SMLA*, *SMLB* and *Bimodal* test functions (top to bottom, solid black lines), obtained with Landweber (left) and 2-times iterated Tikhonov filters in the SLW problem, with $n = 10^5$ and $\gamma = 0.5$.

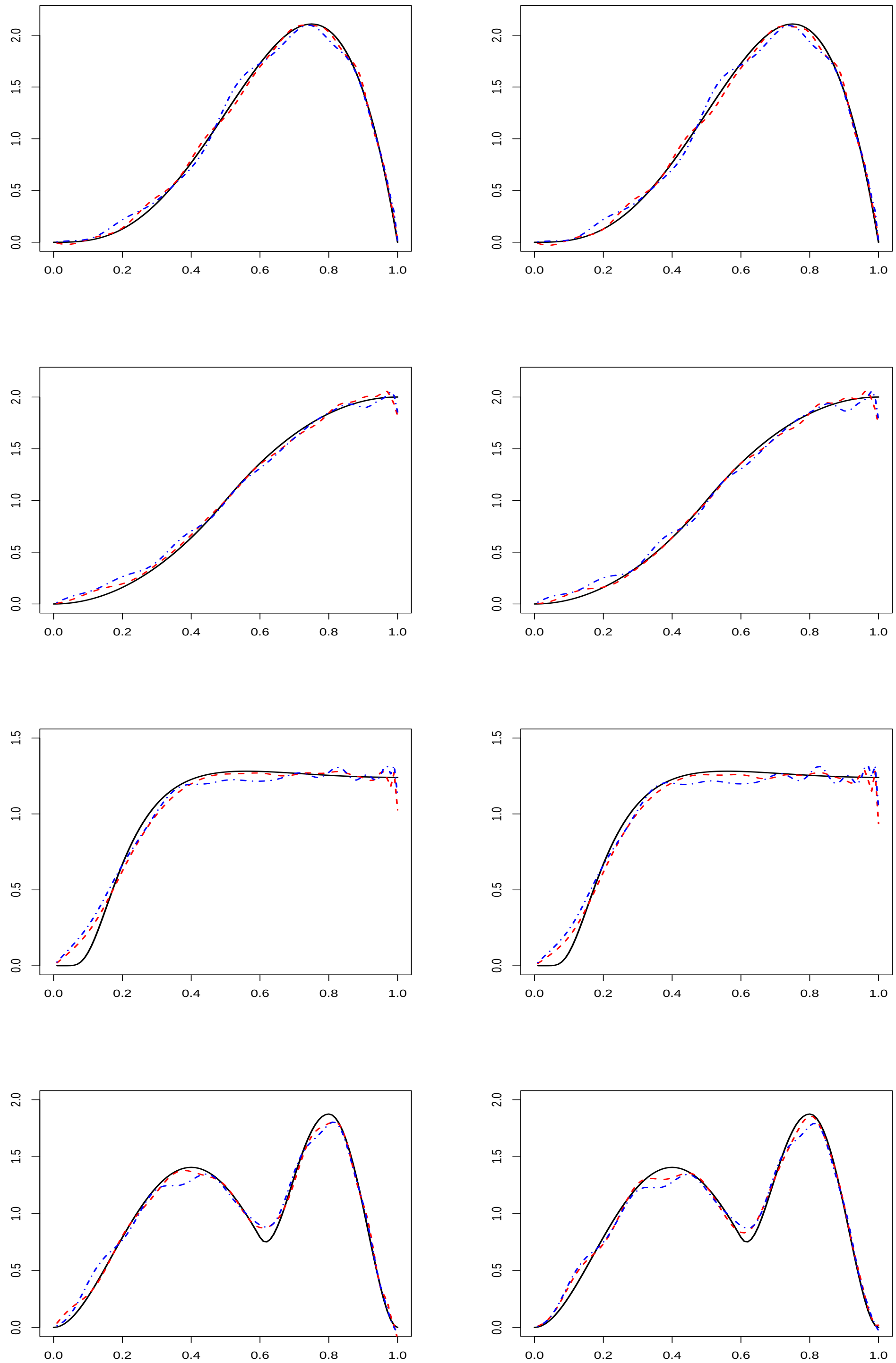

FIGURE 3.5. Best (red, dashed) and worst (blue, dot-dashed) estimators (out of 10 data samples) of $Beta(4, 2)$, $SMLA$, $SMLB$ and $Bimodal$ test functions (top to bottom, solid black lines), obtained with Landweber (left) and 4-times iterated Tikhonov filters in the Wicksell's problem, with $n = 10^5$ and $\gamma = 1$.

test functions. This is not always the case with larger γ's, e.g. for $SMLB$ and $Bimodal$ in the SLW problem and $Bimodal$ in the Wicksell's problem.

