Information bounds for inverse problems with application to deconvolution and Lévy models

Mathias Trabs*†

Humboldt-Universität zu Berlin

Abstract

If a functional in a nonparametric inverse problem can be estimated with parametric rate, then the minimax rate gives no information about the ill-posedness of the problem. To have a more precise lower bound, we study semiparametric efficiency in the sense of Hájek–Le Cam for functional estimation in regular indirect models. These are characterized as models that can be locally approximated by a linear white noise model that is described by the generalized score operator. A convolution theorem for regular indirect models is proved. This applies to a large class of statistical inverse problems, which is illustrated for the prototypical white noise and deconvolution model. It is especially useful for nonlinear models. We discuss in detail a nonlinear model of deconvolution type where a Lévy process is observed at low frequency, concluding an information bound for the estimation of linear functionals of the jump measure.

Keywords: Convolution theorem • Deconvolution • Lévy process • Nonlinear ill-posed inverse problem • Semiparametric efficiency • White noise model

MSC (2000): 60G51 • 60J75 • 62B15 • 62G20 • 62M05

1 Introduction

Inverse problems are a key topic in applied mathematics, in particular models with noise in the data. Typically the parameter which is the target of the statistical inference is not directly observable, but "hidden" by some operator. While upper bounds, like convergence rates for nonparametric inverse problems, are mainly properties of the estimators, lower bounds reveal the deeper information theoretic structure. Instead of the (infinite dimensional) parameter itself, derived quantities are often the final object of interest. On the one hand, they might allow for inference with parametric rate, circumventing typical problems in nonparametric estimation like the choice of the bandwidth, cf. estimating the

*E-mail address: trabs@math.hu-berlin.de

†The author thanks Markus Reiß for many comments and stimulating questions and Nicolas Perkowski, Richard Nickl and Jakob Söhl for helpful remarks. This research was partially supported by the Deutsche Forschungsgemeinschaft through the FOR 1735 "Structural Inference in Statistics".

distribution function instead of the density. In this case minimax convergence rates give no information about the ill-posedness of the problem and we need the much more precise information bounds. On the other hand, many nonparametric statistical procedures rely on basis expansions and model selection strategies, see e.g. Cavalier et al. (2002). For these adaptive methods it is strictly necessary to assess the quality of the estimated coefficient in terms of confidence.

While inverse problems appear in many different shapes in the literature, information bounds are studied only in a few linear cases: Klaassen et al. (2001, 2005) and Khoujmane et al. (2007) considered the linear indirect regression model

$$Y_i = (K\vartheta)(X_i) + \xi_i, \quad i = 1, \dots, n,$$

where the regression function depends on the unknown parameter ϑ via the linear operator K and $(X_1, Y_1), \dots, (X_n, Y_n)$ are observed in presence of the additional errors ξ_i. van Rooij et al. (1999) derived a convolution theorem for linear indirect density estimation, where a sample of i.i.d. random variables $Y_1, \dots, Y_n$ with distribution $K\vartheta$ is observed. If more specifically

$$Y_i = X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where X_i has law ϑ and is corrupted by the noise variable ε_i, K is a convolution operator. Efficiency for this so called deconvolution model was considered by Söhl and Trabs (2012).

Using the polar decomposition or specific properties of the operators, all these studies are restricted to linear models. However, in many situations the operator K might not be linear, see e.g. Engl et al. (1996) and Bissantz et al. (2004). Hence, new mathematical methods are necessary. The aim of the present paper is twofold: (a) to provide a convolution theorem for general inverse problems that are regular in a well specified sense and (b) to study concrete and prototypical examples of linear and nonlinear structure.

A canonical probabilistic and nonlinear inverse problem is the following. Let Y_i be compound Poisson distributed

$$Y_i \sim e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \vartheta^{*k}$$

with intensity $\lambda > 0$ and jump distribution ϑ, writing ϑ^{*k} for the k-fold convolution of the measure ϑ. The distribution of Y_i is a convolution exponential and therefore not linear in ϑ. If Y_i is more generally an increment of a Lévy process $(L_t)_{t \geqslant 0}$, inference on the characteristic triplet of the Lévy process is a nonlinear problem since the dependence of the probability distribution of the marginals on the Lévy triplet is determined by the characteristic exponent, see the review by Reiß (2013). At the same time this model is of practical importance since Lévy processes are the main building blocks for mathematical modeling of stochastic processes. In the related context of diffusion processes, efficient estimation was recently studied by Clément et al. (2013).

In view of the equivalence results by Brown and Low (1996) and Nussbaum (1996), the prototype of an inverse problem is to estimate $\vartheta \in \Theta \subseteq X$, or derived parameters, from observations $y_{\varepsilon,\vartheta}$ in the white noise model

$$y_{\varepsilon,\vartheta} = K(\vartheta) + \varepsilon \dot{W} \quad \text{for a continuous operator } K : X \supseteq \Theta \to Y, \tag{1.1}$$

where X and Y are Hilbert spaces and $\varepsilon\dot{W}$ denotes white noise on Y with noise level $\varepsilon > 0$. For a review of estimation results in this model we refer to Cavalier (2008) and references therein. Studying minimax convergence rates when K is linear, Goldenshluger and Pereverzev (2000, 2003) have shown that the parametric rate ε can be achieved for linear functionals of ϑ whose smoothness is not smaller than the ill-posedness of the operator K.

Inspired by the results by van der Vaart (1991), we restate the classical local asymptotic normality (LAN) theory in a way that is appropriate to capture the inverse structure of the above mentioned models. Here, the linear white noise model (1.1) serves as the local limit experiment in the sense of Le Cam (1972). This leads to the notion of *regular indirect models*, meaning that the white noise model is the locally linear weak approximation of the statistical experiment. The asymptotic linear structure is described by the so called *generalized score operator*. We derive a version of the Hájek–Le Cam convolution theorem for the estimation of derived parameters for regular inverse problems. The tangent set is determined by the range of the generalized score operator and the efficient influence function is given by the Moore–Penrose pseudoinverse of the adjoint score operator. Although we focus on linear functionals in the examples, the theory applies to any parameter which is differentiable in a pathwise sense.

We show that the white noise model with a (possibly) nonlinear operator, the deconvolution model as well as the Lévy model are regular indirect models and thus the convolution theorem applies. In many cases estimators are known that have the optimal limit distribution and consequently the information bound is sharp. The analysis of Lévy processes is most challenging and the second half of this article is devoted to this model. Here, the proofs rely on estimates of the distance of infinitely divisible distributions by Liese (1987) and the Fourier multiplier approach which was introduced by Nickl and Reiß (2012).

We will put some stress on the Lévy model for three reasons: First, it is an important paradigm for nonlinear problems in indirect density estimation. Second, to understand from an efficiency perspective the *auto-deconvolution* structure of the Lévy model which was first reported by Belomestny and Reiß (2006). Third, to answer a conjecture by Nickl and Reiß (2012). Based on low frequency observations of a Lévy process, they have constructed an estimator for the (generalized) distribution function of the jump measure ν and proved asymptotic normality when the parametric rate can be attained. The natural question is whether this estimator is efficient in the Hájek–Le Cam sense. Since Buchmann and Grübel (2003) have constructed for a finite and known jump activity a decompounding estimator with smaller asymptotic variance, an information bound is of particular interest. With the general convolution theorem at hand, we can prove that both estimators are indeed efficient and thus prior knowledge of the jump intensity simplifies the statistical problem significantly. Concerning the information bound in the deconvolution setting, we can relax the assumptions on the functionals and the admissible error densities by van Rooij et al. (1999) and the assumptions on the smoothness and decay behavior of the densities of X_i and ε_i by Söhl and Trabs (2012) substantially. In fact, our abstract approach leads to natural assumptions in the explicit models.

This paper is organized as follows: Starting with the linear white noise model, we develop our general results in Section 2. These are illustrated in the deconvolution setup in Section 3. The theory will be applied to the Lévy model in Section 4. While the previous sections are restricted to $\mathbb{R}^d$-valued functionals, we

discuss the extension to general derived parameters in Section 5. More technical proofs are postponed to Section 6.

2 Regular indirect models

2.1 Linear white noise model

To understand the probabilistic structure of general inverse problems, we start with studying the abstract linear white noise model (1.1), where X and Y are separable real Hilbert spaces with scalar products $\langle \bullet, \bullet \rangle_X$ and $\langle \bullet, \bullet \rangle_Y$, respectively, and $K : X \to Y$ is a linear and bounded operator. To avoid identifiability problems, we additionally assume that K is injective. That is we observe for some unknown $\vartheta \in X$

$$\langle y_{\varepsilon,\vartheta}, \varphi \rangle_Y = \langle K\vartheta, \varphi \rangle_Y + \varepsilon \dot{W}(\varphi) \quad \text{for all} \quad \varphi \in Y$$

where $(\dot{W}(\varphi))_{\varphi \in Y}$ is an iso-normal Gaussian process with mean zero and covariance structure $\mathbb{E}[\dot{W}(\varphi_1)\dot{W}(\varphi_2)] = \langle \varphi_1, \varphi_2 \rangle_Y$ for $\varphi_1, \varphi_2 \in Y$. The law μ of the white noise $\dot{W}$ is defined as symmetric (zero mean) Gaussian measure on $(E, \mathscr{B}(E))$ for a separable Banach space E in which Y can be continuously embedded and where $\mathscr{B}(E)$ denotes the Borel σ-algebra on E. In other words $\dot{W}$ is an isometry from Y into $L^2(E, \mathscr{B}(E), \mu)$. For the construction of the so called abstract Wiener space we refer to Kuo (1975, Thm. 4.1, Lem. 4.7). We denote the law of $y_{\varepsilon,\vartheta}$ by $P_{\varepsilon,\vartheta}$.

Basically, the linear white noise model is a Gaussian shift experiment where the parameter is hidden behind the operator K. The inverse problem is to estimate a derived parameter $\chi(\vartheta)$ from the observation $y_{\varepsilon,\vartheta}$ when $\varepsilon \to 0$. First, let us focus on a linear functional $\chi(\vartheta) = \langle \zeta, \vartheta \rangle_X$ for some $\zeta \in X$. Typically, K is injective but admits no continuous inverse, leading to an ill-posed problem, cf. Goldenshluger and Pereverzev (2000, 2003) or Cavalier (2008) for a recent review of nonparametric estimation.

Following the classical semiparametric approach, we study parametric submodels by perturbing the parameter ϑ in directions $b \in X$. For any $b \in X$ we consider the submodel $t \mapsto P_{\varepsilon,\vartheta_t}$ generated by the path $[0,1) \ni t \to \vartheta_t := \vartheta + tb$. The behavior of the submodel along this path is described by the following lemma.

Lemma 2.1. *Let $P_{\varepsilon,x}$ denote the law of $y_{\varepsilon,x} = K(x) + \varepsilon \dot{W}$ on $(E, \mathscr{B}(E))$ for $x \in X$, and an operator $K : X \to Y$ with $K(0) = 0$, then for all $\vartheta \in X$*

$$\frac{\mathrm{d}P_{\varepsilon,x}}{\mathrm{d}P_{\varepsilon,\vartheta}}(y_{\varepsilon,\vartheta}) = \exp\Big(\dot{W}\Big(\frac{K(x) - K(\vartheta)}{\varepsilon}\Big) - \frac{1}{2\varepsilon^2}\|K(x) - K(\vartheta)\|_Y^2\Big) \quad P_{\varepsilon,\vartheta} - a.s.$$

The proof of this lemma relies on the Cameron–Martin formula for Gaussian measures on Banach spaces (Da Prato and Zabczyk, 1992, Prop. 2.24) and is postponed to Section 6.1. Linearity of K yields $\varepsilon^{-1}(K(\vartheta_\varepsilon) - K(\vartheta)) = Kb$ and thus by Lemma 2.1

$$\log \frac{\mathrm{d}P_{\varepsilon,\vartheta_\varepsilon}}{\mathrm{d}P_{\varepsilon,\vartheta}}(y_{\varepsilon,\vartheta}) = \dot{W}(Kb) - \frac{1}{2}\|Kb\|_Y^2 \quad P_{\varepsilon,\vartheta} - a.s. \tag{2.1}$$

and therefore model (1.1) with linear operator K satisfies the classical LAN condition (even without local and asymptotic) with parameter $h = Kb \in \operatorname{ran} K$. To find an information bound for the derived parameter $\chi(\vartheta) = \langle \zeta, \vartheta \rangle_X$, we express it in terms of $\eta = K\vartheta$ by

$$\psi(\eta) := \langle \zeta, K^{-1}\eta \rangle_X = \chi(\vartheta).$$

Since K^{-1} is typically not continuous, ψ will not be continuous along the path $t \mapsto K\vartheta_t = \eta + tKb$ without further assumptions. Supposing however $\zeta \in \operatorname{ran} K^\star$, where $\operatorname{ran} K^\star$ denotes the range of the adjoint operator $K^\star$, continuity and linearity of ψ follows from

$$\psi(\eta) = \langle K^\star y, K^{-1}\eta \rangle_X = \langle y, \eta \rangle_Y \quad \text{for any } y \in (K^\star)^{-1}(\{\zeta\}).$$

In fact, we will see below that the condition $\zeta \in \operatorname{ran} K^\star$ is equivalent to the regularity of ψ. If $K^\star$ is not injective there are many solutions y of the equation $K^\star y = \zeta$. The unique solution with minimal norm is given by the Moore–Penrose pseudoinverse

$$(K^\star)^\dagger \zeta := (K^\star|_{(\ker K^\star)^\perp})^{-1}(\zeta) = \Pi_{(\ker K^\star)^\perp}(K^\star)^{-1}(\{\zeta\}) \quad \text{for} \quad \zeta \in \operatorname{ran} K^\star, \tag{2.2}$$

where $\Pi_{(\ker K^\star)^\perp}$ denotes the orthogonal projection onto the orthogonal complement $(\ker K^\star)^\perp$ of the kernel of $K^\star$, cf. Engl et al. (1996, Def. 2.2 and Prop. 2.3) for the definition and fundamental properties of the pseudoinverse.

Given the regularity of the parameter, an LAN version of the Hájek–Le Cam convolution theorem (see van der Vaart and Wellner, 1996, Thm. 3.11.2) yields that the variance of any regular estimator is bounded from below by

$$\|\Pi_{\overline{\operatorname{ran}} K}(K^\star)^{-1}(\{\zeta\})\|_Y^2 = \|(K^\star)^\dagger \zeta\|_Y^2 \quad \text{if} \quad \zeta \in \operatorname{ran} K^\star, \tag{2.3}$$

since the closure of the range of K satisfies $\overline{\operatorname{ran}} K = (\ker K^\star)^\perp$.

Remark 2.2. Suppose that the operator K is injective and compact and denote the domain of K by $\operatorname{dom} K$. Then K is adapted to the Hilbert scale $(\operatorname{dom}(K^\star K)^{-\alpha})_{\alpha \geqslant 0}$ generated by $(K^\star K)^{-1}$ and its degree of ill-posedness is $\alpha = 1/2$ (cf. Natterer, 1984). According to Goldenshluger and Pereverzev (2000), the parameter $\chi(\vartheta) = \langle \zeta, \vartheta \rangle_X$ can be estimated with parametric rate ε if and only if $\vartheta \in \operatorname{ran}((K^\star K)^{1/2})$. Noting that $\operatorname{ran} K^\star = \operatorname{ran}((K^\star K)^{1/2})$ (Engl et al., 1996, Prop. 2.18), we recover the condition $\zeta \in \operatorname{ran} K^\star$. Since the existence of a regular estimator implies in particular that $\chi(\vartheta)$ can be estimated with rate ε, this condition is natural for stating a convolution theorem.

Remark 2.3. The information bound in (2.3) is sharp. Usually, regularization methods are necessary to construct estimators in ill-posed problems because $K^\dagger$ is unbounded and the observation $y_{\varepsilon,\vartheta}$ may not be in its domain. Assuming $\zeta \in \operatorname{ran} K^\star$, we can however define the estimator $\widehat{\chi}(\vartheta) := \langle (K^\star)^\dagger \zeta, y_{\varepsilon,\vartheta} \rangle_Y$ which satisfies

$$\begin{aligned}\widehat{\chi}(\vartheta) - \chi(\zeta) &= \langle (K^\dagger)^\star \zeta, K\vartheta \rangle_Y - \langle \zeta, \vartheta \rangle_X + \varepsilon \dot{W}((K^\dagger)^\star \zeta) \\ &= \langle \zeta, (K^\dagger K - \mathrm{Id})\vartheta \rangle_X + \varepsilon \dot{W}((K^\dagger)^\star \zeta) \sim \mathcal{N}(0, \varepsilon^2 \|(K^\dagger)^\star \zeta\|_Y^2)\end{aligned}$$

where we used $K^\dagger K = \Pi_{(\ker K)^\perp} = \mathrm{Id}$ because K is assumed to be injective. Therefore, the estimator $\widehat{\chi}(\vartheta)$ is efficient.

If the operator K in model (1.1) is nonlinear, the situation is more involved and a naive approach may fail as the following example illustrates. We define the Fourier transform of a function $f \in L^1(\mathbb{R}) \cup L^2(\mathbb{R})$ as $\mathcal{F} f(u) := \int e^{iux} f(x)\,\mathrm{d}x$.

Example 2.4. For a given $\vartheta \in \Theta := \{f \in L^2(\mathbb{R}) | f \geqslant 0\} \subseteq X := L^2(\mathbb{R})$ consider the linear differential equation in f

$$f' = -f + \vartheta^2 \quad \text{with} \quad \lim_{t \to -\infty} f(t) = 0, \tag{2.4}$$

which has the explicit solution $f_\vartheta(t) := \int_{-\infty}^t e^{-(t-s)} \vartheta^2(s)\,\mathrm{d}s$. The inverse problem is to estimate a linear functional $\chi(\vartheta) = \langle \vartheta, \zeta \rangle_X, \zeta \in X$, given an observation of the solution $f_\vartheta \in Y := L^2(\mathbb{R})$ of the previous equation corrupted by white noise. Since (2.4) is equivalent to $\mathcal{F}[\vartheta^2] = \mathcal{F}[f + f'] = (1 - iu)\,\mathcal{F} f$, the operator $K : \Theta \to L^2(\mathbb{R})$, which maps ϑ to the solution f_ϑ, can be written as

$$K(\vartheta) = \mathcal{F}^{-1}\left[(1 - iu)^{-1}\,\mathcal{F}[\vartheta^2](u)\right].$$

Note that K is well defined because $\|(1 - i\bullet)^{-1}\,\mathcal{F}[\vartheta^2]\|_{L^2} \leqslant \|(1 - i\bullet)^{-1}\|_{L^2(\mathbb{R})} \|\vartheta\|^2_{L^2(\mathbb{R})}$. We see immediately that K is nonlinear and injective on Θ. Due to the derivative in (2.4), ϑ does not depend continuously on the data f_ϑ and thus the problem is ill-posed.

Following the strategy of the linear model, we introduce the direct parameter $\eta = K(\vartheta)$ and write $\psi(\eta) = \langle \zeta, K^{-1}(\eta) \rangle_X = \chi(\vartheta)$. Note that ψ is nonlinear in η. To study pathwise continuity of ψ, we consider the path $[0, 1) \ni t \mapsto \eta_t = \eta + th$ with direction $h = K(b), b \in \Theta$. Note that $\eta_t \in \operatorname{ran} K$ since

$$\left(\eta_t + \eta_t'\right)^{1/2} = \left(\eta + \eta' + t(h + h')\right)^{1/2} = \left(\vartheta^2 + tb^2\right)^{1/2} \in \Theta.$$

For some intermediate point $\xi \in [0, t]$ the mean value theorem yields

$$\begin{aligned} t^{-1}\big(\psi(\eta_t) - \psi(\eta)\big) &= t^{-1}\langle K^{-1}(\eta_t) - K^{-1}(\eta), \zeta \rangle_X \\ &= \langle \tfrac{1}{2}(\eta + \eta')^{-1/2}(h + h'), \zeta \rangle_X - t\langle \tfrac{1}{4}(\eta_\xi + \eta_\xi')^{-3/2}(h + h')^2, \zeta \rangle_X. \end{aligned} \tag{2.5}$$

The first term is the linearization $\dot{\psi}_\eta(h) = \frac{1}{2}\langle (\eta + \eta')^{-1/2}(h + h'), \zeta \rangle_X = \frac{1}{2}\langle \vartheta^{-1} b^2, \zeta \rangle_X$ where we have to impose suitable conditions on ζ first to compensate the potentially non-integrable singularities of ϑ^{-1} and second to ensure continuity in h. But even if these conditions are satisfied, pathwise continuity of ψ may fail because the integrability problems in the remainder in (2.5) are more serious because b^4 is not integrable for every $b \in X$ and the singularities of $(\vartheta^2 + \xi b^2)^{-3/2}$ are more restrictive.

What went wrong in Example 2.4? Regularity of the parameter ψ depends on two properties: (i) the choice of ζ and (ii) the directions and paths along which we want to show the regularity. In particular the second point has to capture the inverse structure of the problem. The approach in the following section provides a solution to both problems. It gives a clear condition on ζ and it determines appropriate perturbations of the parameter, described by the tangent space.

2.2 Local linear weak approximation

Turning to a much more general model, the following definition will ensure that it behaves locally like the model (1.1) with a linear operator. Let Θ be a

parameter set such that for any $\vartheta \in \Theta$ there is a *tangent set* $\dot{\Theta}_\vartheta$ that is a subset of a Hilbert space with scalar product $\langle \bullet, \bullet \rangle_\vartheta$ such that any element $b \in \dot{\Theta}_\vartheta$ is associated to a path $[0, \tau) \ni t \mapsto \vartheta_t \in \Theta$ starting at ϑ and for some $\tau > 0$. For the sake of brevity we suppress the dependence of the path on b in the notation. In the following $Y_n \overset{P_n}{\Rightarrow} Y$ denotes weak convergence of the law of Y_n under the measure P_n to the law of Y for random variables $Y_1, Y_2, \dots$ and Y.

Definition 2.5. The sequence of statistical experiments $(\mathcal{X}_n, \mathcal{A}_n, P_{n,\vartheta} : \vartheta \in \Theta)$ is called a *locally regular indirect model* at $\vartheta \in \Theta$ with respect to the tangent set $\dot{\Theta}_\vartheta$ if there are a Hilbert space $(H_\vartheta, \langle \bullet, \bullet \rangle_H)$ and a continuous linear operator

$$A_\vartheta : \operatorname{lin} \dot{\Theta}_\vartheta \to H_\vartheta$$

such that for some rate $r_n \downarrow 0$ and for every $b \in \dot{\Theta}_\vartheta$ with associated path $t \mapsto \vartheta_t$ there are random variables $(G_n(h))_{h \in \operatorname{ran} A_\vartheta}$ satisfying

$$\log \frac{\mathrm{d}P_{n,\vartheta_{r_n}}}{\mathrm{d}P_{n,\vartheta}} = G_n(A_\vartheta b) - \frac{1}{2}\|A_\vartheta b\|_H^2 \quad \text{and} \tag{2.6}$$

$$\big(G_n(h_1), \dots, G_n(h_k)\big) \overset{P_{n,\vartheta}}{\Longrightarrow} \big(G(h_1), \dots, G(h_k)\big) \text{ for all } k \in \mathbb{N}, h_1, \dots h_k \in \operatorname{ran} A_\vartheta$$

for a centered Gaussian process $(G(h))_{h \in \operatorname{ran} A_\vartheta}$ with covariance $\mathbb{E}[G(h_1)G(h_2)] = \langle h_1, h_2 \rangle_H$. The operator A_ϑ is called *generalized score operator*.

In the sequel we will use the notation

$$\mathcal{P}_n := \{P_{n,\vartheta} | \vartheta \in \Theta\}.$$

The statistical interpretation of this regularity becomes clear by comparing it to the likelihood ratio (2.1) in the linear white noise model. Condition (2.6) means that locally at ϑ the model $(P_{n,\vartheta_{r_n}})$ converges to a limit experiment which is a linear inverse problem (1.1) in white noise with operator $K = A_\vartheta$ on the Hilbert space H_ϑ and with noise level $\varepsilon_n = r_n$. In other words at ϑ, the model converges weakly to the linear inverse problem in the sense of Le Cam (1972). Therefore, the classical white noise model (1.1) serves as a *locally linear weak approximation* of the general model $\mathcal{P}_n$. The difference to the classical theory is that the limit experiment is not a direct Gaussian shift experiment, but an indirect Gaussian shift, preserving the inverse structure of the problem. In that sense property (2.6) generalizes the classical local asymptotic normality, which corresponds to the identity operator $A_\vartheta = \mathrm{Id}$, to *local asymptotic indirect normality (LAIN)*.

The derived parameter $\chi : \Theta \to \mathbb{R}^d$, which is the aim of the statistical inference, should then be regular in the following sense.

Definition 2.6. The function $\chi : \Theta \mapsto \mathbb{R}^d, d \in \mathbb{N}$, is *pathwise differentiable* at $\vartheta \in \Theta$ with respect to the tangent set $\dot{\Theta}_\vartheta$ if there is a continuous linear operator $\dot{\chi}_\vartheta : \operatorname{lin} \dot{\Theta}_\vartheta \to \mathbb{R}^d$ such that for every $b \in \dot{\Theta}_\vartheta$ with associated path $[0, \tau) \ni t \mapsto \vartheta_t \in \Theta$ it holds

$$\frac{1}{t}\big(\chi(\vartheta_t) - \chi(\vartheta)\big) \to \dot{\chi}_\vartheta b \quad \text{as } t \downarrow 0.$$

By the Riesz representation theorem we can write $\dot{\chi}_\vartheta b = \langle \widetilde{\chi}_\vartheta, b \rangle_\vartheta$ for all $b \in \dot{\Theta}_\vartheta$ and some gradient $\widetilde{\chi}_\vartheta \in \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$. Recall that the sequence of parameter

functions $\psi_n : \mathcal{P}_n \to \mathbb{R}^d$ given by $\psi_n(P_{n,\vartheta}) = \chi(\vartheta)$ is called regular at ϑ relative to $A_\vartheta \dot{\Theta}_\vartheta$ if for any $h \in A_\vartheta \dot{\Theta}_\vartheta$ and any submodel $t \mapsto P_{n,\vartheta_t}$ satisfying (2.6) with $h = A_\vartheta b$ for some $b \in \dot{\Theta}_\vartheta$, it holds

$$\frac{\psi_n(P_{n,\vartheta_{r_n}}) - \psi_n(P_{n,\vartheta})}{r_n} \to \dot{\psi}_\vartheta(h) \tag{2.7}$$

for some continuous linear map $\dot{\psi}_\vartheta : H \to \mathbb{R}^d$. Again the Riesz representation theorem determines a unique $\widetilde{\psi}_\vartheta \in \overline{\operatorname{ran}} A_\vartheta = \overline{\operatorname{lin}} A_\vartheta \dot{\Theta}_\vartheta$ such that $\dot{\psi}_\vartheta(h) = \langle \widetilde{\psi}_\vartheta, h \rangle_H$ for all $h \in \operatorname{ran} A_\vartheta$. $\widetilde{\psi}_\vartheta$ is called *efficient influence function* in the classical semiparametric theory. As the last ingredient we recall that a sequence of estimators $T_n : \mathcal{X}_n \to \mathbb{R}^d$ is called *regular* at ϑ with respect to the rate r_n and relative to the directions $\dot{\Theta}_\vartheta$ if there is a limit distribution L on the Borel measurable space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ such that

$$\frac{1}{r_n}\big(T_n - \chi(\vartheta_{r_n})\big) \overset{P_{n,\vartheta_{r_n}}}{\Longrightarrow} L$$

for every $b \in \dot{\Theta}_\vartheta$ and any corresponding submodel $t \mapsto P_{n,\vartheta_t}$. We recall the definition (2.2) of the Moore–Penrose pseudoinverse $K^\dagger$ of an operator K on its range and obtain the following convolution theorem.

Theorem 2.7. *Let $(\mathcal{X}_n, \mathcal{A}_n, P_{n,\vartheta} : \vartheta \in \Theta)$ be a locally regular indirect model at $\vartheta \in \Theta$ and $\chi : \Theta \to \mathbb{R}^d$ be pathwise differentiable at ϑ with respect to $\dot{\Theta}_\vartheta$. Then the sequence $\psi_n : \mathcal{P}_n \to \mathbb{R}^d$ is regular at ϑ relative to $\dot{\Theta}_\vartheta$ if and only if each coordinate function of $\widetilde{\chi}_\vartheta = (\widetilde{\chi}_\vartheta^{(1)}, \dots, \widetilde{\chi}_\vartheta^{(d)})$ is contained in the range of the adjoint score operator $A_\vartheta^\star : H \to \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$.*

In this case the efficient influence function is given by $\widetilde{\psi}_\vartheta = (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta = ((A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta^{(1)}, \dots, (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta^{(d)})$ and for any regular estimator sequence $T_n : \mathcal{X}_n \to \mathbb{R}^d$ the limit distribution satisfies $L = \mathcal{N}(0, \Sigma) * M$ for some Borel probability distribution M and covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$ with

$$\Sigma_{k,l} = \big\langle \widetilde{\psi}_\vartheta^{(k)}, \widetilde{\psi}_\vartheta^{(l)} \big\rangle_H = \big\langle (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta^{(k)}, (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta^{(l)} \big\rangle_H, \quad k, l \in \{1, \dots, d\}. \tag{2.8}$$

Proof. The characterization of regular parameter functions can be proved analogously to the i.i.d. setting studied by van der Vaart (1991, Thm. 3.1). Let us first consider $d = 1$. For any $b \in \dot{\Theta}_\vartheta$ there is a path $[0, \tau) \ni t \mapsto \vartheta_t$ in direction b generating a submodel $t \mapsto P_{n,\vartheta_t}$ with $h = A_\vartheta b$. If ψ_n is regular,

$$\dot{\psi}_\vartheta(A_\vartheta b) = \lim_{n\to\infty} \frac{\psi_n(P_{n,\vartheta_{r_n}}) - \psi_n(P_{n,\vartheta})}{r_n} = \lim_{n\to\infty} \frac{\chi(\vartheta_{r_n}) - \chi(\vartheta)}{r_n} = \langle \widetilde{\chi}_\vartheta, b \rangle_\vartheta.$$

Since the equality $\langle \widetilde{\psi}_\vartheta, A_\vartheta b \rangle_H = \dot{\psi}_\vartheta(A_\vartheta b) = \langle \widetilde{\chi}_\vartheta, b \rangle_\vartheta$ holds for all $b \in \dot{\Theta}_\vartheta$, it follows $A_\vartheta^\star \widetilde{\psi}_\vartheta = \widetilde{\chi}_\vartheta$. To conclude the converse direction, we can use the previous display as definition of $\dot{\psi}_\vartheta$ and have to verify that it is indeed linear and continuous. But this follows because by assumption there is some $\psi \in H$ such that $A_\vartheta^\star \psi = \widetilde{\chi}_\vartheta$ and thus $\dot{\psi}_\vartheta(h) = \langle \psi, h \rangle_H$ for all $h \in H$. Because $\overline{\operatorname{ran}} A_\vartheta = (\ker A_\vartheta^\star)^\perp$, there is exactly one solution of $A_\vartheta^\star \psi = \widetilde{\chi}_\vartheta$ in $\overline{\operatorname{ran}} A_\vartheta$ and this is given by $(A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta = \Pi_{(\ker A_\vartheta^\star)^\perp} (A_\vartheta^\star)^{-1}(\{\widetilde{\chi}_\vartheta\})$. For $d > 1$ it is sufficient to consider the coordinate functions separately.

To conclude the second part of the theorem, we consider $A_\vartheta \dot{\Theta}_\vartheta$ as local parameter set and identify the local parameter $\kappa_n(A_\vartheta b) := \psi_n(P_{n,\vartheta_{r_n}}) = \chi(\vartheta_{r_n})$ with $\kappa_n(0) := \psi_n(P_{n,\vartheta})$. Then κ_n is regular relative to $A_\vartheta \dot{\Theta}_\vartheta$ and we can apply the convolution theorem in van der Vaart and Wellner (1996, Thm. 3.11.2). Hereby, we have to note that it is sufficient if the local parameter set $A_\vartheta \dot{\Theta}_\vartheta$ is only a subset of a Hilbert space, and thus (2.6) may not hold for all linear combinations of elements in $\dot{\Theta}_\vartheta$, as long as the weak convergence $G_n(h) \Rightarrow G(h)$ under $P_{n,\vartheta}$ holds true for all $h \in \operatorname{lin} A_\vartheta \dot{\Theta}_\vartheta$. □

The theorem implies immediately that the asymptotic covariance of every regular estimator of $\chi(\vartheta)$ is bounded from below by (2.8) in the order of nonnegative definite matrices. If $\widetilde{\chi}_\vartheta$ is contained in the smaller range of the *information operator* $A_\vartheta^\star A_\vartheta$, then the efficient influence function can be obtained by

$$\widetilde{\psi}_\vartheta = (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta = A_\vartheta (A_\vartheta^\star A_\vartheta)^\dagger \widetilde{\chi}_\vartheta,$$

owing to $\ker(A_\vartheta^\star A_\vartheta) = \ker A_\vartheta$. Therefore, in this case the hardest parametric subproblem is given by the direction $(A_\vartheta^\star A_\vartheta)^\dagger \widetilde{\chi}_\vartheta \in \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$. In the finite dimensional linear model this lower bound coincides with the minimal variance of the Gauß–Markov theorem. Let us illustrate Theorem 2.7 in several examples.

Example 2.8 (Indirect regression model)**.** With $X = Y = L^2(\mathbb{R})$ and a linear, bounded, injective operator $K : X \to Y$ consider the indirect regression model with deterministic design

$$Y_i = (Kf)\Big(\frac{i}{n}\Big) + \varepsilon_i, \quad i = 1, \dots, n, \quad \text{with unkown } f \in X$$

and with i.i.d. errors $\varepsilon_1, \dots, \varepsilon_n \sim \mu$ for some law μ. Therefore, $(Y_1, \dots, Y_n) \sim P_{n,f} = \Pi_{i=1}^n \mu(\bullet - (Kf)(i/n))$. Khoujmane et al. (2007) have proved a convolution theorem for estimating $\chi(f) = \langle \zeta, f \rangle_{L^2(\mathbb{R})}, \zeta \in X$ with $\|\zeta\|_{L^2(\mathbb{R})} = 1$, where f and the error distribution μ are unknown. Let us focus on the submodel with standard normal errors $\varepsilon_i \sim \mathcal{N}(0,1)$. Under smoothness conditions it is stated in this submodel (Khoujmane et al., 2007, Thm. 2 with $\delta = 0$) that the asymptotic distribution of any regular estimator is given by the convolution $\mathcal{N}(0, \|K\zeta\|_{L^2(\mathbb{R})}^{-2}) * M$ for some Borel probability measure M.