It should be noticed that, strictly speaking, values of γ greater than 1 are not allowed in Theorem 2.3, if one uses $\eta(\alpha) \asymp \alpha^2$ as the filter truncating function, because the truncated Landweber and Tikhonov filters are then not γ-bounded ($\eta(\alpha) \asymp \alpha$ would do). The deterioration of errors and convergence rates with increasing non-optimal γ seems to be more serious for the Landweber than for the Tikhonov filter.

Standard pre-conditioning, as used, e.g., in Mika and Szkutnik (2021) in Poisson inverse problems, corresponds to $\gamma = 1$ and leads to non-optimally adaptive estimators, unless $b = 1/2$, as in the Wicksell's problem. Theorem 3 in Mika and Szkutnik (2021) is obtained from our Theorem 2.3 with $\gamma = 1$. When $b > 1/2$, as in the SLW problem, our Theorem 2.3 offers a significant improvement over Theorem 3 in Mika and Szkutnik (2021): less smoothed residuals obtained with $\gamma \searrow 1/(2b)$ lead to essentially optimally adaptive estimators.

The Landweber and Tikhonov filters, combined with DP with optimally smoothed residuals, behave very similarly in the SLW and Wicksell's problems. It should be noticed here that data for each subfigure in Figures 3.2–3.5 were generated with the same initial seed for the random numbers generation, which means that the corresponding left and right subfigures are based on the same data samples.

Exact and tractable analytical forms of the folding operator SVD are available only in rather exceptional cases. An alternative is to consider a finite-dimensional, approximate model and the corresponding matrix SVD. If, as in Blanchard et al. (2018a,b); Stankewitz (2020), $\mu_i' s$ in the sequence space model (1.4) are not genuine Fourier coefficients w.r.t. the basis of right singular elements but, instead, some sort of initial discretization is performed, e.g. via a finite element method combined with Galerkin projection, and matrix SVD is computed for the finite-dimensional approximate model, then the direct connection with minimax rates over $\mathcal{F}_{\nu,\rho}$ is lost, because the discretization effects are neglected, but the estimators with the weighted DP, proposed in this article, may still reasonably be expected to be approximately optimal. This conjecture needs, however, further investigations.

5. Proofs

The general idea of the proof of Theorem 2.3 is similar to that of Theorem 3 in Mika and Szkutnik (2021). For completeness, we provide all necessary details.

Define, for brevity, $S_\alpha := \mathcal{K}^*\mathcal{K}F_\alpha(\mathcal{K}^*\mathcal{K}) - Id = F_\alpha(\mathcal{K}^*\mathcal{K})\mathcal{K}^*\mathcal{K} - Id$, with identity operator Id and recall that $q = \mathcal{K}^*\mathcal{K}f$.

Lemma 5.1. *Let $\{F_\alpha, \alpha > 0\}$ be a regularizing filter and let $f \in \mathcal{F}_{\nu,\rho}$. Then, with some constant C and with any $\gamma > -1$*

(1) $\|S_\alpha f\| \le C\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_\alpha q\|^{\frac{\nu}{\nu+\gamma+1}} \rho^{\frac{\gamma+1}{\nu+\gamma+1}}$.

If, additionally, the filter has qualification $\nu_0 \ge \nu + \gamma + 1$, then

(2) $\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_\alpha q\| \le C_{\nu+\gamma+1}\rho\alpha^{\frac{\nu+\gamma+1}{2}}$.

Proof. Set $\xi = \sum_i (F_\alpha(\sigma_i^2)\sigma_i^2 - 1)\sigma_i^{-\nu}\langle f, v_i\rangle v_i$. Then, $(\mathcal{K}^*\mathcal{K})^{\nu/2}\xi = S_\alpha f$. The interpolation inequality ((2.49) in Engl et al., 1996 with $r = \nu/2$ and $q = (\nu + \gamma + 1)/2$) gives

$$\|S_\alpha f\| = \|(\mathcal{K}^*\mathcal{K})^{\nu/2}\xi\| \le \|(\mathcal{K}^*\mathcal{K})^{\frac{\nu+\gamma+1}{2}}\xi\|^{\frac{\nu}{\nu+\gamma+1}}\|\xi\|^{\frac{\gamma+1}{\nu+\gamma+1}}$$
$$= \|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma+1}{2}} S_\alpha f\|^{\frac{\nu}{\nu+\gamma+1}}\|\xi\|^{\frac{\gamma+1}{\nu+\gamma+1}} = \|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_\alpha q\|^{\frac{\nu}{\nu+\gamma+1}}\|\xi\|^{\frac{\gamma+1}{\nu+\gamma+1}}.$$

Moreover,

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

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

For the second statement, write

$$\begin{aligned}
\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_\alpha q\| &= \Big(\sum_i \sigma_i^{2\gamma+2}\Big(F_\alpha(\sigma_i^2)\sigma_i^2 - 1\Big)^2 \langle f, v_i\rangle^2\Big)^{1/2} \\
&\le \sup_{\lambda\in(0,\|K\|^2]} \lambda^{\frac{\nu+\gamma+1}{2}} \Big|\lambda F_\alpha(\lambda) - 1\Big| \Big(\sum_i \sigma_i^{-2\nu}\langle f, v_i\rangle^2\Big)^{1/2} \\
&\le C_{\nu+\gamma+1}\alpha^{\frac{\nu+\gamma+1}{2}}\rho.
\end{aligned}$$

□

Lemma 5.2. *Let $\{F_\alpha, \alpha > 0\}$ be a regularizing filter with qualification $\nu_0 > \max\{2, \gamma+1\}$ that satisfies Assumption 2. Then, for any fixed f, there exists a function $g : \mathbb{R}^+ \to \mathbb{R}^+$ such that $g(\alpha) \to \infty$ and $\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{\alpha g(\alpha)} q\| = o(\alpha)$, as $\alpha \to 0$.*

Proof. We first prove that $\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_\alpha q\| = o(\alpha)$. Let $\epsilon > 0$ and fix L such that $C_2^2\sigma_{L+1}^{2\gamma}\sum_{i=L+1}^\infty \langle f, v_i\rangle^2 < \epsilon$. Then

$$\begin{aligned}
\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_\alpha q\|^2\alpha^{-2} &= \alpha^{-2}\sum_i \Big(F_\alpha(\sigma_i^2)\sigma_i^2 - 1\Big)^2 \sigma_i^{2\gamma+2}\langle f, v_i\rangle^2 \\
&\le \alpha^{-2}\sum_{i=1}^{L} \Big(F_\alpha(\sigma_i^2)\sigma_i^2 - 1\Big)^2 \sigma_i^{2\nu_0}\sigma_i^{2(\gamma+1-\nu_0)}\langle f, v_i\rangle^2 \\
&\quad + \alpha^{-2}\sum_{i=L+1}^{\infty} \Big(F_\alpha(\sigma_i^2)\sigma_i^2 - 1\Big)^2 \sigma_i^2\sigma_i^{2\gamma}\langle f, v_i\rangle^2 \\
&\le \alpha^{-2}\sup_{\lambda\in(0,\|\mathcal{K}\|^2]} (\lambda^{\nu_0/2}|F_\alpha(\lambda)\lambda - 1|)^2\|f\|^2\sigma_L^{2(\gamma+1-\nu_0)} \\
&\quad + \alpha^{-2}\sup_{\lambda\in(0,\|K\|^2]} \Big(\lambda|F_\alpha(\lambda)\lambda - 1|\Big)^2 \sum_{i=L+1}^{\infty}\langle f, v_i\rangle^2\sigma_{L+1}^{2\gamma} \\
&\le C_{\nu_0}^2\|f\|^2\sigma_L^{2(\gamma+1-\nu_0)}\alpha^{\nu_0-2} + C_2^2\sigma_{L+1}^{2\gamma}\sum_{i=L+1}^{\infty}\langle f, v_i\rangle^2 < 2\epsilon
\end{aligned}$$

for α sufficiently small, which proves that $\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_\alpha q\| = o(\alpha)$. Further, define

$$g(\alpha) := \sup\Big\{t > 0 : \|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{\alpha t} q\|/\alpha \le t^{-1}\Big\}$$