To apply Theorem 2.7, we have to check that this model is a regular indirect model. Choosing the tangent space $\dot{\Theta}_f = X$ with linear paths $f_t = f + tb$ in directions $b \in \dot{\Theta}_f$, we calculate

$$\begin{aligned}
&\log \frac{\mathrm{d}P_{n,f_{1/\sqrt{n}}}}{\mathrm{d}P_{n,f}}(Y_1, \dots, Y_n) \\
&\quad = -\frac{1}{2n} \sum_{i=1}^n (Kb)^2\Big(\frac{i}{n}\Big) + \frac{1}{n^{1/2}} \sum_{i=1}^n (Kb)\Big(\frac{i}{n}\Big)\Big(Y_i - (Kf)\Big(\frac{i}{n}\Big)\Big) \\
&\quad \overset{P_{n,f}}{\Longrightarrow} \mathcal{N}\Big(-\frac{1}{2}\|Kb\|_{L^2(\mathbb{R})}^2, \|Kb\|_{L^2(\mathbb{R})}^2\Big).
\end{aligned}$$

Therefore, the generalized score operator is given by $A_f = K$. Assuming for simplicity that $\zeta \in \operatorname{ran}(K^\star K)$, the asymptotic distribution of any regular estimator is given by a convolution $\mathcal{N}(0, \|K(K^\star K)^\dagger \zeta\|_{L^2(\mathbb{R})}^2) * M$. Since the Cauchy–Schwarz inequality yields $\|K\zeta\|_{L^2(\mathbb{R})} \|K(K^\star K)^\dagger \zeta\|_{L^2(\mathbb{R})} \geqslant \|\zeta\|_{L^2(\mathbb{R})}^2 = 1$, the

bound by Khoujmane et al. (2007) achieves our information bound if and only if $K^\star K = \lambda\,\mathrm{Id}$ for some $\lambda > 0$. Therefore, their information bound may not be optimal. The reason is that f has been perturbed in direction ζ instead of the the least favorable direction $(K^\star K)^\dagger \zeta$.

Example 2.9 (Nonlinear white noise model)**.** Suppose we observe $y_{n,\vartheta} = K(\vartheta) + \varepsilon_n \dot{W}$ with $\varepsilon_n \to 0$ as $n \to \infty$ on the Hilbert space Y for some $\vartheta \in X$ and for a not necessarily linear operator $K : X \supseteq \Theta \to Y$ with $K(0) = 0$ which is Gâteaux differentiable at the inner point $\vartheta \in \Theta$. That is there is a continuous linear operator $\dot{K}_\vartheta : X \to Y$ with

$$\lim_{t\to 0} \frac{1}{t}\big(K(\vartheta + tb) - K(\vartheta)\big) = \dot{K}_\vartheta b \quad \text{for all} \quad b \in X.$$

By the Hilbert space structure, the tangent space can be chosen as $\dot{\Theta}_\vartheta = X$ by considering the path $[0,1) \ni t \mapsto \vartheta_t := \vartheta + tb$ for $b \in \dot{\Theta}_\vartheta$. Lemma 2.1 yields for any $b \in X$ with associated path $t \mapsto \vartheta_t$

$$\log \frac{\mathrm{d}P_{n,\vartheta_{\varepsilon_n}}}{\mathrm{d}P_{n,\vartheta}}(y_{n,\vartheta}) = \dot{W}\Big(\frac{K(\vartheta + \varepsilon_n b) - K(\vartheta)}{\varepsilon_n}\Big) - \frac{1}{2\varepsilon_n^2}\|K(\vartheta + \varepsilon_n b) - K(\vartheta)\|_Y^2.$$

Therefore, the LAIN property (2.6) is satisfied with generalized score operator chosen as the Gâteaux derivative $A_\vartheta = \dot{K}_\vartheta$ at ϑ and

$$G_n(A_\vartheta b) = \dot{W}\Big(\frac{K(\vartheta + \varepsilon_n b) - K(\vartheta)}{\varepsilon_n}\Big) - \frac{1}{2}\Big(\Big\|\frac{K(\vartheta + \varepsilon_n b) - K(\vartheta)}{\varepsilon_n}\Big\|_Y^2 - \|\dot{K}_\vartheta b\|_Y^2\Big)$$

where $G_n(A_\vartheta b) \Rightarrow \mathcal{N}(0, \|\dot{K}_\vartheta b\|_Y^2)$, since the variance of the first term of G_n converge to $\|\dot{K}_\vartheta b\|_Y^2$ and the second term converges deterministically to zero by the Gâteaux differentiability. The convergence of the finite dimensional distributions follows likewise.

Along the path $t \mapsto \vartheta_t$ the linear functional $\chi(\vartheta) = \langle \zeta, \vartheta\rangle_X$ possesses the derivative

$$\lim_{t\to 0} \frac{1}{t}\big(\chi(\vartheta_t) - \chi(\vartheta)\big) = \langle \zeta, b\rangle_X.$$

Hence, $\dot{\chi}_\vartheta b = \langle \widetilde{\chi}_\vartheta, b\rangle_X$ for the gradient $\widetilde{\chi}_\vartheta = \zeta$ and all $b \in \dot{\Theta}_\vartheta$. Applying Theorem 2.7 shows in particular that the asymptotic variance of every regular sequence of estimators T_n (with respect to the rate ε_n) of the functional $\chi(\vartheta)$ is bounded from below by

$$\|(\dot{K}_\vartheta^\star)^\dagger \zeta\|_Y^2 \quad \text{whenever} \quad \zeta \in \operatorname{ran} \dot{K}_\vartheta^\star.$$

If K is a linear bounded operator the score operator is $A_\vartheta = \dot{K}_\vartheta = K$ and thus the statement of Theorem 2.7 coincides with the previous result (2.3).

In Remark 2.3 we saw that that this information bound can be achieved if K is linear. For nonlinear operators an upper bound is beyond the scope of this paper.

Example 2.10 (Example 2.4 continued)**.** Let us come back to the ill-posed inverse problem in Example 2.4 related to the differential equation (2.4). The corresponding nonlinear operator $K : X \to Y$ with $\Theta = \{f \in L^2(\mathbb{R}) | f \geqslant 0\} \subseteq X$ and $X = Y = L^2(\mathbb{R})$ was given by $K(\vartheta) = \mathcal{F}^{-1}[(1 - i\bullet)^{-1}\,\mathcal{F}[\vartheta^2]]$. For $\vartheta \in \Theta$

and any $b \in \dot{\Theta}_\vartheta := \Theta$ the functional $\chi(\vartheta) = \langle \zeta, \vartheta \rangle_X, \zeta \in L^2(\mathbb{R})$, is pathwise differentiable along the path $[0,1) \mapsto \vartheta_t = \vartheta + tb$ with gradient $\widetilde{\chi}_\vartheta = \zeta$. K is pathwise differentiable with respect to the tangent set $\dot{\Theta}_\vartheta$ at ϑ with derivative

$$\dot{K}_\vartheta b = \mathcal{F}^{-1}\left[(1-iu)^{-1}\,\mathcal{F}[2b\vartheta](u)\right].$$

Since $\dot{K}$ is well defined on $\operatorname{lin}\dot{\Theta}_\vartheta = L^2(\mathbb{R})$, the generalized score operator $A_\vartheta : \operatorname{lin}\dot{\Theta}_\vartheta \to H := L^2(\mathbb{R})$ is given by $A_\vartheta b = \dot{K}_\vartheta b$ as in the previous example. The "directions" in which we perturb the direct parameter $K(\vartheta)$ are then given by $A_\vartheta \dot{\Theta}_\Theta = \{K(\sqrt{\vartheta b}) | b \in X\}$. Applying Plancherel's identity twice, the adjoint of A_ϑ can be calculated via

$$\begin{aligned}\langle \dot{K}_\vartheta b, h\rangle &= \frac{1}{2\pi}\int (1-iu)^{-1}\,\mathcal{F}[2b\vartheta](u)\overline{\mathcal{F}\,h}(u)\,\mathrm{d}u \\ &= \langle b, 2\vartheta\,\mathcal{F}^{-1}\left[(1-iu)^{-1}\,\mathcal{F}\,h(-u)\right]\rangle_X,\end{aligned}$$

for $b, h \in L^2(\mathbb{R})$. Therefore, $A_\vartheta^\star h = 2\vartheta\,\mathcal{F}^{-1}[(1-iu)^{-1}\,\mathcal{F}\,h(-u)]$ and regularity of the parameter function follows for any $\zeta \in \operatorname{ran} A_\vartheta^\star = \{\vartheta f | f \in H^1(\mathbb{R})\}$ with the Sobolev space $H^1(\mathbb{R}) = \{f \in L^2(\mathbb{R}) | \int (1+u^2)|\,\mathcal{F}\,f(u)|^2\,\mathrm{d}u < \infty\}$.

As Example 2.10 indicates, the adjoint score operator $A_\vartheta^\star$ has usually no closed range. In these cases it is a difficult problem to determine the range of $A_\vartheta^\star$. As the following characterization shows, it is sufficient to know $A_\vartheta^\star$ on a dense subspace. This approximation argument will turn out to be very useful for more complex models.

Proposition 2.11. *Let $A_\vartheta^\star : H \to \overline{\operatorname{lin}}\dot{\Theta}_\vartheta$ be injective and let $\mathcal{G}$ be a dense subspace in H. Then $\widetilde{\chi}_\vartheta \in \overline{\operatorname{lin}}\dot{\Theta}_\vartheta$ is contained in $\operatorname{ran} A_\vartheta^\star$ if and only if the following is satisfied*

(i) there exists a sequence $\chi_n \in \operatorname{ran} A_\vartheta^\star|_{\mathcal{G}}$ such that $\chi_n \to \widetilde{\chi}_\vartheta$ as $n \to \infty$ and

(ii) $(A_\vartheta^\star)^{-1}\chi_n$ converges to some $\psi \in H$.

In this case $A_\vartheta^\star \psi = \widetilde{\chi}_\vartheta$ and thus $\Pi_{\overline{\operatorname{ran}} A_\vartheta}\psi = \psi$ is the efficient influence function.

Proof. "if": Since $A_\vartheta^\star$ is a bounded operator, its graph

$$\{(g, A_\vartheta^\star g) : g \in H\} \subseteq H \times \overline{\operatorname{lin}}\dot{\Theta}_\vartheta$$

is closed. Therefore, the inverse operator $(A_\vartheta^\star)^{-1}|_{\operatorname{ran} A_\vartheta^\star}$ is closed, too. Consequently, (i) and (ii) imply $\widetilde{\chi}_\vartheta \in \operatorname{dom}(A_\vartheta^\star)^{-1} = \operatorname{ran} A_\vartheta^\star$ with $(A_\vartheta^\star)^{-1}\widetilde{\chi}_\vartheta = \psi$.

"only if": Since $\widetilde{\chi}_\vartheta \in \operatorname{ran} A_\vartheta^\star$ there is some $\psi \in H$ such that $A_\vartheta^\star \psi = \widetilde{\chi}_\vartheta$. Moreover, there is a sequence $(g_n) \subseteq \mathcal{G}$ with $g_n \to \psi$ because $\mathcal{G}$ is dense in H. The continuity of $A_\vartheta^\star$ yields then $\chi_n := A_\vartheta^\star g_n \to A_\vartheta^\star \psi = \widetilde{\chi}_\vartheta$. □

2.3 I.i.d. observations

When the observations are given by n independent and identically distributed random variables $Y_1, \ldots, Y_n$, the model simplifies to the product space $(\mathcal{X}^n, \mathcal{A}^{\otimes n}, P_\vartheta^{\otimes n} : \vartheta \in \Theta)$ such that the probability measure is completely described by the family of marginal distributions $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$. We will

rephrase the conditions of the previous section in terms of the marginal measure P_ϑ. This setting appears quite often in applications and, in particular, the deconvolution model and the Lévy model which we study in Sections 3 and 4 will be two examples. That is why, we will give some details for the i.i.d. case.

Recall that a *tangent set* $\dot{\mathcal{P}}_{P_\vartheta}$ at P_ϑ is a set of score functions g of submodels $[0,\tau) \ni t \mapsto P_{\vartheta_t}$ starting at P_ϑ and for some $\tau > 0$. In the present situation the derived parameter can be written as $\psi(P_\vartheta) = \chi(\vartheta)$, independent of n. The classical Hajék–Le Cam convolution theorem (cf. Bickel et al., 1998, Thm. 3.3.2) applies if ψ is differentiable at P_ϑ relative to $\dot{\mathcal{P}}_{P_\vartheta}$, that is, there exists a continuous linear map $\dot{\psi} : L^2(P_\vartheta) \to \mathbb{R}^k$ such that

$$\lim_{t\to 0} \frac{\psi(P_{\vartheta_t}) - \psi(P_\vartheta)}{t} = \dot{\psi} g \quad \text{for all } g \in \dot{\mathcal{P}}_{P_\vartheta}.$$

This differentiability corresponds to the general assumption (2.7). In the i.i.d. setting local asymptotic normality follows from Hellinger regularity of the submodel $t \mapsto P_{\vartheta_t}$. Therefore, we can reformulate the conditions in Definition 2.5 to the following

Assumption A. At $\vartheta \in \Theta$ let the parameter set give rise to a tangent set $\dot{\Theta}_\vartheta$. Furthermore, let there be a continuous linear operator

$$A_\vartheta : \operatorname{lin} \dot{\Theta}_\vartheta \to L_0^2(P_\vartheta) := \Big\{ g \in L^2(P_\vartheta) : \int g \, \mathrm{d}P_\vartheta = 0 \Big\}$$

such that for every $b \in \dot{\Theta}_\vartheta$ with associated path $t \mapsto \vartheta_t$

$$\int \Big(\frac{\mathrm{d}P_{\vartheta_t}^{1/2} - \mathrm{d}P_\vartheta^{1/2}}{t} - \frac{1}{2} A_\vartheta b \, \mathrm{d}P_\vartheta^{1/2} \Big)^2 \to 0 \quad \text{as } t \downarrow 0. \tag{2.9}$$

In (2.9) $\mathrm{d}P_{\vartheta_t}$ denotes the Radon–Nikodym μ-density of P_{ϑ_t} for some dominating measure μ and the integration is with respect to μ. Since the integral does not depend on μ, it is suppressed in the notation.

Lemma 2.12. *If the product model $(\mathcal{X}^n, \mathcal{A}^{\otimes n}, P_\vartheta^{\otimes n} : \vartheta \in \Theta)$ satisfies Assumption A at $\vartheta \in \Theta$, then it is a locally regular indirect model at $\vartheta \in \Theta$ with respect to the tangent set $\dot{\Theta}_\vartheta$, with rate $r_n = n^{-1/2}$ and (generalized) score operator A_ϑ.*

Proof. The Hellinger regularity in Assumption A implies local asymptotic normality, since it yields, for instance, see Bickel et al. (1998, Prop. 2.1.2),

$$\sum_{j=1}^n \log \frac{\mathrm{d}P_{\vartheta_{1/\sqrt{n}}}}{\mathrm{d}P_\vartheta}(X_j) = \frac{1}{\sqrt{n}} \sum_{j=1}^n A_\vartheta b(X_j) - \frac{1}{2} \|A_\vartheta b\|_{L^2(P_\vartheta)} + R_n$$

for a remainder R_n that converges in $P_\vartheta^{\otimes n}$-probability to zero. Hence, the LAIN property (2.6) is statisfied with rate $1/\sqrt{n}$ and with the score operator A_ϑ mapping into the Hilbert space $H = L_0^2(P_\vartheta)$. The convergence of the finite dimensional distributions in Definition 2.5 follows from the Cramér–Wold device and the linearity of A_ϑ. □

Note that $L_0^2(P_\nu)$ is the orthogonal complement of $\operatorname{lin} 1$ and thus it is a closed subspace of the Hilbert space $L^2(P_\nu)$. The operator A_ϑ maps directions $b \in \dot{\Theta}_\vartheta$ into score functions at P_ϑ and thus it is called *score operator* which explains the name given in the general case. It generates the tangent set $\dot{\mathcal{P}}_{P_\vartheta} = A_\vartheta \dot{\Theta}_\vartheta$ of the model $\mathcal{P}$ at P_ϑ. Note that the range of A_ϑ is a subset of the maximal tangent set as the following example shows.

Example 2.13 (Maximal tangent set)**.** Let $\mathcal{P}$ be the model of all probability measures on some sample space. The maximal tangent set of the model $\mathcal{P}$ at some distribution P is given by $L_0^2(P)$. This can be seen as follows: Score functions are necessarily centered and square integrable. For any score $g \in L_0^2(P)$ a one-dimensional submodel is $t \to c(t)k\big(tg(x)\big)\,\mathrm{d}P(x)$ with a $C^2(\mathbb{R})$-function $k : \mathbb{R} \to \mathbb{R}_+$ which satisfies $k(0) = k'(0) = 1$ and such that k'/k is bounded and with normalization constant $c(t) = \|k(tb)\|_{L^1(\nu)}^{-1}$, for instance, $k(y) = 2/(1 + e^{-2y})$ (cf. van der Vaart, 1998, Ex. 25.16).

Theorem 2.7 yields then the following convolution theorem, which was already obtained by van der Vaart (1991).

Corollary 2.14. *Suppose the product model with marginal distributions $\mathcal{P} = \{P_\vartheta | \vartheta \in \Theta\}$ satisfies Assumption A and let $\chi : \Theta \to \mathbb{R}^d$ be pathwise differentiable with respect to $\dot{\Theta}_\vartheta$. The map $\psi : \mathcal{P} \to \mathbb{R}^d$ is differentiable at P_ϑ relative to the tangent set $\dot{\mathcal{P}}_{P_\vartheta} = A_\vartheta \dot{\Theta}_\vartheta$ if and only if each coordinate function of $\widetilde{\chi}_\vartheta$ is contained in the range of the adjoint score operator $A_\vartheta^\star : L_0^2(P_\vartheta) \to \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$. In this case the efficient influence function is given by $\widetilde{\psi}_\vartheta = (A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta$.*

In particular, for $\zeta \in \operatorname{ran} A_\vartheta^\star$ the asymptotic covariance matrix of every regular estimator is bounded from below by

$$\mathbb{E}_\vartheta \left[\widetilde{\psi}_{P_\vartheta} \widetilde{\psi}_{P_\vartheta}^\top\right] = \mathbb{E}_\vartheta \left[\big((A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta\big)\big((A_\vartheta^\star)^\dagger \widetilde{\chi}_\vartheta\big)^\top\right].$$

If $\widetilde{\chi}_\vartheta \notin \operatorname{ran} A_\vartheta^\star$, van der Vaart (1991) shows that there exists no regular estimator of the functional $\chi(\vartheta)$.