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

$$\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{\alpha g(\alpha)} q\|/\alpha \le 1/g(\alpha) \to 0,$$

as $\alpha \to 0$. $\square$

Proof of Theorem 2.3 Let $\Pi_g^{(i)}$, $i = 1, \dots, n$, be independent copies of a Poisson process Π_g with intensity function g. Then $\Pi_{ng} \overset{d}{=} \sum_{i=1}^n \Pi_g^{(i)}$. Under Assumption 1, we thus have by the law of large numbers

$$n\hat{\delta}_{n,\gamma}^2 = \frac{1}{n}\int w(y)\,\mathrm{d}\Pi_{ng} \overset{d}{=} \frac{1}{n}\sum_{i=1}^n \int w(y)\,\mathrm{d}\Pi_g^{(i)} \xrightarrow{P} E\Big(\int w(y)\,\mathrm{d}\Pi_g\Big) \tag{5.1}$$
$$= \beta_{g,\gamma}^2 := \sum_i \sigma_i^{2\gamma}\langle g, u_i^2\rangle < \infty,$$

which implies that $n\hat{\delta}_{n,\gamma}^2 \xrightarrow{P} \beta_{g,\gamma}^2$, because convergence in distribution to a constant is equivalent to convergence in probability to that constant.

For further use, define $\Delta_{n,\gamma}^2 := \|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}}(\hat{q} - q)\|^2$ and note that

$$(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}}(\hat{q} - q) = \frac{1}{n}\sum_{j=1}^{N_n}\sum_i \sigma_i^\gamma u_i(Y_j)v_i - \sum_i \sigma_i^\gamma\langle g, u_i\rangle v_i \overset{d}{=} \frac{1}{n}\sum_{k=1}^n (T_k - E(T_k)),$$

where $T_k(\cdot)$ are independent copies of

$$T := \sum_i \sigma_i^\gamma \int u_i(y)\,\mathrm{d}\Pi_g(y)\, v_i.$$

Since

$$\|E(T)\|^2 = \sum_i \sigma_i^{2\gamma}\langle g, u_i\rangle^2 \le \sigma_1^{2\gamma}\|g\|^2 < \infty$$

and

$$E\Big(\|T - E(T)\|^2\Big) = \sum_i \sigma_i^{2\gamma}\mathrm{Var}\Big(\int u_i\,\mathrm{d}\Pi_g\Big) = \sum_i \sigma_i^{2\gamma}\langle g, u_i^2\rangle < \infty$$

by Campbell's theorem and Assumption 1, we conclude that $\|T - E(T)\| < \infty$ a.s. and, consequently, $\|T\| < \infty$ a.s., so that T is an H_1-valued random element. Moreover, since

$$\begin{aligned} E\Big(\|T\|^2\Big) &= \sum_i \sigma_i^{2\gamma} E\Big[\Big(\int u_i\,\mathrm{d}\Pi_g\Big)^2\Big] \\ &= \sum_i \sigma_i^{2\gamma}\Big[\mathrm{Var}\Big(\int u_i\,\mathrm{d}\Pi_g\Big) + E^2\Big(\int u_i\,\mathrm{d}\Pi_g\Big)\Big] \\ &= \sum_i \sigma_i^{2\gamma}\Big[\langle g, u_i^2\rangle + \langle g, u_i\rangle^2\Big] \le \beta_{g,\gamma}^2 + \sigma_1^{2\gamma}\|g\|^2 < \infty, \end{aligned}$$

we have

$$\sqrt{n}(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}}(\hat{q} - q) \xrightarrow{d} W,$$

where W is a centered Gaussian random element in H_1 with the same covariance operator as T (cf. Thm 10.5 in Ledoux and Talagrand, 2011 and Corollary 3.2 in Hoffmann-Jørgensen and Pisier, 1976). This implies that

$$n\Delta_{n,\gamma}^2 \xrightarrow{d} \|W\|^2 \tag{5.2}$$

by continuous mapping theorem and, by Slutsky lemma,

$$\Delta_{n,\gamma}^2/\hat{\delta}_{n,\gamma}^2 \xrightarrow{d} \|W\|^2/\beta_{g,\gamma}^2. \tag{5.3}$$

We start the proof from the second assertion of Theorem 2.3. Define

$$\Omega_n = \Big\{|\sqrt{n}\hat{\delta}_{n,\gamma} - \beta_{g,\gamma}| < \frac{\beta_{g,\gamma}}{2},\ \|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}}S_{c_n}(\hat{q} - q)\| < \frac{\sqrt{\tau}\beta_{g,\gamma}}{4\sqrt{n}}\Big\},$$

and observe that, by Markov inequality with $\epsilon = \sqrt{\tau}\beta_{g,\gamma}/4$,

$$\mathrm{P}\Big(\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{c_n}(\hat{q}-q)\| \geq \frac{\epsilon}{\sqrt{n}}\Big) \leq \frac{nE\Big[\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{c_n}(\hat{q}-q)\|^2\Big]}{\epsilon^2}$$
$$= \frac{1}{\epsilon^2}\sum_i \Big(F_{c_n}(\sigma_i^2)\sigma_i^2 - 1\Big)^2 \sigma_i^{2\gamma}\langle u_i^2, g\rangle \to 0$$

with any $c_n \to 0$, because of Assumption 1. This and (5.1) imply that $\mathrm{P}(\Omega_n) \to 1$, as $n \to \infty$. To get control over the size of the elements of the DP-sequence $(\alpha_n)_n$, set $c_n = [\sqrt{\tau}\beta_{g,\gamma}/(4\rho C_{\nu+\gamma+1}\sqrt{n})]^{\frac{2}{\nu+\gamma+1}}$ and write, using Lemma 5.1

$$\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{c_n}\hat{q}\| I(\Omega_n) \leq \|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{c_n} q\| I(\Omega_n) + \|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{c_n}(\hat{q}-q)\| I(\Omega_n)$$
$$\leq \Big[C_{\nu+\gamma+1}\,\rho\, c_n^{\frac{\nu+\gamma+1}{2}} + \frac{\sqrt{\tau}\beta_{g,\gamma}}{4\sqrt{n}}\Big] I(\Omega_n) = \frac{\sqrt{\tau}\beta_{g,\gamma}}{2\sqrt{n}} I(\Omega_n) < \sqrt{\tau}\hat{\delta}_{n,\gamma} I(\Omega_n).$$

This means that, on the set Ω_n,

$$\alpha_n \geq c_n \geq \Big(\frac{\sqrt{\tau}\hat{\delta}_{n,\gamma}}{6\rho C_{\nu+\gamma+1}}\Big)^{\frac{2}{\nu+\gamma+1}}.$$

Further,

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

and, by Lemma 5.1 and inequality (2.4)

$$\|F_{\alpha_n}(\mathcal{K}^*\mathcal{K})q - f\| \leq C\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{\alpha_n} q\|^{\frac{\nu}{\nu+\gamma+1}} \rho^{\frac{\gamma+1}{\nu+\gamma+1}}$$
$$\leq C\rho^{\frac{\gamma+1}{\nu+\gamma+1}}\Big(\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{\alpha_n}\hat{q}\| + \|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{\alpha_n}(\hat{q}-q)\|\Big)^{\frac{\nu}{\nu+\gamma+1}}$$
$$\leq C\rho^{\frac{\gamma+1}{\nu+\gamma+1}}\Big(\sqrt{\tau}\hat{\delta}_{n,\gamma} + (1+C_R)\Delta_{n,\gamma}\Big)^{\frac{\nu}{\nu+\gamma+1}},$$