In the i.i.d. case we find the following statistical interpretation of Proposition 2.11, adopting the Cramér–Rao point of view. Let $\mathcal{G}$ be a dense subset of $L_0^2(P_\vartheta)$ and let $\chi(\vartheta)$ be a one-dimensional derived parameter with gradient $\widetilde{\chi}_\vartheta$. Consider an approximating sequence $\chi_n \to \widetilde{\chi}_\vartheta$ satisfying $\mathcal{G} \ni g_n := (A_\vartheta^\star)^\dagger \chi_n \to \widetilde{\psi}_{P_\vartheta}$. Assuming $\mathcal{G} \subseteq \operatorname{ran} A_\vartheta$, we can define $b_n := A_\vartheta^\dagger g_n = I^\dagger \chi_n$ where $I := A_\vartheta^\star A_\vartheta$ is the information operator. The information bound can be read as a Cramér–Rao bound in the least favorable submodel

$$\begin{aligned}\mathbb{E}[\widetilde{\psi}_{P_\vartheta}^2] &= \sup_{g \in \operatorname{lin} \dot{\mathcal{P}}_{P_\vartheta}} \frac{\langle \widetilde{\psi}_{P_\vartheta}, g\rangle_{P_\vartheta}^2}{\langle g, g\rangle_{P_\vartheta}} = \sup_{b \in \operatorname{lin} \dot{\Theta}_\vartheta} \frac{\langle \widetilde{\chi}_\vartheta, b\rangle_\vartheta^2}{\langle A_\vartheta b, A_\vartheta b\rangle_{P_\vartheta}} \\ &\geqslant \frac{\big(\langle \chi_n, b_n\rangle_\vartheta - \langle \widetilde{\chi}_\vartheta - \chi_n, b_n\rangle_\vartheta\big)^2}{\langle A_\vartheta b_n, A_\vartheta b_n\rangle_{P_\vartheta}}, \end{aligned} \tag{2.10}$$

where we plugged in the direction $b_n = I^\dagger \chi_n$. The term $\langle \chi_n, b_n\rangle_\vartheta^2 / \langle A_\vartheta b_n, A_\vartheta b_n\rangle_{P_\vartheta} = \langle g_n, g_n\rangle_{P_\vartheta}$ is the Cramér–Rao bound for the estimation problem of a functional, which approximates $\chi(\vartheta)$, with gradient χ_n. The approximation error $\langle \widetilde{\chi}_\vartheta - \chi_n, b_n\rangle_\vartheta$ should be understood as bias. Since

b_n does not have to be bounded, $\chi_n \to \widetilde{\chi}_\vartheta$ is not sufficient to conclude that the bias vanishes. However, Proposition 2.11(ii) implies that this error converges to zero owing to the Cauchy–Schwarz inequality:

$$\begin{aligned}|\langle \widetilde{\chi}_\vartheta - \chi_n, b_n\rangle_\vartheta| &= |\langle (A_\vartheta^\star)^\dagger(\widetilde{\chi}_\vartheta - \chi_n), A_\vartheta b_n\rangle_{P_\vartheta}| \\ &\leqslant \|\widetilde{\psi}_{P_\vartheta} - g_n\|_{P_\vartheta}\|g_n\|_{P_\vartheta} \to 0.\end{aligned}$$

Hence, the Cramér–Rao bound (2.10) converges to the information bound $\langle \widetilde{\psi}_{P_\vartheta}, \widetilde{\psi}_{P_\vartheta}\rangle_{P_\vartheta}$. A similar perspective was taken by Söhl and Trabs (2012, Lem. 3).

3 Deconvolution

Let us discuss the previous results in the classical nonparametric deconvolution setup, which has many applications, e.g., measurement-error problems (see Meister (2009)). We observe an i.i.d. sample

$$Y_i = X_i + \varepsilon_i, \quad i = 1, \dots, n. \tag{3.1}$$

Let X_i and ε_i be independent and have distributions ν and μ, respectively. If μ is known, the model is $\mathcal{P} = \{P_\nu = \nu * \mu | \nu \in \Theta\}$ where the parameter set Θ is given by the set of all probability measures. We aim for a convolution theorem for estimating the linear functional $\psi(P_\nu) = \chi(\nu) := \int \zeta \,\mathrm{d}\nu$ with $\zeta \in L^2(\nu)$. One of the most interesting examples is the estimation of the distribution function of X_1, corresponding to $\zeta = \mathbb{1}_{(-\infty,t]}$ for $t \in \mathbb{R}$.

In a general linear indirect density estimation setting, a convolution theorem was already proved by van Rooij et al. (1999), who use the spectral decomposition of the operator. Their approach applies however only for a restricted class of functionals, depending on the polar decomposition, and they need an abstract condition on the density of ν which is difficult to verify. It implicitly assumes an appropriate decay behavior on this density. Their application to the deconvolution setting is restricted to a specific example. Studying deconvolution in more detail, Söhl and Trabs (2012) have shown an information bound, assuming a polynomial decay behavior of a sufficiently regular Lebesgue density of ν and a bit more than second moments. They described the class of admissible functionals analytically, including the estimation of the distribution function of ν. Here, we are able to relax the conditions on ν and μ considerably, see Theorem 3.2 and Remark 4.12 below.

For any $\nu \in \Theta$ we choose the tangent space

$$\dot{\Theta}_\nu := L_0^2(\nu) = \overline{\mathrm{lin}}\dot{\Theta}_\nu.$$

According to Example 2.13, $\dot{\Theta}_\nu$ coincides with the maximal tangent set for direct observations. For any direction $b \in \dot{\Theta}_\nu$ and some sufficiently small $\tau > 0$ the path $[0, \tau) \ni t \to \nu_t$ where $\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu} = k(tb)/\int k(tb)\,\mathrm{d}\nu$ with $k : \mathbb{R} \to \mathbb{R}_+$ as in Example 2.13 is a submodel of Θ with $b = \frac{\partial}{\partial t}|_{t=0} \log(\mathrm{d}\nu_t)$. Using $|k(tb)| \leqslant t|b| \in L^2(\nu)$ and dominated convergence, the pathwise derivative of χ along $t \mapsto \nu_t$ at

$t = 0$ is given by

$$\lim_{t\to 0} t^{-1}(\chi(\nu_t) - \chi(\nu)) = \lim_{t\to 0} \int \zeta(x) t^{-1}\Big(\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}(x) - 1\Big)\,\mathrm{d}\nu(x)$$
$$= \int \zeta(x) b(x) \nu(\,\mathrm{d}x) = \langle \zeta, b\rangle_\nu =: \dot\chi_\nu b.$$

Hence, the derivative can be represented by $\dot\chi_\nu b = \langle \widetilde\chi_\nu, b\rangle_\nu$ for $\widetilde\chi_\nu = \zeta - \int \zeta\,\mathrm{d}\nu \in \dot\Theta_\nu$. The path $t \mapsto \nu_t$ induces a regular submodel $t \mapsto P_{\nu_t} = \nu_t * \mu$ which is shown by the following lemma.

Lemma 3.1. *For any nonzero $b \in \dot\Theta_\nu = L_0^2(\nu)$ the submodel $[0,\tau) \ni t \mapsto P_{\nu_t} = \nu_t * \mu$, for $\tau > 0$ sufficiently small, is Hellinger differentiable, that is* (2.9) *holds with continuous score operator*

$$A_\nu : \quad \dot\Theta_\nu \to L_0^2(P_\nu), \quad b \mapsto \mathbb{E}\left[b(X)|X+\varepsilon\right] = \frac{\mathrm{d}\big((b\nu)*\mu\big)}{\mathrm{d}P_\nu}, \tag{3.2}$$

where the expectation is taken with respect to the product measure measure $P^{(X,\varepsilon)} = \nu \otimes \mu$.

Proof. First, note that the (signed) measure $(f\nu) * \mu$ is absolutely continuous with respect to $\nu * \mu$ for any $f \in L^1(\nu)$, written as $(f\nu) * \mu \ll \nu * \mu$. In particular, the Radon–Nikodym density in (3.2) is well defined and $P_{\nu_t} \ll P_\nu$ for all $t > 0$. Let us write $\mathbb{E}_t[\bullet]$ for the expectation under $P_t^{(X,\varepsilon)} = \nu_t \otimes \mu$. We define $p_t(y) := \frac{\mathrm{d}P_{\nu_t}}{\mathrm{d}P_\nu}(y)$, $n_t(x) := \frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}(x) = k(tb(x))/\int k(tb)\,\mathrm{d}\nu$. Let $\mathbb{R} \times \mathscr{B}(\mathbb{R}) \ni (y, A) \mapsto \kappa_{X,X+\varepsilon}(y, A)$ be the regular conditional probability of $P^{(X,\varepsilon)}(X \in \bullet|X+\varepsilon)$ that is

$$\kappa_{X,X+\varepsilon}(y, A) = P^{(X,\varepsilon)}(X \in A|X+\varepsilon = y)$$

for P^ν-a.e. $y \in \mathbb{R}$ and all $A \in \mathscr{B}(\mathbb{R})$. We claim

$$p_t(Y) = \mathbb{E}_0\left[n_t(X)\big|X+\varepsilon = Y\right] = \int n_t(x)\kappa_{X,X+\varepsilon}(Y, \mathrm{d}x) \quad P_\nu - a.s. \tag{3.3}$$

To verify (3.3), we note for any Borel set $A \in \mathscr{B}(\mathbb{R})$

$$\mathbb{E}_0[\mathbb{1}_A(Y)p_t(Y)] = \mathbb{E}_t[\mathbb{1}_A(Y)] = \mathbb{E}_0[\mathbb{1}_A(X+\varepsilon)n_t(X)]$$
$$= \mathbb{E}_0\left[\mathbb{1}_A(Y)\,\mathbb{E}_0[n_t(X)|X+\varepsilon = Y]\right] \tag{3.4}$$

which shows the first equality in (3.3). The second one follows from the choice of $\kappa_{X,X+\varepsilon}$.

We will show regularity of the submodel $(-\tau,\tau) \ni t \mapsto P_{\nu_t} = \nu_t * \mu$ for a sufficiently small $\tau > 0$ by applying Proposition 2.1.1 by Bickel et al. (1998). Using the properties of k,

$$\dot n_t(x) := \frac{\partial}{\partial t} n_t(x) = \frac{b(x)k'(tb(x))\int k(tb)\,\mathrm{d}\nu - k(tb(x))\int bk'(tb)\,\mathrm{d}\nu}{(\int k(tb)\,\mathrm{d}\nu)^2}$$

can be uniformly in $t \in (-\tau,\tau)$ bounded by $c_b(b(x)+1)$ for a constant $c_b > 0$, depending on b, and it is continuous in t on $(-\tau,\tau)$ for some sufficiently small

$\tau > 0$. Since $b \in L^2(\nu) \subseteq L^2(\nu \otimes \mu)$, dominated convergence and (3.3) yield that $p_t(y)$ is continuously differentiable in $t \in (-\tau, \tau)$ for P_ν-almost all $y \in \mathbb{R}$ with derivative

$$\dot{p}_t(y) = \frac{\partial}{\partial t} p_t(y) = \int \dot{n}_t(x) \kappa_{X,X+\varepsilon}(y, \mathrm{d}x) = \mathbb{E}_0[\dot{n}_t(X) | X + \varepsilon = y].$$

By Jensen's inequality we see that

$$\begin{aligned} \|\dot{p}_t\|^2_{L^2(P_t)} &= \mathbb{E}_0 \left[p_t(Y) \big| \, \mathbb{E}_0[\dot{n}_t(X) | X + \varepsilon = Y] \big|^2 \right] \\ &\leqslant \mathbb{E}_0 \left[p_t(Y) |\dot{n}_t(X)|^2 \right]. \end{aligned} \tag{3.5}$$

Since $n_t(x)$ can be bounded uniformly in $t \in (-\tau, \tau)$ and $x \in \mathbb{R}$, the density $p_t(Y)$ is P_ν-a.s. bounded by some constant $C > 0$ owing to (3.3). Therefore, we conclude from the previous estimate together with the bound $|\dot{n}_t| \leqslant c_b(b+1)$ and $b \in L^2(\nu)$ that

$$\|\dot{p}_t\|^2_{L^2(P_t)} \leqslant C \, \mathbb{E}_0[|\dot{n}_t(X)|^2] \leqslant C c_b^2 \, \mathbb{E}_0[(b(X)+1)^2] < \infty.$$

In particular, the Fisher information $I_t := \mathbb{E}_t[\dot{p}_t(Y)^2]$ is finite. Using (3.5), we infer that I_t is continuous. Noting that

$$\frac{\partial}{\partial t}\Big|_{t=0} \int k(tb) \, \mathrm{d}\nu = \int b \, \mathrm{d}\nu = 0 \quad \text{and thus} \quad \dot{n}_0(y) = \frac{\partial}{\partial t}\Big|_{t=0} n_t(x) = b(x),$$

we have $I_0 = \mathbb{E}_0[b(X)^2]$ and I_t is therefore nonzero for $b \neq 0$ and t small enough. In combination with the continuous differentiability of p_t, Proposition 2.1.1 by Bickel et al. (1998) yields the Hellinger differentiability (2.9) at $t = 0$ with derivative $\frac{1}{2}\dot{p}_0$. We obtain the score operator

$$A_\nu b := \dot{p}_0 = \mathbb{E}_0[b(X) | X + \varepsilon].$$

To see that $A_\nu : \dot{\Theta}_\nu \to L^2_0(P_\nu)$ is well defined and continuous, we again use Jensen's inequality which yields

$$\mathbb{E}_0[A_\nu b] = \mathbb{E}_0[b(X)] = 0 \quad \text{and} \quad \mathbb{E}_0[|A_\nu b|^2] \leqslant \mathbb{E}_0[|b(X)|^2] = \|b\|^2_{L^2(\nu)}.$$

Finally, a similar calculation as (3.4) shows $A_\nu b = \mathbb{E}_0[b(X) | X + \varepsilon] = \frac{\mathrm{d}((b\nu) * \mu)}{\mathrm{d}P_\nu}$. □

Lemma 3.1 shows that Assumption A is satisfied and thus Lemma 2.12 yields regularity of the deconvolution model with rate $r_n = 1/\sqrt{n}$. In order to apply Corollary 2.14, we have to determine the adjoint of the score operator. For any $g \in L^2_0(P_\nu)$ and any $b \in \dot{\Theta}_\vartheta \subseteq L^2(\nu)$ the map $\mathbb{R}^2 \ni (x, y) \mapsto g(x+y)b(x)$ is $\nu \otimes \mu$-integrable due to the Cauchy–Schwarz inequality. $(b\nu) * \mu \ll P_\nu$ and Fubini's theorem thus yield

$$\begin{aligned} \langle A_\nu b, g \rangle_{P_\nu} &= \int (A_\nu b) g \, \mathrm{d}P_\nu = \int g \, \mathrm{d}\big((b\nu) * \mu\big) \\ &= \int \int g(x+y) b(x) \nu(\mathrm{d}x) \mu(\mathrm{d}y) = \langle \mu(-\bullet) * g, b \rangle_\nu. \end{aligned}$$

Noting that Jensen's inequality shows

$$\|\mu(-\bullet) * g\|^2_{L^2(\nu)} = \int \Big(\int g(x+y)\mu(\,\mathrm{d}y)\Big)^2 \nu(\,\mathrm{d}x) \leqslant \int g^2 \,\mathrm{d}(\nu * \mu) = \|g\|^2_{L^2(P_\nu)}$$

and that $\int (\mu(-\bullet) * g)\,\mathrm{d}\nu = \int g\,\mathrm{d}(\nu * \mu) = 0$, the adjoint score operator equals

$$A^\star_\nu : \quad L^2_0(P_\nu) \to \dot{\Theta}_\nu, \quad g \mapsto \mu(-\bullet) * g. \tag{3.6}$$

Under weak conditions on the measures ν and μ we conclude the following convolution theorem. Thereby we extend the definition of the Fourier transform to finite measure μ on the Borel measurable space $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ by $\mathcal{F}\,\mu(u) = \int e^{iux}\mu(\,\mathrm{d}x)$.

Theorem 3.2. *In the deconvolution model* (3.1) *suppose that* $\varphi_\varepsilon(u) := \mathbb{E}[e^{iu\varepsilon_1}] = \mathcal{F}\,\mu(u) \neq 0$ *for all* $u \in \mathbb{R}$ *and that* ν *admits a Lebesgue density. Then* $A^\star_\nu$ *as given in* (3.6) *is injective. Let moreover* $\zeta^{(1)}, \ldots, \zeta^{(d)} \in L^2(\nu)$ *satisfy* $\zeta^{(j)} \in \operatorname{ran} A^\star_\nu$ *for* $j = 1, \ldots, d$ *and* $d \in \mathbb{N}$. *Then the limit distribution of any estimator of the parameter* $(\int \zeta^{(1)}\,\mathrm{d}\nu, \ldots, \int \zeta^{(d)}\,\mathrm{d}\nu)$ *which is regular with respect to the rate* $n^{-1/2}$ *equals* $\mathcal{N}(0,\Sigma) * M$ *for some Borel probability measure* M *and with covariance matrix* $\Sigma \in \mathbb{R}^{d\times d}$ *given by*

$$\Sigma_{j,k} = \int \Big((A^\star_\nu)^{-1}\zeta^{(j)}\Big)\Big((A^\star_\nu)^{-1}\zeta^{(k)}\Big)\,\mathrm{d}P_\nu - \Big(\int \zeta^{(j)}\,\mathrm{d}\nu\Big)\Big(\int \zeta^{(k)}\,\mathrm{d}\nu\Big) \tag{3.7}$$

for $j, k = 1, \ldots, d$.

Proof. Let us first show that on the assumptions the adjoint operator $A^\star_\nu$ is injective. Since ν admits a Lebesgue density, the equivalence classes with respect to the Lebesgue measure embed into the equivalence classes with respect to ν and with respect to P_ν. Hence, we can consider the subset $\mathcal{G} := L^2(\mathbb{R}) \cap L^2_0(P_\nu)$, which is dense in $L^2_0(P_\nu)$ (cf. Section 6.4 (ii)). Since the kernel of the continuous operator $A^\star_\nu$ is closed, it is sufficient to show that the restricted operator $A^\star_\nu|_{\mathcal{G}}$ is injective. For any $g \in \mathcal{G}$ it holds $0 = A^\star_\nu g = \mu(-\bullet) * g$ if and only if $0 = \mathcal{F}[\mu(-\bullet) * g] = \varphi_\varepsilon(-\bullet)\,\mathcal{F}\,g$ which is equivalent to $\mathcal{F}\,g = 0$ since $|\varphi_\varepsilon| > 0$ by assumption. Hence, the kernel of $A^\star_\nu$ equals $\{0\}$.