because $\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{\alpha_n}(\hat{q}-q)\| = \|S_{\alpha_n}(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}}(\hat{q}-q)\|$ and $\|S_{\alpha_n}\| \leq 1 + C_R$. Moreover, using the γ-boundedness of the filter,

$$\|F_{\alpha_n}(\mathcal{K}^*\mathcal{K})(\hat{q}-q)\| = \|F_{\alpha_n}(\mathcal{K}^*\mathcal{K})(\mathcal{K}^*\mathcal{K})^{\frac{1-\gamma}{2}}(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}}(\hat{q}-q)\|$$
$$\leq \sup_{\lambda\in[0,\|\mathcal{K}\|^2]} \lambda^{\frac{1-\gamma}{2}} F_{\alpha_n}(\lambda)\Delta_{n,\gamma} \leq C_{F,\gamma}\Delta_{n,\gamma}/\alpha_n^{\frac{1+\gamma}{2}}.$$

Combining everything,

$$\|\hat{f}_{\alpha_n} - f\| \leq C\rho^{\frac{\gamma+1}{\nu+\gamma+1}}\Big(\sqrt{\tau}\hat{\delta}_{n,\gamma} + (1+C_R)\Delta_{n,\gamma}\Big)^{\frac{\nu}{\nu+\gamma+1}}$$
$$+ C_{F,\gamma}\Big(\frac{6\rho C_{\nu+\gamma+1}}{\sqrt{\tau}}\Big)^{\frac{1+\gamma}{\nu+\gamma+1}} \hat{\delta}_{n,\gamma}^{-\frac{1+\gamma}{\nu+\gamma+1}} \Delta_{n,\gamma}$$
$$\leq D\max\Big\{\hat{\delta}_{n,\gamma}^{\frac{\nu}{\nu+\gamma+1}}, \Delta_{n,\gamma}^{\frac{\nu}{\nu+\gamma+1}}\Big(\frac{\Delta_{n,\gamma}}{\hat{\delta}_{n,\gamma}}\Big)^{\frac{1+\gamma}{\nu+\gamma+1}}\Big\},$$

with some constant D, which completes the proof of the second assertion of Theorem 2.3, because of (5.1), (5.2) and (5.3), and because $\mathrm{P}(\Omega_n) \to 1$.

Now we turn to the consistency part of the theorem. If f has a finite expansion in terms of the singular elements v_i, then it obviously satisfies the source condition with any positive ν. Consequently, $\|\hat{f}_{\alpha_n} - f\|^2 = \mathcal{O}_P(n^{-\nu/(\nu+\gamma+1)})$, if $\nu+\gamma+1 \leq \nu_0$, which is possible with some positive ν, if $\nu_0 > \gamma+1$. Hence, $\hat{f}_{\alpha_n} \to f$ in probability, as $n \to \infty$. If the expansion of f is infinite, so is the expansion of $q = \mathcal{K}^*\mathcal{K}f$. Fix some $\epsilon' > 0$ 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 for bounded filters 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

$$\Omega_n^1 = \left\{\beta_{g,\gamma}/2 \leq \sqrt{n}\,\hat{\delta}_{n,\gamma} \leq 2\beta_{g,\gamma},\ \langle \hat{q}, v_L\rangle^2 \geq \langle q, v_L\rangle^2/2\right\}$$

and note that $\mathrm{P}(\Omega_n^1) \to 1$ because of (5.1) and because, by the law of large numbers, $\langle \hat{q}, v_L\rangle = \sigma_L Z_L \xrightarrow{P} \sigma_L^2\mu_L = \langle q, v_L\rangle$. Then, for $n \geq 64\tau\beta_{g,\gamma}^2/(\langle q, v_L\rangle^2\sigma_L^{2\gamma-2})$ and for all $\alpha \geq \epsilon'$, one has

$$\begin{aligned}\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_\alpha \hat{q}\|^2 I(\Omega_n^1) &\geq \sigma_L^{2\gamma-2}\Big[(\sigma_L^2 F_\alpha(\sigma_L^2) - 1\Big]^2 \langle \hat{q}, v_L\rangle^2 I(\Omega_n^1) \\ &\geq \sigma_L^{2\gamma-2}\langle \hat{q}, v_L\rangle^2/4 \geq \sigma_L^{2\gamma-2}\langle q, v_L\rangle^2/8 \geq \frac{8\tau\beta_{g,\gamma}^2}{n} \geq 2\tau\hat{\delta}_{n.\gamma}^2 I(\Omega_n^1).\end{aligned}$$

This means that on Ω_n^1 we have $\alpha_n \leq \epsilon'$, because of (2.3). Consequently, $\alpha_n \xrightarrow{P} 0$.