To infer the information bound, recall that the gradient of the linear functional $\int \zeta^{(j)}\,\mathrm{d}\nu$ is $\widetilde{\chi}^{(j)}_\nu = \zeta^{(j)} - \int \zeta^{(j)}\,\mathrm{d}\nu$. Furthermore, note that $\zeta^{(j)} \in \operatorname{ran} A^\star_\nu$ implies $\widetilde{\chi}^{(j)}_\nu \in \operatorname{ran} A^\star_\nu$ since $A^\star_\nu a = \mu(-\bullet) * a = \int a\,\mathrm{d}\mu = a$ for any real number $a \in \mathbb{R}$. Therefore, Lemma 3.1, Corollary 2.14 and the injectivity of $A^\star_\nu$ yield the vector of efficient influence functions

$$\widetilde{\psi}^{(j)}_{P_\nu} = (A^\star_\nu)^{-1}\widetilde{\chi}^{(j)}_\nu = (A^\star_\nu)^{-1}\zeta^{(j)} - \int \zeta^{(j)}\,\mathrm{d}\nu, \quad j = 1, \ldots, d,$$

and the assertion follows from Theorem 2.7. □

Remark 3.3. The assumption $\varphi_\varepsilon(u) \neq 0, u \in \mathbb{R}$, is not sufficient for the injectivity of $A^\star_\nu$ as the following counterexample shows: Let $\nu = \delta_0$ be the Dirac measure in zero and $\mu = \mathcal{N}(0,1)$ be standard normal such that $P_\nu = \nu * \mu = \mathcal{N}(0,1)$. Consider $g \in L^2_0(P_\nu)$ with $g(x) = x^3, x \in \mathbb{R}$. Then, $A^\star_\nu g(x) = \mathbb{E}[(x+\varepsilon_1)^3]$ is zero at the origin. Hence, $A^\star_\nu g = 0$ ν-a.s. and $0 \neq g \in \ker A^\star_\nu$. A sufficient condition for $A^\star_\nu$ being injective is given in Theorem 3.2 by assuming additionally a

Lebesgue density of ν, which is a natural assumption. In particular, we obtain $\overline{\operatorname{ran}} A_\nu = L_0^2(P_\nu)$ implying that the tangent space $\dot{\mathcal{P}}_{P_\nu} = A_\nu \dot{\Theta}_\nu$ is dense in the set of all score functions. Without injectivity the convolution theorem remains true if the inverse $(A_\nu^\star)^{-1}$ in (3.7) is replaced by the Moore–Penrose pseudoinverse $(A_\nu^\star)^\dagger$.

In view of Theorem 2.7 our condition $\zeta^{(j)} \in \operatorname{ran} A_\nu^\star, j = 1, \ldots, d$, is necessary and sufficient for the regularity of the parameter. The remaining question is under which conditions $\mathbb{1}_{(-\infty,t]} \in \operatorname{ran} A_\nu^\star, t \in \mathbb{R}$, and how the pre-image looks like in this case. Applying the approach by Nickl and Reiß (2012), we will give an answer in the more involved Lévy setting. Under certain assumptions on μ it can be carried over to the deconvolution case and leads to a similar result, but with weaker conditions on μ than by Söhl and Trabs (2012) (see Remark 4.12 below). In particular, these conditions will imply that $\mathbb{1}_{(-\infty,t]}$ can be estimated with parametric rate.

4 Application to Lévy processes

4.1 Setting and regularity

Recall that a Lévy process $(L_t)_{t\geqslant 0}$ is a stochastic process which is stochastically continuous with $L_0 = 0$ and which has stationary and independent increments. Let (L_t) be real-valued. For some distance $\Delta > 0$ we observe the Lévy process at equidistant time points $t_k = \Delta k$ with $k = 0, \ldots, n$. In the so called low-frequency regime Δ remains fixed as n goes to infinity. The Lévy process is uniquely determined by its characteristic triplet consisting of the volatility $\sigma^2 \geqslant 0$, of the drift parameter $\gamma \in \mathbb{R}$ and of the Lévy or jump measure ν on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ which satisfies $\int_{\mathbb{R}} (|x|^2 \wedge 1)\nu(\,\mathrm{d}x) < \infty$ and $\nu(\{0\}) = 0$ (cf. Sato, 1999, Chap. 2). Our aim is to derive a convolution theorem for the estimation of the linear functional of the jump measure

$$\chi(\nu) := \int \zeta \,\mathrm{d}\nu \quad \text{for } \zeta \in L^1(\nu) \cap L^2(\nu). \tag{4.1}$$

If ζ is $\mathbb{R}^d$-valued, the scalar product in (4.1) has to be interpreted coordinatewise. As a relevant example the reader should have in mind the generalized distribution function of ν. It corresponds to $\zeta = \mathbb{1}_{(-\infty,t]}$ for $t < 0$ and $\zeta = \mathbb{1}_{[t,\infty)}$ for $t > 0$. This adaptation of the standard distribution function is necessary owing to the possibly existing singularity of ν at the origin. In order for the estimation of $\chi(\nu)$ to be possible with parametric rate, we restrict on processes with finite variation in view of the lower bounds by Neumann and Reiß (2009). That means $\sigma^2 = 0$ and $\int_{\mathbb{R}} (|x| \wedge 1)\nu(\,\mathrm{d}x) < \infty$ are assumed. For a recent review on the statistical inference on Lévy processes we refer to Reiß (2013).

Due to the stationary and independent increments of (L_t), the random variables $Y_k := L_{\Delta k} - L_{\Delta(k-1)}, k = 1, \ldots, n$, are independent and identically distributed. Their characteristic function is given by the Lévy-Khintchine formula

$$\varphi_\nu(u) = \mathbb{E}\left[e^{iuL_\Delta}\right] = \exp\Big(\Delta\big(i\gamma u + \int (e^{iux} - 1)\nu(\,\mathrm{d}x)\big)\Big). \tag{4.2}$$

Fixing the drift γ, the model is given by

$$\mathcal{P} = \big\{P_\nu = \mathcal{F}^{-1}\,\varphi_\nu | \nu \in \Theta\big\} \quad \text{with} \tag{4.3}$$
$$\Theta = \Big\{\nu \text{ jump measure on } (\mathbb{R}, \mathscr{B}(\mathbb{R}))\Big| \int (|x| \wedge 1)\nu(\,\mathrm{d}x) < \infty\Big\}.$$

Compared to tangents at the set of probability measures in Example 2.13, directions for the Lévy measures do not need to be centered since Lévy measures are not normalized. In general, jump measures are even not finite such that $L^2(\nu)$, which gives the Hilbert space structure, is still too large. We should intersect with $L^1(\nu)$ to include linear functionals as (4.1). Hence, we define the tangent space at $\nu \in \Theta$ as

$$\dot{\Theta}_\nu := L^1(\nu) \cap L^2(\nu) = \operatorname{lin} \dot{\Theta}_\nu. \tag{4.4}$$

Using the function $k(y) = 2/(1 + e^{-2y})$ from Example 2.13, for any $b \in \dot{\Theta}_\nu$ the path $[0, 1) \ni t \mapsto \nu_t$ with $\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}(x) = k\big(tb(x)\big)$ is contained in Θ and satisfies

$$b(x) = \frac{\partial}{\partial t}\Big|_{t=0} \log\Big(\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}(x)\Big).$$

On this path the derivative of the functional (4.1) can be calculated with use of dominated convergence, noting that $|k(tb) - 1| \leqslant t|b| \in L^2(\nu)$. Hence,

$$\begin{aligned}\lim_{t\to 0} t^{-1}(\chi(\nu_t) - \chi(\nu)) &= \lim_{t\to 0} \int \zeta(x) t^{-1}\Big(\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}(x) - 1\Big)\,\mathrm{d}\nu(x)\\ &= \int \zeta(x) b(x) \nu(\,\mathrm{d}x) = \langle \zeta, b\rangle_\nu =: \dot{\chi}_\nu b\end{aligned}$$

and thus the gradient is given by $\widetilde{\chi}_\nu = \zeta$. Compared to the deconvolution setting, we do not need to center $\widetilde{\chi}_\nu$ because the total mass of the Lévy measure is allowed to change along the path.

To apply Corollary 2.14, we need to verify Assumption A for the Lévy model. By the Lévy–Khintchine representation the laws P_{ν_t} satisfy

$$\begin{aligned}P_{\nu_t} &= \mathcal{F}^{-1}\Big[\exp\Big(\Delta\Big(i\gamma u + \int (e^{iux} - 1)\nu_t(\,\mathrm{d}x)\Big)\Big)\Big]\\ &= \mathcal{F}^{-1}\Big[\exp\Big(\Delta \int (e^{iux} - 1)\Big(k\big(tb(x)\big) - 1\Big)\nu(\,\mathrm{d}x)\Big)\varphi_\nu(u)\Big].\end{aligned} \tag{4.5}$$

Owing to $(k(tb) - 1) \in L^1(\nu)$, the measure P_{ν_t} is a convolution of P_ν and a compound Poisson type measure with signed jump measure $\Delta(k(tb) - 1)\,\mathrm{d}\nu$. To see that the submodel $t \mapsto P_{\nu_t}$ is dominated, we check that the Hellinger distance of the jump measures $\int(\sqrt{\mathrm{d}\nu_t} - \sqrt{\mathrm{d}\nu})^2 = \int(\sqrt{k(tb(x))} - 1)^2\nu(\,\mathrm{d}x)$ is finite for all t. Since the drift γ remains constant, Theorem 33.1 by Sato (1999) yields absolute continuity of P_{ν_t} with respect to P_ν, denoted as $P_{\nu_t} \ll P_\nu$, for all t. To find the Hellinger derivative of the path $t \mapsto \frac{\mathrm{d}P_{\nu_t}}{\mathrm{d}P_\nu}$ at $t = 0$, we note by dominated convergence

$$\begin{aligned}&\mathcal{F}^{-1}\Big[\frac{\partial}{\partial t}\Big|_{t=0} \exp\Big(\Delta \int (e^{iux} - 1)\big(k\big(tb(x)\big) - 1\big)\nu(\,\mathrm{d}x)\Big)\varphi_\nu(u)\Big]\\ =&\Delta\,\mathcal{F}^{-1}\Big[\varphi_\nu(u)\Big(\int (e^{iux} - 1)b(x)\nu(\,\mathrm{d}x)\Big)\Big]\\ =&\Delta\Big(P_\nu * (b\nu) - \int b(x)\nu(\,\mathrm{d}x)P_\nu\Big).\end{aligned}$$

This indicates how the score operator should look like. The following proposition determines the score operator A_ν and shows Hellinger regularity of the parametric submodel $t \mapsto P_{\nu_t}$. This is the key result to apply the theory of Section 2 to the Lévy model.

Proposition 4.1. *Let the model $\mathcal{P}$ be given by* (4.3) *with the tangent space $\dot{\Theta}_\nu$ at $\nu \in \Theta$ as defined in* (4.4)*. Then $P_\nu * (b\nu) \ll P_\nu$ for all $b \in \dot{\Theta}_\nu \cap L^\infty(\nu)$. Moreover, the linear operator*

$$A_\nu\big|_{\dot{\Theta}_\nu \cap L^\infty(\nu)} : \dot{\Theta}_\nu \cap L^\infty(\nu) \to L_0^2(P_\nu), \quad b \mapsto \Delta \frac{\mathrm{d}\big(P_\nu * (b\nu)\big) - \big(\int b \,\mathrm{d}\nu\big)\,\mathrm{d}P_\nu}{\mathrm{d}P_\nu} \tag{4.6}$$

is bounded. $\dot{\Theta}_\nu \cap L^\infty(\nu)$ is dense in $\dot{\Theta}_\vartheta$ and thus $A_\nu : \dot{\Theta}_\nu \to L_0^2(P_\nu)$ can be defined as its unique continuous extension. Then for all $b \in \dot{\Theta}_\nu$ the associated submodel $[0,1) \ni t \mapsto P_{\nu_t}$ is Hellinger differentiable at zero with derivative $A_\nu b$, that means (2.9) *is fulfilled.*

The proof of this proposition is given in Section 6.2. An essential ingredient is an estimate of the Hellinger integral of two infinitely divisible distributions by Liese (1987). More precisely, his results imply (for details see Section 6.2)

$$\int \Big(\frac{\mathrm{d}P_{\nu_t}}{\mathrm{d}P_\nu}\Big)^2 \mathrm{d}P_\nu \leqslant \exp\Big(\frac{1}{2}\int \big(\frac{\mathrm{d}\nu_t}{\mathrm{d}\nu} - 1\big)^2 \mathrm{d}\nu\Big) \leqslant \exp\Big(\frac{1}{2}t^2\|b\|^2_{L^2(\nu)}\Big). \tag{4.7}$$

Remark 4.2. Relying on the semimartingale structure of the model, an alternative strategy to prove Proposition 4.1 is as follows: Observing L_Δ is a sub-experiment of observing $(L_t)_{0\leqslant t\leqslant\Delta}$ in continuous time. Hence, by proving Hellinger differentiability of the latter model, we find a score process, say $(V_t)_{0\leqslant t\leqslant\Delta}$, and the score operator is then given by $A_\nu b = \mathbb{E}[V_\Delta | L_\Delta]$.

Theorem 2.34 by Jacod (1990) yield local differentiability of the experiment with continuous observations corresponding to the score process

$$V_t = \sum_{s\leqslant t} b(\Delta L_s) - t\int b\,\mathrm{d}\nu.$$

Noting that the Hellinger processes are deterministic (cf. Jacod and Shiryaev, 2003, Rem. IV.1.25), local differentiability implies Hellinger differentiability, see Jacod (1989).[1]

Proposition 4.1 shows that the Lévy model $\mathcal{P}$, defined in (4.3) equipped with the tangent space $\dot{\Theta}_\nu$, given in (4.4) satisfies Assumption A. In particular, it is a regular indirect model at any $\nu \in \Theta$ with respect to the rate $n^{-1/2}$ by Lemma 2.12. Having in mind the regularity Lemma 3.1 in the deconvolution model, the score operators look very similar in both models. Since the gradient does not have to integrate to zero in the Lévy model, the centering is incorporated in the operator A_ν. Apart from that the convolution structure is the same. Therefore, the Lévy model can locally be weakly approximated with a linear white noise model whose operator is of convolution type.

By Proposition 4.1 the score operator A_ν is characterized by (4.6). To prove information bounds, we will combine this result with Proposition 2.11 which

[1] Thanks to an anonymous referee for pointing out this approach.

shows that it is sufficient to study $A_\nu^\star$ on a nicely chosen, dense subset of $L_0^2(P_\nu)$. Then Theorem 2.7 provides the convolution theorem for all $\zeta \in \operatorname{ran} A_\vartheta^\star$. In the following we will discuss Lévy processes with finite and infinite jump activity separately because the analytical properties of the score operator are quite different: In the compound Poisson case the inverse adjoint score operator can be explicitly expressed as a convolution with a finite signed measure. If the jump intensity is infinite, the distribution P_ν possesses a Lebesgue density and thus $A_\nu^\star$ will be a smoothing operator.

4.2 Compound Poisson processes

Let (L_t) be a compound Poisson process with jump intensity $\lambda := \nu(\mathbb{R}) < \infty$. Consequently, the tangent space simplifies to $\dot{\Theta}_\vartheta = L^2(\nu)$ and the measure P_ν can be written as the convolution exponential (cf. Sato, 1999, Rem. 27.3)

$$P_\nu = \delta_{\Delta\gamma} * \Big(e^{-\Delta\lambda} \sum_{k=0}^{\infty} \frac{\Delta^k}{k!} \nu^{*k}\Big), \tag{4.8}$$

where δ_x denotes the Dirac measure in $x \in \mathbb{R}$. Define the subsets

$$\begin{aligned} \mathcal{G} &:= L^\infty(P_\nu) \cap L_0^2(P_\nu) \subseteq L_0^2(P_\nu) \quad \text{and} \\ \mathcal{H} &:= \big\{h \in L^2(\nu) \big| \sup_{k=0,1,\dots} \|h\|_{L^\infty(\nu^{*k})} < \infty\big\} \subseteq \dot{\Theta}_\nu \end{aligned} \tag{4.9}$$

which are dense in $L_0^2(P_\nu)$ and $L^2(\nu)$, respectively. Let $g \in \mathcal{G}$ and $b \in \mathcal{H}$. By Proposition 4.1 we know $(b\nu) * P_\nu \ll P_\nu$ which implies $g \in L^\infty((b\nu) * P_\nu)$. Hence, $\int g \,\mathrm{d}P_\nu = 0$ and Fubini's theorem yield

$$\begin{aligned} \langle A_\nu b, g\rangle_{P_\nu} = \int (A_\nu b) g \,\mathrm{d}P_\nu &= \Delta \int g \,\mathrm{d}\big(P_\nu * (b\nu)\big) - \Delta\Big(\int b \,\mathrm{d}\nu\Big)\Big(\int g \,\mathrm{d}P_\nu\Big) \\ &= \Delta \int\int g(x+y) b(x) \,\mathrm{d}\nu(x) \,\mathrm{d}P_\nu(y) \\ &= \Delta \langle P_\nu(-\bullet) * g, b\rangle_\nu. \end{aligned} \tag{4.10}$$

Therefore, the adjoint score operator on $\mathcal{G}$ is $A_\nu^\star|_{\mathcal{G}} : \mathcal{G} \to L^2(\nu), g \mapsto \Delta P_\nu(-\bullet) * g$.