Consider the set $\Omega_n \cap \Omega_n^1$ with $c_n = n^{-1/2}g(n^{-1/2})$, where $g(u)$ is defined in Lemma 5.2. Note that $c_n\sqrt{n} \to \infty$ and $\mathrm{P}(\Omega_n \cap \Omega_n^1) \to 1$. On the set $\Omega_n \cap \Omega_n^1$, one has for large n,

$$\begin{aligned}\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{c_n}\hat{q}\| &\leq \|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{c_n} q\| + \|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{c_n}(\hat{q} - q)\| \\ &\leq \frac{\sqrt{\tau}\beta_{g,\gamma}}{4\sqrt{n}} + \frac{\sqrt{\tau}\beta_{g,\gamma}}{4\sqrt{n}} \leq \sqrt{\tau}\hat{\delta}_n,\end{aligned}$$

utilizing Lemma 5.2 ($\|(\mathcal{K}^*\mathcal{K})^{\frac{\gamma-1}{2}} S_{c_n} q\| = o(n^{-1/2})$), and this means that, on the set $\Omega_n \cap \Omega_n^1$, one has $\alpha_n \geq c_n$. Having control over the size of the DP sequence, one can write

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

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

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

and $\alpha_n\sqrt{n} \geq c_n\sqrt{n} \to \infty$. □

Acknowledgements

The author would like to thank an anonymous referee for insightful remarks, which helped to improve the presentation of results.

References

Anderssen, R. S. and de Hoog, F. R. Abel integral equations. In *Numerical solution of integral equations*, volume 42 of *Math. Concepts Methods Sci. Engrg.*, pp. 373–410. Plenum, New York (1990). MR1067155.

Antoniadis, A. and Bigot, J. Poisson inverse problems. *Ann. Statist.*, **34** (5), 2132–2158 (2006). MR2291495.

Bardsley, J. M. and Goldes, J. Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation. *Inverse Problems*, **25** (9), 095005, 18 (2009). MR2540885.

Bertero, M., Boccacci, P., Talenti, G., Zanella, R., and Zanni, L. A discrepancy principle for Poisson data. *Inverse Problems*, **26** (10), 105004, 20 (2010). MR2679468.

Bissantz, N., Hohage, T., Munk, A., and Ruymgaart, F. Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numer. Anal.*, **45** (6), 2610–2636 (2007). MR2361904.

Blanchard, G., Hoffmann, M., and Reiß, M. Early stopping for statistical inverse problems via truncated SVD estimation. *Electron. J. Stat.*, **12** (2), 3204–3231 (2018a). MR3859376.

Blanchard, G., Hoffmann, M., and Reiß, M. Optimal adaptation for early stopping in statistical inverse problems. *SIAM/ASA J. Uncertain. Quantif.*, **6** (3), 1043–1075 (2018b). MR3829522.

Blanchard, G. and Mathé, P. Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. *Inverse Problems*, **28** (11), 115011, 23 (2012). MR2992966.

Cavalier, L. and Koo, J.-Y. Poisson intensity estimation for tomographic data using a wavelet shrinkage approach. *IEEE Trans. Inform. Theory*, **48** (10), 2794–2802 (2002). MR1930348.

Ćmiel, B. An adaptive wavelet shrinkage approach to the Spektor-Lord-Willis problem. *J. Multivariate Anal.*, **101** (6), 1458–1470 (2010). MR2609506.

Ćmiel, B. Poisson intensity estimation for the Spektor-Lord-Willis problem using a wavelet shrinkage approach. *J. Multivariate Anal.*, **112**, 194–206 (2012). MR2957296.

Ćmiel, B., Szkutnik, Z., and Wojdyła, J. Asymptotic confidence bands in the Spektor-Lord-Willis problem via kernel estimation of intensity derivative. *Electron. J. Stat.*, **12** (1), 194–223 (2018). MR3760886.

Dudek, A. and Szkutnik, Z. Minimax unfolding of the spheres size distribution from linear sections. *Statist. Sinica*, **18** (3), 1063–1080 (2008). MR2440404.

Engl, H. W., Hanke, M., and Neubauer, A. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht (1996). ISBN 0-7923-4157-0. MR1408680.

Harrach, B., Jahn, T., and Potthast, R. Beyond the Bakushinkii veto: regularising linear inverse problems without knowing the noise distribution. *Numer. Math.*, **145** (3), 581–603 (2020). MR4119042.

Hoffmann-Jørgensen, J. and Pisier, G. The law of large numbers and the central limit theorem in Banach spaces. *Ann. Probability*, **4** (4), 587–599 (1976). MR423451.

Hohage, T. and Werner, F. Inverse problems with Poisson data: statistical regularization theory, applications and algorithms. *Inverse Problems*, **32** (9), 093001, 56 (2016). MR3543331.

Jahn, T. A modified discrepancy principle to attain optimal convergence rates under unknown noise. *Inverse Problems*, **37** (9), Paper No. 095008, 23 (2021). MR4304519.

Jahn, T. Optimal convergence of the discrepancy principle for polynomially and exponentially ill-posed operators under white noise. *Numer. Funct. Anal. Optim.*, **43** (2), 145–167 (2022). MR4413866.

Johnstone, I. M. and Silverman, B. W. Discretization effects in statistical inverse problems. *J. Complexity*, **7** (1), 1–34 (1991). MR1096170.

Kroll, M. Nonparametric intensity estimation from noisy observations of a Poisson process under unknown error distribution. *Metrika*, **82** (8), 961–990 (2019). MR4018433.

Last, G. and Penrose, M. *Lectures on the Poisson process*, volume 7 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, Cambridge (2018). ISBN 978-1-107-45843-7; 978-1-107-08801-6. MR3791470.

Ledoux, M. and Talagrand, M. *Probability in Banach spaces. Isoperimetry and processes.* Classics in Mathematics. Springer-Verlag, Berlin (2011). ISBN 978-3-642-20211-7. MR2814399.

Mair, B. A. and Ruymgaart, F. H. Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, **56** (5), 1424–1444 (1996). MR1409127.

Mika, G. and Szkutnik, Z. Towards adaptivity via a new discrepancy principle for Poisson inverse problems. *Electron. J. Stat.*, **15** (1), 2029–2059 (2021). MR4255321.

Reiss, R.-D. *A course on point processes.* Springer Series in Statistics. Springer-Verlag, New York (1993). ISBN 0-387-97924-7. MR1199815.

Sixou, B., Hohweiller, T., and Ducros, N. Morozov principle for Kullback-Leibler residual term and Poisson noise. *Inverse Probl. Imaging*, **12** (3), 607–634 (2018). MR3810172.

Stankewitz, B. Smoothed residual stopping for statistical inverse problems via truncated SVD estimation. *Electron. J. Stat.*, **14** (2), 3396–3428 (2020). MR4152147.

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

Szkutnik, Z. B-splines and discretization in an inverse problem for Poisson processes. *J. Multivariate Anal.*, **93** (1), 198–221 (2005). MR2119771.

Szkutnik, Z. Unfolding spheres size distribution from linear sections with B-splines and EMDS algorithm. *Opuscula Math.*, **27** (1), 151–165 (2007). MR2320844.

Vakhania, N. N., Tarieladze, V. I., and Chobanyan, S. A. *Probability distributions on Banach spaces*, volume 14 of *Mathematics and its Applications (Soviet Series)*. D. Reidel Publishing Co., Dordrecht (1987). ISBN 90-277-2496-2. MR1435288.

Werner, F. and Hohage, T. Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data. *Inverse Problems*, **28** (10), 104004, 15 (2012). MR2987899.

Zanella, R., Boccacci, P., Zanni, L., and Bertero, M. Efficient gradient projection methods for edge-preserving removal of Poisson noise. *Inverse Problems*, **25** (4), 045010, 24 (2009). MR2482161.