Lemma 4.3. *The map $A_\nu^\star : \mathcal{G} \to \mathcal{H}, g \mapsto P_\nu(-\bullet) * g$ is well defined.*

This lemma is proved in Section 6.3. Although the centering of $g \in L_0^2(P_\nu)$ implies $A_\nu^\star g(0) = \int g \,\mathrm{d}P_\nu = 0$, it does not cause an additional constraint owing to $\nu(\{0\}) = 0$. In general, $A_\nu^\star$ is not injective as the following example shows:

Example 4.4 (Poisson process)**.** Setting $\nu = \delta_1, \gamma = 0$ and $\Delta = \lambda = 1$, the law $P_\nu = e^{-1} \sum_{k=0}^\infty \delta_k/(k!)$ is the Poisson distribution and the adjoint score operator is given by

$$A_\nu^\star g(x) = e^{-1} \sum_{k=0}^{\infty} g(x+k)/(k!), \quad x \in \mathbb{R}\,.$$

Consider the function $g = \mathbb{1}_{\{0\}} - 2\mathbb{1}_{\{1\}} + 2\mathbb{1}_{\{2\}}$ which is a nonzero element of $\mathcal{G}$ by construction. However, $A_\nu^\star g(1) = 0$ and thus $0 \neq g \in \ker A_\nu^\star$ contradicting injectivity.

As in the deconvolution model we assume therefore that ν admits a Lebesgue density concluding injectivity of $A_\nu^\star$ exactly as in Theorem 3.2. Since $|\varphi_\nu(u)| = e^{\Delta \int (\cos(ux)-1)\nu(\,\mathrm{d}x)} \geqslant e^{-2\Delta\lambda}$ for all $u \in \mathbb{R}$ by (4.2), the inverse of $A_\nu^\star$ is then the deconvolution operator $h \mapsto \Delta^{-1}\,\mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)] * h$ with the finite signed measure

$$\mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)] = \delta_{-\Delta\gamma} * \Big(e^{\Delta\lambda} \sum_{k=0}^{\infty} \frac{(-\Delta)^k}{k!} \nu(-\bullet)^{*k}\Big)$$

which is well defined on $\mathcal{H}$. In particular, the pre-image of the indicator function $\zeta = \mathbb{1}_{(-\infty,t]}$ (or equivalently $\mathbb{1}_{(-\infty,t]}\mathbb{1}_{\mathbb{R}\setminus\{0\}}$) is well defined for any $t \in \mathbb{R}$. Consequently, Corollary 2.14 and Theorem 2.7 yield

Corollary 4.5. *Let (L_t) be a pure jump process of compound Poisson type with jump measure ν which is absolutely continuous with respect to the Lebesgue measure and with drift $\gamma \in \mathbb{R}$. Then the limit distribution of any regular estimator of the distribution function $\mathbb{R}^d \ni (t_1, \dots, t_d) \mapsto (\nu((-\infty, t_1]), \dots, \nu((-\infty, t_d]))$, for $d \in \mathbb{N}$, is a convolution $\mathcal{N}(0, \Sigma) * M$ for some Borel probability measure M and with covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ given by*

$$\Sigma_{i,j} = \Delta^{-2} \int \Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\nu(-\bullet)}\Big] * \mathbb{1}_{(-\infty,t_i]}\Big)\Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\nu(-\bullet)}\Big] * \mathbb{1}_{(-\infty,t_j]}\Big)\,\mathrm{d}P_\nu$$

for $i, j \in \{1, \dots, d\}$.

Considering the negative half line, this lower bound coincides with the asymptotic variance of the kernel estimator by Nickl and Reiß (2012). An interesting deviation is obtained by restricting the model on compound Poisson processes with known jump intensity $\lambda > 0$ as studied in the *decompounding* problem by Buchmann and Grübel (2003). Similarly to the deconvolution model in Section 3 the tangent space is then given by $L_0^2(\nu)$ and thus the gradient of the functional $\chi(\nu) = \nu((-\infty, t])$ equals $\widetilde{\chi}_\nu = \mathbb{1}_{(-\infty,t]} - \nu((-\infty, t])$. We obtain the smaller information bound, setting $d = 1$ for simplicity,

$$\Delta^{-2} \int \Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\nu(-\bullet)}\Big] * \mathbb{1}_{(-\infty,t]}\Big)^2 \,\mathrm{d}P_\nu - \Delta^{-2}\nu((-\infty, t])^2.$$

That means an efficient estimator which "knows" the jump intensity should have a smaller variance than for unknown λ and the statistical problem is significantly simpler. Indeed, the estimator by Buchmann and Grübel (2003) is asymptotically normal with the above variance.

4.3 Lévy processes with infinite jump activity

If the Lévy process has infinite jump activity, the analysis is more difficult. However, we can profit from the absolute continuity of the infinite divisible distribution P_ν with respect to the Lebesgue measure (Sato, 1999, Thm. 27.4). To apply Fourier methods, we will again assume that ν admits a Lebesgue density which implies in particular that the set of Lebesgue-a.e. equivalence classes embeds into the ν-a.e. and into the P_ν-a.e. equivalence classes. Keeping the Hilbert space structure, we can then define

$$\mathcal{G} := H^\infty(\mathbb{R}) \cap L_0^2(P_\nu) \quad \text{and} \quad \mathcal{H} := \big\{b \in H^\infty(\mathbb{R}) | b(0) = 0\big\} \cap \dot{\Theta}_\nu, \tag{4.11}$$

where $H^\infty(\mathbb{R}) := \bigcap_{s\geqslant 0} H^s(\mathbb{R})$ with Sobolev spaces $H^s(\mathbb{R}) := \{f \in L^2(\mathbb{R}) | \|(1+|u|^2)^{s/2}\,\mathcal{F}f(u)\|_{L^2} < \infty\}$ of regularity $s \geqslant 0$. For $b \in \mathcal{H}$ the condition $b(0) = 0$ should hold for the continuous version of b. To allow that the generalized distribution function of ν can be estimated with parametric rate, we concentrate on mildly ill-posed problems leading to the assumption that $|\varphi_\nu(u)|$ decays polynomially as $|u| \to \infty$ (cf. Neumann and Reiß, 2009).

Lemma 4.6. *Let the finite variation Lévy process (L_t) with $\nu \in \Theta$ have infinite jump activity satisfying $|\varphi_\nu(u)| \gtrsim (1+|u|)^{-\beta}$ for some $\beta > 0$ and let ν be absolutely continuous with respect to the Lebesgue measure. Then*

(i) on $\mathcal{G}$ from (4.11) the adjoint score operator $A_\nu^\star|_{\mathcal{G}}$ is a bijection onto $\mathcal{H}$ satisfying

$$A_\nu^\star\big|_{\mathcal{G}}: \quad \mathcal{G} \to \mathcal{H} \subseteq \dot{\Theta}_\nu, \qquad g \mapsto \Delta(P_\nu(-\bullet) * g), \tag{4.12}$$

$$(A_\nu^\star)^{-1}\big|_{\mathcal{H}}: \quad \mathcal{H} \to \mathcal{G} \subseteq L_0^2(P_\nu), \qquad b \mapsto \Delta^{-1}\,\mathcal{F}^{-1}\left[\mathcal{F}b/\varphi_\nu(-\bullet)\right], \tag{4.13}$$

(ii) $\mathcal{G}$ is dense in $L_0^2(P_\nu)$ and $A_\nu^\star$ is the unique continuous extension of $A_\nu^\star\big|_{\mathcal{G}}$.

The proof of this lemma uses that by the polynomial decay $(1+|u|)^{-\beta} \lesssim |\varphi_\nu(u)| \lesssim 1$ both operators, $A_\nu^\star$ and $(A_\nu^\star)^{-1}$, are Fourier multipliers on Sobolev spaces and thus $H^\infty(\mathbb{R})$-functions are mapped into $H^\infty(\mathbb{R})$ again. The smoothness will be used to show $|(A_\nu^\star g)(x)| \lesssim 1 \wedge |x|$ for $x \in \mathbb{R}$ in order to verify that $A_\nu^\star$ is well defined. For details we refer to Section 6.4.

Comparing the adjoint score operator (4.13) in the Lévy model to $A_\nu^\star$ in the deconvolution model (3.2), we see that both operators have exactly the same structure. To invert $A_\nu^\star$, we have to deconvolve with the observation measure itself in the Lévy case. From our lower bounds perspective we clearly recover this *auto-deconvolution* phenomenon, which was already described by Belomestny and Reiß (2006) as well as Nickl and Reiß (2012). For convenience we will write throughout $\mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)] * b = \mathcal{F}^{-1}[\mathcal{F}b/\varphi_\nu(-\bullet)]$ which is justified in distributional sense. In combination with the results for the compound Poisson case Lemma 4.6 has two immediate consequences.

Remark 4.7. If the Lévy process is of finite variation, has an absolutely continuous jump measure and has either finite jump activity or has a polynomial decreasing characteristic function, then

(i) $A_\nu^\star$ is injective and therefore $\overline{\operatorname{ran}} A_\nu = L_0^2(P_\nu)$. This means that the tangent set $\dot{\mathcal{P}}_{P_\nu} = A_\nu \dot{\Theta}_\nu$ is dense in $L^2(P_\nu)$.

(ii) For any linear functional $\chi(\nu) = \int \zeta \,\mathrm{d}\nu$ satisfying $\zeta \in \mathcal{H}$, where $\mathcal{H} \subseteq \dot{\Theta}_\nu$ is defined in (4.9) and (4.11), respectively, the information bound is given by

$$\Delta^{-2} \int \left(\mathcal{F}^{-1}\left[\frac{1}{\varphi_\nu(-\bullet)}\right] * \zeta\right)^2 \mathrm{d}P_\nu.$$

The subset $\mathcal{H}$ of arbitrary large Sobolev smoothness is obviously very restrictive. Let us extend the information bound to a larger class of functionals by using Proposition 2.11. This is illustrated in the following example.

Example 4.8 (Gamma process)**.** Let (L_t) be a gamma process with $Y_k \sim \Gamma(\alpha\Delta, \lambda)$ for all $k = 1, \dots, n$. For simplicity set $\lambda = 1$. The probability density, the characteristic function and the Lévy measure are given by

$$\gamma_{\alpha\Delta}(x) := \frac{1}{\Gamma(\alpha\Delta)} x^{\alpha\Delta-1} e^{-x} \mathbb{1}_{[0,\infty)}(x), \qquad \varphi_\nu(u) = (1-iu)^{-\alpha\Delta} \quad \text{and}$$
$$\nu(\,\mathrm{d}x) = \alpha x^{-1} e^{-x} \mathbb{1}_{[0,\infty)}(x)\,\mathrm{d}x, \quad \text{for } x, u \in \mathbb{R},$$

respectively. Therefore, $|\varphi_\nu|$ decays with polynomial rate $\beta = \Delta\alpha$ and we can apply Lemma 4.6. The estimation of the generalized distribution function $\chi(\nu) = \int_t^\infty \mathrm{d}\nu$ for some fixed $t > 0$, induces the gradient $\widetilde{\chi}_\nu = \mathbb{1}_{[t,\infty)}$. To approximate $\widetilde{\chi}_\nu$ with a sequence in $\mathcal{H}$, we construct $\chi_n(x) = \int_{-\infty}^x (\delta_n(y-t) - \delta_n(y-n))\,\mathrm{d}y$ for a Dirac sequence (δ_n). More precisely, let $(\delta_n) \subseteq C^\infty(\mathbb{R})$ be a family of smooth nonnegative functions satisfying $\int_\mathbb{R} \delta_n = 1$ and $\operatorname{supp}\delta_n \subseteq [-1/n, 1/n]$. Obviously, $(\chi_n) \subseteq \mathcal{H}$. Since ν is a finite measure on $\mathbb{R}\setminus(-\varepsilon,\varepsilon)$ for any $\varepsilon > 0$, dominated convergence shows $\|\widetilde{\chi}_\nu - \chi_n\|_{L^2(\nu)} \to 0$. Denoting the distribution function of $\Gamma(\beta, 1)$ by Γ_β, we obtain for $\alpha\Delta < 1/2$

$$\begin{aligned}
(A_\nu^\star)^{-1}\chi_n &= \Delta^{-1}\,\mathcal{F}^{-1}[(1+iu)(1+iu)^{\alpha\Delta-1}] * \chi_n \\
&= \Delta^{-1}\gamma_{1-\alpha\Delta}(-\bullet) * (\chi_n - \chi_n') \\
&= \Delta^{-1}\big(\gamma_{1-\alpha\Delta}(-\bullet) * \chi_n - \gamma_{1-\alpha\Delta}(-\bullet) * \delta_n(\bullet - t) \\
&\qquad + \gamma_{1-\alpha\Delta}(-\bullet) * \delta_n(\bullet - n)\big) \\
&\to \Delta^{-1}\big((1 - \Gamma_{1-\alpha\Delta}(t-\bullet)) - \gamma_{1-\alpha\Delta}(t-\bullet)\big) \\
&= \Delta^{-1}\gamma_{1-\alpha\Delta}(-\bullet) * (\mathbb{1}_{(-\infty,t]} - \delta_t) =: \psi
\end{aligned} \tag{4.14}$$

where the convergence holds in $L^2(\mathbb{R})$ owing to $\gamma_{1-\alpha\Delta} \in L^2(\mathbb{R})$ for $\alpha\Delta < 1/2$. Therefore, in a natural way the limiting object is $\psi = \mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)] * \mathbb{1}_{(-\infty,t]}$. When does this limit hold in $L^2(P_\nu)$, too? As we saw above, the probability density of P_ν is bounded everywhere except for the singularity at zero which is of order $1 - \alpha\Delta$. For any $t > 0$ and n large enough $\gamma_{1-\alpha\Delta} * \delta_n(\bullet - t)$ is uniformly bounded in a small neighborhood of zero such that dominated convergence around zero together with the $L^2(\mathbb{R})$-convergence on the real line yields $\|(A_\nu^\star)^{-1}\chi_n - \psi\|_{L^2(P_\nu)} \to 0$. Hence, Proposition 2.11 shows $\mathbb{1}_{[t,\infty)} \in \operatorname{ran} A_\nu^\star$. Therefore, the information bound is given by

$$\int (\mathcal{F}[1/\varphi_\nu(-\bullet)] * \mathbb{1}_{[t,\infty)})^2\,\mathrm{d}P_\nu,$$

which can be understood via definition (4.14) or equivalently as the limit $\lim_{n\to\infty} \|(A_\nu^\star)^{-1}\chi_n\|^2_{L^2(P_\nu)}$. For $\alpha\Delta > 1/2$ Neumann and Reiß (2009) show that $\nu([t,\infty))$ cannot be estimated with $\sqrt{n}$-rate.

This example shows the importance of the pseudo-locality for the devolution operator which was discussed by Nickl and Reiß (2012) in detail: If the singularity of the pointwise limit ψ as in (4.14) and the singularity of the distribution P_ν would coincide, the $L^2(P_\nu)$-norm of any approximating sequence $(A_\nu^\star)^{-1}\chi_n$ would diverge such that $\widetilde{\chi}_\nu$ cannot be an element of $\operatorname{ran} A_\nu^\star$ by Proposition 2.11. A simple example is given by the convolution of a Gamma process and a Poisson process (cf. Nickl and Reiß, 2012, Sect. 3.2).

For $\delta > 0$ and l_δ the largest integer which is strictly smaller than δ, let $C^\delta(\mathbb{R})$ denote the set of functions f that possess l_δ continuous derivatives with $f^{(l_\delta)}$ being $(\delta - l_\delta)$-Hölder continuous. We will show that for suitable regularity $\delta > 0$ the class

$$Z^\delta(\mathbb{R}) := \Big\{\zeta = \zeta^s + \zeta^c \Big| (1+|x|^{-1})\zeta^s(x) \in H^\delta(\mathbb{R}), \\ \zeta^c \in C^{\delta+\varepsilon}(\mathbb{R}) \text{ for some } \varepsilon > 0, \zeta^c(0) = 0\Big\}$$

intersected with $L^1(\nu) \cap L^2(\nu)$ is a subset of $\operatorname{ran} A_\nu^\star$.

Example 4.9 (Generalized distribution function)**.** Recall that the generalized distribution function of ν corresponds to the functionals $\zeta_t := \mathbb{1}_{(-\infty,t]}$ for $t < 0$ and $\zeta_t := \mathbb{1}_{[t,\infty)}$ for $t > 0$. It is easy to check that ζ_t can be decomposed for all $t \neq 0$ in a way such that it is contained in $Z^\delta(\mathbb{R})$ for any $\delta < 1/2$. For instance, write $\mathbb{1}_{[t,\infty)} = \zeta_t^s + \zeta_t^c$ with $\zeta_t^s(x) := e^{t-x}\mathbb{1}_{[t,\infty)}(x)$ and $\zeta_t^c(x) := (1-e^{t-x})\mathbb{1}_{[t,\infty)}(x)$ for $t > 0$. Then ζ_t^s is a translation of the gamma density γ_1 such that its Fourier transform decays with polynomial rate one. The factor $(1+|x|^{-1})$ is harmless since ζ_t^s equals zero around the origin. Moreover, ζ_t^c is Lipschitz continuous. On the negative half line an analogous decomposition applies.

For this analytic description of the range of the adjoint score operator, we apply the approach by Nickl and Reiß (2012) as well as Söhl and Trabs (2012). They study the deconvolution operator $\mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)]$ as Fourier multiplier on Besov spaces. We suppose that the Lévy process L with jump density ν satisfies the following.

Assumption B. For some $\beta > 0$ assume for all $u \in \mathbb{R}$

(i) $|\varphi_\nu(u)| \gtrsim (1+|u|)^{-\beta}$ and

(ii) $x\nu$ has a bounded Lebesgue density with $|\mathcal{F}[x\nu](u)| \lesssim (1+|u|)^{-1}$.

This assumption is satisfied by the gamma process discussed in Example 4.8 and in view of Lemma 2.1 by Trabs (2013) for much larger class of Lévy processes as well, including self-decomposable processes.

Proposition 4.10. *Let the finite variation Lévy process (L_t) with $\nu \in \Theta$ satisfy Assumption B for some $\beta > 0$. Then for any $\zeta \in Z^\beta(\mathbb{R}) \cap L^1(\nu) \cap L^2(\nu)$ it holds $\zeta \in \operatorname{ran} A_\nu^\star$ with $(A_\nu^\star)^{-1}\zeta = \mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)] * \zeta$.*

To prove this proposition, we combine the analysis of Lévy processes by Nickl and Reiß (2012) with the insights on the interplay of the smoothness of ζ and the decay of the deconvolution operator by Söhl and Trabs (2012) and with the characterization in Proposition 2.11. The proof is postponed to Section 6.5.

The formula $(A_\nu^\star)^{-1}\zeta = \mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)] * \zeta$ can be either understood in distributional sense or as the limit of an approximating sequence as illustrated in Example 4.8. Applying Theorem 2.7 and Proposition 4.10 on Example 4.9, we get the following convolution theorem.

Theorem 4.11. *Let the finite variation Lévy process (L_t) with $\nu \in \Theta$ satisfy Assumption B for some $\beta < 1/2$. Then the limit distribution of any regular estimator of the generalized distribution function $(\mathbb{R}\setminus\{0\})^d \ni (t_1,\dots,t_d) \mapsto$*

$(\langle\zeta_{t_i},\nu\rangle,\dots,\langle\zeta_{t_d},\nu\rangle)$ is a convolution $\mathcal{N}(0,\Sigma)*M$ for some Borel probability measure M and with the covariance matrix $\Sigma\in\mathbb{R}^{d\times d}$ given by

$$\Sigma_{i,j}=\frac{1}{\Delta^2}\int_{\mathbb{R}}\Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\nu(-\bullet)}\Big]*\zeta_{t_i}\Big)\Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\nu(-\bullet)}\Big]*\zeta_{t_j}\Big)\,\mathrm{d}P_\nu$$

for $i,j\in\{1,\dots,d\}$.

In the situations of Corollary 4.5 and Theorem 4.11 the estimator constructed by Nickl and Reiß (2012) is therefore efficient.

Remark 4.12. Let us finish the considerations of Section 3 for the deconvolution model $Y_j=X_j+\varepsilon_j\sim P_\nu=\nu*\mu$. If the characteristic function of the error distribution μ satisfies for some $\beta>0$

$$|\varphi_\varepsilon(u)|>0\quad\text{and}\quad|(1/\varphi_\varepsilon)'(u)|\lesssim(1+|u|)^{\beta-1},\quad u\in\mathbb{R},\tag{4.15}$$

Lemma 5(i) by Söhl and Trabs (2012) shows that φ_ε^{-1} is a Fourier multiplier on Besov spaces. Therefore, an analogous result as Proposition 4.10 applies in the deconvolution setup which can be combined with Theorem 3.2. We can recover Theorem 4 by Söhl and Trabs (2012) under weaker assumptions on the distributions of ν and μ: If the distribution ν of X_j possesses a Lebesgue density and if (4.15) is satisfied for some $\beta>0$, then the asymptotic variance of every regular estimator of the linear function $\chi(\nu)=\int\zeta\,\mathrm{d}\nu$ for some $\zeta\in Z^\delta(\mathbb{R}),\delta>\beta$, is bounded from below by

$$\int\Big(\mathcal{F}^{-1}\Big[\frac{1}{\varphi_\varepsilon(-\bullet)}\Big]*\zeta\Big)^2\,\mathrm{d}P_\nu-\Big(\int\zeta\,\mathrm{d}\nu\Big)^2.$$

As we saw in Example 4.9, we need $\delta<1/2$ to apply this information bound to distribution function estimation. Therefore, we need $|\varphi_\varepsilon(u)|\gtrsim(1+|u|)^{-\beta}$ for $\beta<1/2$ which coincides with the classical condition under which the distribution function can be estimated with the parametric rate, cf. Fan (1991).

5 Extension to Banach space valued functions

So far we considered $\mathbb{R}^d$-valued derived parameters χ. The aim of the section is to generalize Theorem 2.7 to functions $\chi:\Theta\to\mathbb{B}$ for a Banach space $(\mathbb{B},\|\bullet\|)$. As pointed out by van der Vaart (1988) for the estimation of parameters in infinite dimensional spaces efficiency means essentially efficiency for the marginals plus tightness of the limit law of the sequence of estimator.

Let $(\mathcal{X}_n,\mathcal{A}_n,P_{n,\vartheta}:\vartheta\in\Theta)$ be a locally regular indirect model at $\vartheta\in\Theta$ with respect to the tangent set $\dot\Theta_\vartheta$ and with generalized score operator $A_\vartheta:\operatorname{lin}\dot\Theta_\vartheta\to H_\vartheta$ for some Hilbert space $(H_\vartheta,\langle\bullet,\bullet\rangle_H)$. First, we have to generalize the notion of regularity to Banach space valued functions. The derived parameter $\chi:\Theta\to\mathbb{B}$ is *pathwise differentiable* at $\vartheta\in\Theta$ with respect to the tangent set $\dot\Theta_\vartheta$ if for all $b\in\dot\Theta_\vartheta$ with associated path $[0,\tau)\ni t\mapsto\vartheta_t$

$$t^{-1}\big(\chi(\vartheta_t)-\chi(\vartheta)\big)\to\dot\chi_\vartheta b$$

holds true for some continuous linear map $\dot\chi_\vartheta:\operatorname{lin}\dot\Theta_\vartheta\to\mathbb{B}$. The gradient of $\dot\chi_\vartheta$ is then defined as by van der Vaart (1991) using the dual space $\mathbb{B}^\star$, which

is the space of all continuous linear functions $b^\star : \mathbb{B} \to \mathbb{R}$. The composition $b^\star \circ \dot{\chi}_\vartheta : \operatorname{lin} \dot{\Theta}_\vartheta \to \mathbb{R}$ is linear and continuous and thus it can be represented by some $\widetilde{\chi}_{\vartheta,b^\star} \in \overline{\operatorname{lin}} \dot{\Theta}_\vartheta$:

$$(b^\star \circ \dot{\chi}_\vartheta) b = \langle \widetilde{\chi}_{\vartheta,b^\star}, b \rangle_\vartheta \quad \text{for all } b \in \operatorname{lin} \dot{\Theta}_\vartheta.$$

Similarly, the parameter $\psi_n(P_{n,\vartheta}) = \chi(\vartheta)$ is *regular* if (2.7) holds for some continuous linear map $\dot{\psi}_\vartheta : H \to \mathbb{B}$. The efficient influence functions $\widetilde{\psi}_{\vartheta,b^\star} \in \overline{\operatorname{ran}} A_\vartheta$ are defined by $(b^\star \circ \dot{\psi}_\vartheta) h = \langle \widetilde{\psi}_{\vartheta,b^\star}, h \rangle_H$ for all $h \in \operatorname{ran} A_\vartheta$. The sequence of estimators $T_n : \mathcal{X}_n \to \mathbb{B}$ is called *regular* at $\vartheta \in \Theta$ with respect to the rate r_n if there is a fixed tight Borel probability measure L on $\mathbb{B}$ such that for all $b \in \dot{\Theta}_\vartheta$ with corresponding submodel $t \mapsto P_{n,\vartheta_t}$

$$\frac{1}{r_n}\big(T_n - \chi(\vartheta_{r_n})\big) \overset{P_{n,\vartheta_{r_n}}}{\Longrightarrow} L,$$

where weak convergence is defined in terms of outer probability to avoid measurability problems, that is,

$$\mathbb{E}^*_{P_{n,\vartheta_{r_n}}} \big[f\big(r_n^{-1}(T_n - \chi(\vartheta_{r_n}))\big] \to \int f \, \mathrm{d}L$$

for all bounded, continuous function $f : \mathbb{B} \to \mathbb{R}$ (cf. van der Vaart and Wellner, 1996, Def. 1.3.3). Now we can state the following generalization of Theorem 2.7.

Theorem 5.1. *Let $(\mathcal{X}_n, \mathcal{A}_n, P_{n,\vartheta} : \vartheta \in \Theta)$ be a locally regular indirect model at $\vartheta \in \Theta$ with respect to $\dot{\Theta}_\vartheta$ and let $\chi : \Theta \to \mathbb{B}$ be pathwise differentiable at ϑ with respect to $\dot{\Theta}_\vartheta$. Then the sequence $\psi_n : \mathcal{P}_n \to \mathbb{B}$ is regular at ϑ relative to $\dot{\Theta}_\vartheta$ if and only if*

$$\widetilde{\chi}_{\vartheta,b^\star} \in \operatorname{ran} A^\star_\vartheta \quad \textit{for all} \quad b^\star \in \mathbb{B}^\star. \tag{5.1}$$

In this case the efficient influence functions is given by the Moore–Penrose pseudoinverse $\widetilde{\psi}_{\vartheta,b^\star} = (A^\star_\vartheta)^\dagger \widetilde{\chi}_{\vartheta,b^\star}$. For any regular sequence of estimators T_n the limit distribution L of $r_n^{-1}(T_n - \chi(\vartheta_{r_n}))$ is given by the law of a sum $N + W$ of independent, tight, Borel measurable random elements in $\mathbb{B}$ such that

$$b^\star N \sim \mathcal{N}\big(0, \|\widetilde{\psi}_{\vartheta,b^\star}\|_H^2\big) = \mathcal{N}\big(0, \|(A^\star_\vartheta)^\dagger \widetilde{\chi}_{\vartheta,b^\star}\|_H^2\big).$$

The proof of this Theorem is analogous to Theorem 2.7 with an additional application of Lemma A.2 by van der Vaart (1991). We omit the details.

Remark 5.2. The type of regularity which we used for the parameters $\psi_n(P_{n,\vartheta})$ to apply the convolution theorem is quite strong because the derivative $\dot{\psi}_\vartheta$ has to be continuous with respect to the norm topology of $\mathbb{B}$. Necessarily, the range condition (5.1) has to hold for all $b^\star \in \mathbb{B}^\star$ which may fail if the dual space is large. This problem can be solved by using a weaker topology on $\mathbb{B}$ which is generated by a subspace $B' \subseteq \mathbb{B}^\star$ as shown by van der Vaart (1988, Sect. 3).

To show tightness of the limit distribution may be a difficult problem for inverse problems. In the i.i.d. setting the theory of smoothed empirical processes by Giné and Nickl (2008) turns out to be useful as the following example shows.

Example 5.3. Let us consider again the estimation of the distribution function $\mathbb{R} \ni t \mapsto \nu((-\infty, t])$ in the deconvolution model in Section 3. On the whole real line the parameter $\chi(\nu) = (\nu((-\infty, t]))_{t\in\mathbb{R}}$ maps into the space of bounded functions $\mathbb{B} = \ell^\infty(\mathbb{R})$ which is equipped with the supremums norm. Under suitable conditions Söhl and Trabs (2012) have shown a uniform central limit theorem in $\ell^\infty(\mathbb{R})$ for the canonical plug-in kernel estimator. In particular, their limit distribution is tight and the finite dimensional distributions coincide the the information bound from Theorem 3.2 and Remark 4.12. In the Lévy model Nickl and Reiß (2012) have proved a Donsker theorem for the estimation of the generalized distribution function in $\ell^\infty((-\infty, -\delta] \cup [\delta, \infty))$ for $\delta > 0$ which proves tightness of the limit process in Theorem 4.11.

6 Remaining proofs

6.1 Proof of Lemma 2.1

As discussed above the law of the white noise $\varepsilon\dot{W} = y_{\varepsilon,0}$ is a symmetric (zero mean) Gaussian measure on $(E, \mathscr{B}(E))$. Its (unique) reproducing kernel Hilbert space is Y with norm $\|\bullet\|_\varepsilon := \varepsilon^{-1}\|\bullet\|_Y$. To see this, note that every functional $\varphi \in E^\star$ can be represented by $\varphi = \langle y, \bullet\rangle_Y = \langle \widetilde{y}, \bullet\rangle_\varepsilon$ for some $y \in Y$ and $\widetilde{y} = \varepsilon^2 y$. Then

$$\varphi(\varepsilon\dot{W}) \sim \mathcal{N}(0, \|\widetilde{y}\|_\varepsilon^2) = \mathcal{N}(0, \varepsilon^2\|y\|_Y^2).$$

The Cameron–Martin formula (Da Prato and Zabczyk, 1992, Prop. 2.24) yields that $P_{\varepsilon,x}$ and $P_{\varepsilon,0}$ are equivalent measures on $(E, \mathscr{B}(E))$ with Radon-Nikodym derivative

$$\frac{\mathrm{d}P_{\varepsilon,x}}{\mathrm{d}P_{\varepsilon,0}}(y_{\varepsilon,0}) = \exp\Big(\langle y_{\varepsilon,0}, K(x)\rangle_\varepsilon - \frac{1}{2}\|K(x)\|_\varepsilon^2\Big) \quad P_{\varepsilon,0} - a.s.$$

and thus

$$\begin{aligned}
\frac{\mathrm{d}P_{\varepsilon,0}}{\mathrm{d}P_{\varepsilon,\vartheta}}(y_{\varepsilon,\vartheta}) &= \exp\Big(-\langle y_{\varepsilon,\vartheta}, K(\vartheta)\rangle_\varepsilon + \frac{1}{2}\|K(\vartheta)\|_\varepsilon^2\Big) \quad \text{and} \\
\frac{\mathrm{d}P_{\varepsilon,x}}{\mathrm{d}P_{\varepsilon,\vartheta}}(y_{\varepsilon,\vartheta}) &= \frac{\mathrm{d}P_{\varepsilon,x}}{\mathrm{d}P_{\varepsilon,0}}(y_{\varepsilon,\vartheta})\frac{\mathrm{d}P_{\varepsilon,0}}{\mathrm{d}P_{\varepsilon,\vartheta}}(y_{\varepsilon,\vartheta}) \\
&= \exp\Big(\langle \varepsilon\dot{W}, K(x) - K(\vartheta)\rangle_\varepsilon - \frac{1}{2}\|K(x) - K(\vartheta)\|_\varepsilon^2\Big) \\
&= \exp\Big(\Big\langle \dot{W}, \frac{K(x) - K(\vartheta)}{\varepsilon}\Big\rangle_Y - \frac{1}{2}\Big\|\frac{K(x) - K(\vartheta)}{\varepsilon}\Big\|_Y^2\Big) \quad P_{\varepsilon,\vartheta} - a.s.
\end{aligned}$$

6.2 Proof of Proposition 4.1

Without loss of generality we assume $\Delta = 1$ in this and the following subsections.

For any $b \in \dot{\Theta}_\nu$ let $P_\nu * (b\nu) = (P_\nu * (b\nu))^a + (P_\nu * (b\nu))^\perp$ be the Lebesgue decomposition of $P_\nu * (b\nu)$ with respect to P_ν, that is the first and second measure are absolutely continuous and singular with respect to P_ν, respectively. $\frac{\mathrm{d}P_\nu * (b\nu)}{\mathrm{d}P_\nu}$ is then defined as the Radon–Nikodym density of $(P_\nu * (b\nu))^a$ with respect to P_ν. Therefore, $A_\nu b = \frac{\mathrm{d}P_\nu * (b\nu)}{\mathrm{d}P_\nu} - \int b\,\mathrm{d}\nu$ is well-defined without further assumptions.

In a first step we will show L^2-differentiability of the submodels corresponding to some direction $b \in \dot{\Theta}_\vartheta \cap L^\infty(\nu)$. The extension to the whole tangent set is

proved in the second step. In the last step we will see that even $P_\nu * (b\nu) \ll P_\nu$ holds true.

Step 1: Let $b \in L^1(\nu) \cap L^\infty(\nu)$. We will show that the associated model $[0,1) \ni t \mapsto P_{\nu_t}$ is L^2-differentiable at 0 with derivative $A_\nu b \in L^2_0(P_\nu)$ as defined in (4.6), that is

$$\int \Big(\frac{\mathrm{d}P_{\nu_t} - \mathrm{d}P_\nu}{t\,\mathrm{d}P_\nu} - A_\nu b \Big)^2 \mathrm{d}P_\nu \to 0 \quad \text{as} \quad t \to 0. \tag{6.1}$$

Note that $\frac{\mathrm{d}P_{\nu_t}}{\mathrm{d}P_\nu} \in L^2(P_\nu)$ by (4.7), which follows by a similar argument as the one following (6.2). Applying Proposition 1.199 by Witting (1985), the L^2-regularity (6.1) implies the proposed Hellinger differentiability for all $b \in \dot{\Theta}_\nu \cap L^\infty(\nu)$.

Defining the measure ν_t^c via the density $\frac{\mathrm{d}\nu_t^c}{\mathrm{d}\nu} = (k(tb) - 1) =: f_{\nu_t^c}$, we write as a consequence of (4.5) and Remark 27.3 by Sato (1999)

$$P_{\nu_t} = e^{-\nu_t^c(\mathbb{R})} \sum_{k=0}^\infty \frac{1}{k!} (\nu_t^c)^{*k} * P_\nu$$

Arranging terms

$$\begin{aligned}
&\int \Big(\frac{\mathrm{d}P_{\nu_t} - \mathrm{d}P_\nu}{t\,\mathrm{d}P_\nu} - A_\nu b \Big)^2 \mathrm{d}P_\nu \\
&= \int \Big(\frac{\mathrm{d}P_{\nu_t} - \mathrm{d}P_\nu - t\,\mathrm{d}(P_\nu * (b\nu)) + t\big(\int b\nu\big)\,\mathrm{d}P_\nu}{t\,\mathrm{d}P_\nu} \Big)^2 \mathrm{d}P_\nu \\
&= \int \Bigg(\frac{(e^{-\nu_t^c(\mathbb{R})} - 1 + t\int b\,\mathrm{d}\nu)\,\mathrm{d}P_\nu + \mathrm{d}\big(((e^{-\nu_t^c(\mathbb{R})} f_{\nu_t^c} - tb)\nu) * P_\nu\big)}{t\,\mathrm{d}P_\nu} \\
&\qquad + \frac{e^{-\nu_t^c(\mathbb{R})} \sum_{k=2}^\infty (k!)^{-1}\,\mathrm{d}\big((\nu_t^c)^{*k} * P_\nu\big)}{t\,\mathrm{d}P_\nu} \Bigg)^2 \mathrm{d}P_\nu,
\end{aligned}$$

all three terms in the numerator turn out to be of order t^2. Note that we can dominate $|f_{\nu_t^c}| \leqslant t|b|$ as well as $|f_{\nu_t^c} - tb| \leqslant t^2|b^2|$ by $|k(y) - 1| \leqslant |y|$ and $|k(y) - 1 - y| \leqslant y^2$. Therefore,

$$\begin{aligned}
\Big| e^{-\nu_t^c(\mathbb{R})} - 1 + t\int b\,\mathrm{d}\nu \Big| &= \Big| \sum_{k\geqslant 2} \frac{(-\nu_t^c(\mathbb{R}))^k}{k!} - \Big(\nu_t^c(\mathbb{R}) - t\int b\,\mathrm{d}\nu \Big) \Big| \\
&\leqslant \sum_{k\geqslant 2} \frac{t^k \|b\|^k_{L^1(\nu)}}{k!} + t^2 \|b\|^2_{L^2(\nu)} \\
&\leqslant \big(e^{\|b\|_{L^1(\nu)}} + \|b\|^2_{L^2(\nu)} \big) t^2, \\
\big| e^{-\nu_t^c(\mathbb{R})} f_{\nu_t^c} - tb \big| &= \Big| \big(e^{-\nu_t^c(\mathbb{R})} - 1 \big) f_{\nu_t^c} + f_{\nu_t^c} - tb \Big| \\
&\leqslant t^2 \big(e^{\|b\|_{L^1(\nu)}} |b| + |b|^2 \big), \\
\Big| e^{-\nu_t^c(\mathbb{R})} \sum_{k=2}^\infty \frac{(\nu_t^c)^{*k}}{k!} \Big| &\leqslant t^2 \sum_{k=2}^\infty \frac{(|b|\nu)^{*k}}{k!}.
\end{aligned}$$

Hence, we estimate

$$\begin{aligned}
&\int \Big(\frac{\mathrm{d}P_{\nu_t} - \mathrm{d}P_\nu}{t\,\mathrm{d}P_\nu} - A_\nu b\Big)^2 \,\mathrm{d}P_\nu \\
\leqslant & t^2 \int \Bigg(\frac{(e^{\|b\|_{L^1(\nu)}} + \|b\|^2_{L^2(\nu)})\,\mathrm{d}P_\nu + \big(e^{\|b\|_{L^1(\nu)}}|b| + |b^2|\big)\,\mathrm{d}(\nu * P_\nu)}{\mathrm{d}P_\nu} \\
& \qquad + \frac{\sum_{k=2}^\infty (k!)^{-1}\,\mathrm{d}\big((|b|\nu)^{*k} * P_\nu\big)}{\mathrm{d}P_\nu}\Bigg)^2 \mathrm{d}P_\nu \\
\leqslant & t^2 \big(e^{\|b\|_{L^1(\nu)}} + \|b\|^2_{L^2(\nu)}\big)^2 \int \Bigg(\frac{\mathrm{d}\Big(\sum_{k=0}^\infty ((|b| + |b|^2)\nu)^{*k}/(k!)\Big) * P_\nu}{\mathrm{d}P_\nu}\Bigg)^2 \mathrm{d}P_\nu.
\end{aligned}$$

Introducing an infinite divisible distribution μ without diffusion component, without drift and with finite jump measure $(|b| + |b|^2)\nu$, the previous line can be written as

$$t^2 \underbrace{\big(e^{\|b\|_{L^1(\nu)}} + \|b\|^2_{L^2(\nu)}\big)^2 e^{2\|b\|_{L^1(\nu)} + 2\|b\|^2_{L^2(\nu)}}}_{=:C(\|b\|_{L^1(\nu)}, \|b\|_{L^2(\nu)})} \int \Big(\frac{\mathrm{d}(\mu * P_\nu)}{\mathrm{d}P_\nu}\Big)^2 \mathrm{d}P_\nu \tag{6.2}$$

Therefore, the assertion holds true provided the Hellinger integral $H_2(\mu * P_\nu, P_\nu) = \int (\,\mathrm{d}(\mu * P_\nu)/\,\mathrm{d}P_\nu)^2\,\mathrm{d}P_\nu$ is finite. To show this, we apply the bound of Renyi's distance R_2 for infinite divisible distributions by Liese (1987, Thm. 2.6). Using that both distributions, $\mu * P_\nu$ and P_ν, have the same drift and have finite variation, we obtain (for $\alpha = 2$)

$$\tfrac{1}{2}\log H_2(\mu * P_\nu, P_\nu) = R_2(\mu * P_\nu, P_\nu) \leqslant \tfrac{1}{2}\chi^2\big((|b| + b^2 + 1)\nu, \nu\big)$$

where the χ^2-distance of the jump measures satisfies

$$\begin{aligned}
\chi^2\big((|b| + b^2 + 1)\nu, \nu\big) :=& \frac{1}{2}\int \Big(\frac{\mathrm{d}((|b| + b^2 + 1)\nu)}{\mathrm{d}\nu} - 1\Big)^2 \,\mathrm{d}\nu \\
=& \frac{1}{2}\int \big(|b| + b^2\big)^2\,\mathrm{d}\nu \leqslant \frac{1}{2}\big(1 + \|b\|^2_{L^\infty(\nu)}\big)\|b\|^2_{L^2(\nu)} < \infty.
\end{aligned}$$

The combination with the bound (6.2) yields

$$\int \Big(\frac{\mathrm{d}P_{\nu_t} - \mathrm{d}P_\nu}{t\,\mathrm{d}P_\nu} - A_\nu b\Big)^2 \mathrm{d}P_\nu \leqslant t^2 C\big(\|b\|_{L^1(\nu)}, \|b\|_{L^2(\nu)}\big) e^{\frac{1}{2}(1 + \|b\|^2_{L^\infty(\nu)})\|b\|^2_{L^2(\nu)}}.$$

As $t \to 0$ this upper bound converges to zero which shows the L^2-differentiability. We conclude $\int A_\nu b\,\mathrm{d}P_\nu = 0$ for all $b \in \dot{\Theta}_\nu \cap L^\infty(\nu)$.

Step 2: To show continuity of $A_\nu|_{L^1(\nu)\cap L^\infty(\nu)}$, let $\varepsilon > 0$ and $b \in L^1(\nu) \cap L^\infty(\nu)$ with $\|b\|^2_{L^2(\nu)} < \varepsilon$. By (6.1), $\frac{1}{2}A_\nu b$ is the L^2-limit of $t^{-1}(\sqrt{\mathrm{d}P_{\nu_t}} - \sqrt{\mathrm{d}P_\nu})$ and thus for t small enough

$$\|A_\nu b\|^2_{L^2(P_\nu)} \leqslant \frac{2}{t^2}\int \big(\sqrt{\mathrm{d}P_{\nu_t}} - \sqrt{\mathrm{d}P_\nu}\big)^2 + \varepsilon.$$

As above Theorem 2.6 by Liese (1987) for $\alpha = 1/2$ yields the estimate for the Hellinger distance of the infinite divisible distributions

$$\begin{aligned}
\int \big(\sqrt{\mathrm{d}P_{\nu_t}} - \sqrt{\mathrm{d}P_\nu}\big)^2 &\leqslant 2\Big(1 - \exp\Big(-2\int (\sqrt{\mathrm{d}\nu_t} - \sqrt{\mathrm{d}\nu})^2\Big)\Big) \\
&= 2\Big(1 - \exp\Big(-2\int (\sqrt{k(tb)} - 1)^2\,\mathrm{d}\nu\Big)\Big).
\end{aligned} \tag{6.3}$$

Since $|\sqrt{k(y)}-1| \leqslant |(\sqrt{k(y)}+1)(\sqrt{k(y)}-1)| = |k(y)-1| \leqslant |y|$ and $1-e^{-y} \leqslant |y|$ for $y \in \mathbb{R}$, the previous display can be bounded by

$$2\Big(1-\exp\Big(-2\int (tb)^2\,\mathrm{d}\nu\Big)\Big) \leqslant 4t^2\|b\|^2_{L^2(\nu)}.$$

Because $\varepsilon > 0$ was arbitrary, we conclude $\|A_\nu b\|_{L^2(P_\nu)} \lesssim \|b\|_{L^2(\nu)}$ which is equivalent to the continuity of the linear operator $A_\nu|_{L^1(\nu)\cap L^\infty(\nu)}$. Since $L^1(\nu)\cap L^\infty(\nu)$ is dense in $\dot\Theta_\nu$, there is a unique continuous extension A_ν on $\dot\Theta_\nu$ satisfying $A_\nu b = \frac{\mathrm{d}P_\nu * (b\nu) - \int b\nu\,\mathrm{d}P_\nu}{\mathrm{d}P_\nu}$ for all $b \in L^1(\nu)\cap L^\infty(\nu)$.

Now, for any $b \in \dot\Theta_\nu$ with associated path $t \mapsto \mathrm{d}\nu_t = k(tb)\,\mathrm{d}\nu$ and for any positive null sequence $(t_m)_{m\in\mathbb{N}}$ and let $\widetilde{b}_m \in L^1(\nu)\cap L^\infty(\nu)$ with path $t \mapsto \mathrm{d}\widetilde{\nu}_t := k(t\widetilde{b}_m)\,\mathrm{d}\nu$ such that $\|b-\widetilde{b}_m\|_{L^2(\nu)} \to 0$ and $\|\widetilde{b}_m\|_{L^\infty(\nu)} = o(|\log t_m|^{1/2})$ as $m\to\infty$. Then

$$\begin{aligned}
&t_m^{-2}\int\Big(\sqrt{\frac{\mathrm{d}P_{\nu_{t_m}}}{\mathrm{d}P_\nu}}-1-\frac{t_m}{2}A_\nu b\Big)^2\,\mathrm{d}P_\nu\\
\leqslant&\frac{3}{t_m^2}\int\Big(\sqrt{\frac{\mathrm{d}P_{\nu_{t_m}}}{\mathrm{d}P_\nu}}-\sqrt{\frac{\mathrm{d}P_{\widetilde{\nu}_{t_m}}}{\mathrm{d}P_\nu}}\Big)^2\,\mathrm{d}P_\nu+\frac{3}{t_m^2}\int\Big(\sqrt{\frac{\mathrm{d}P_{\widetilde{\nu}_{t_m}}}{\mathrm{d}P_\nu}}-1-\frac{t_m}{2}A_\nu\widetilde{b}_m\Big)^2\,\mathrm{d}P_\nu\\
&\quad+\frac{3}{4}\int\Big(A_\nu\widetilde{b}_m-A_\nu b\Big)^2\,\mathrm{d}P_\nu
\end{aligned}\tag{6.4}$$

The frist term is the Hellinger distance between P_{ν_t} and $P_{\widetilde{\nu}_t}$ which can be bounded as in (6.3)

$$\begin{aligned}
&t_m^{-2}\int\Big(\sqrt{\frac{\mathrm{d}P_{\nu_{t_m}}}{\mathrm{d}P_\nu}}-\sqrt{\frac{\mathrm{d}P_{\widetilde{\nu}_{t_m}}}{\mathrm{d}P_\nu}}\Big)^2\,\mathrm{d}P_\nu\\
&\quad\leqslant 2t_m^{-2}\Big(1-\exp\Big(-2\int\big(\sqrt{\mathrm{d}\nu_{t_m}}-\sqrt{\mathrm{d}\widetilde{\nu}_{t_m}}\big)^2\Big)\Big)\\
&\quad= 2t_m^{-2}\Big(1-\exp\Big(-2\int\Big(\sqrt{k(t_m b(x))}-\sqrt{k(t_m\widetilde{b}_m(x))}\Big)^2\nu(\,\mathrm{d}x)\Big)\Big).
\end{aligned}$$

An easy calculation shows $|(\sqrt{k})'(x)| \leqslant 1$ for all $x\in\mathbb{R}$ and thus the above display can be bounded by the mean value theorem

$$2t_m^{-2}\Big(1-\exp\big(-2t_m^2\|b-\widetilde{b}_m\|^2_{L^2(\nu)}\big)\Big) \leqslant 4\|b-\widetilde{b}_m\|^2_{L^2(\nu)} \to 0.$$

The second term in (6.4) converges to zero according to Step 1 provided $\|\widetilde{b}\|_{L^\infty(\nu)} = o(|\log t|^{1/2})$. Applying continuity of A_ν, the third term in (6.4) vanishes as well. Therefore, we have shown that $A_\nu b$ is the Hellinger derivative of P_{ν_t} for all $b\in\dot\Theta_\nu$.

Step 3: Finally, we will show $P_\nu * (b\nu) \ll P_\nu$ for all $b \in \dot\Theta_\nu \cap L^\infty(\nu)$. By construction $|b| \in \dot\Theta_\nu\cap L^\infty(\nu)$, too. Let $P_\nu*(|b|\nu) = (P_\nu*(|b|\nu))^a + (P_\nu*(|b|\nu))^\perp$ be Lebesgue's decomposition with respect to P_ν where both measures can be chosen to be nonnegative and finite. According to Step 1, $\int A_\nu|b|\,\mathrm{d}P_\nu = 0$ which yields together with the nonnegativity of the measures and Fubini's theorem

$$\int |b|\,\mathrm{d}\nu = \int \frac{\mathrm{d}(P_\nu*(|b|\nu))}{\mathrm{d}P_\nu}\,\mathrm{d}P_\nu = \int \mathrm{d}(P_\nu*(|b|\nu))^a \leqslant \int \mathrm{d}P_\nu*(|b|\nu) = \int |b|\,\mathrm{d}\nu.$$

We conclude $(P_\nu * (|b|\nu))^a = P_\nu * (|b|\nu)$ or equivalently $P_\nu * (|b|\nu) \ll P_\nu$. Now for any event $A \in \mathscr{B}(\mathbb{R})$ with $P_\nu * (|b|\nu)(A) = 0$ we have

$$\begin{aligned}|P_\nu * (b\nu)(A)| &= \Big| \int_{\mathbb{R}^2} \mathbb{1}_A(x+y) b(x) \nu(\,\mathrm{d}x) P_\nu(\,\mathrm{d}y) \Big| \\ &\leqslant \int_{\mathbb{R}^2} \mathbb{1}_A(x+y)|b(x)|\nu(\,\mathrm{d}x)P_\nu(\,\mathrm{d}y) = P_\nu * (|b|\nu)(A) = 0.\end{aligned}$$

Consequently, $P_\nu * (b\nu) \ll P_\nu * (|b|\nu) \ll P_\nu$. □

6.3 Proof of Lemma 4.3

First, we show $\nu^{*l} * P_\nu \ll P_\nu$ for any $l \in \mathbb{N}$. Let $A \in \mathscr{B}(\mathbb{R})$ satisfy $P_\nu(A) = 0$. (4.8) yields

$$0 = e^{-\Delta\lambda} \sum_{k=0}^{\infty} \frac{\Delta^k}{k!} \int \mathbb{1}_A(x + \Delta\gamma)\,\mathrm{d}\nu^{*k}(\,\mathrm{d}x)$$

and thus $\int \mathbb{1}_A(x+\Delta\gamma)\,\mathrm{d}\nu^{*k}(\,\mathrm{d}x) = 0$ for all $k \in \mathbb{N}$. But this implies by linearity of the convolution that $\nu^{*l} * P_\nu(A) = 0$.

To see that $A_\nu^\star$ is well defined on equivalence classes with respect to P_ν zero sets, note that $\nu * P_\nu \ll P_\nu$ implies that for any function g with $g(x) = 0$ for P_ν-a.e. $x \in \mathbb{R}$ it holds $P_\nu(-\bullet) * g(y) = 0$ for ν-a.e. $y \in \mathbb{R}$. It remains to show $A_\nu^\star g \in \mathcal{H}$ for $g \in \mathcal{G}$. For any $g \in L^\infty(P_\nu)$ there is a set $A \in \mathscr{B}(\mathbb{R})$ with $P_\nu(A) = 0$ such that $g(y) \leqslant C$ for some constant $C > 0$ and for all $y \notin A$. Using

$$0 = \nu^{*l} * P_\nu(A) = \int\int \mathbb{1}_{A-\{x\}}(y) P_\nu(\,\mathrm{d}y)\nu^{*l}(\,\mathrm{d}x)$$

we infer $P_\nu(A-\{x\}) = 0$ for ν^{*l}-a.e. $x \in \mathbb{R}$ and therefore $P_\nu(-\bullet)*g(y) = \int g(x+y)P_\nu(\,\mathrm{d}x)$ is bounded by C for ν^{*l}-a.e. $y \in \mathbb{R}$. Hence, $\|P_\nu(-\bullet) * g\|_{L^\infty(\nu^{*l})} \leqslant C$ for any $l \in \mathbb{N}$. □

6.4 Proof of Lemma 4.6

(i) We will determine the adjoint score operator and its inverse on the subsets $\mathcal{G}$ and $\mathcal{H}$ as defined in (4.11). In the case of infinite jump activity the application of Fubini's theorem in (4.10) holds as well. Hence, the adjoint score operator on $\mathcal{G}$ is given by $A_\nu^\star g = P_\nu(-\bullet)*g$. To verify that $A_\nu^\star|_{\mathcal{G}}$ is well-defined, we note first that by the Sobolev embedding any $g \in \mathcal{G}$ has a version in $C^1(\mathbb{R})$. Throughout we can identify g with this smooth version. Then, we obtain $A_\nu^\star g(0) = \int g\,\mathrm{d}P_\nu = 0$ and $\|(A_\nu^\star g)^{(l)}\|_\infty = \|P_\nu(-\bullet) * (g^{(l)})\|_\infty \leqslant \|g^{(l)}\|_\infty \leqslant \|g\|_{C^1}$ for $l = 0, 1$. Hence, $A_\nu^\star g$ is a bounded function and

$$\begin{aligned}\int |A_\nu^\star g(x)|\nu(x)\,\mathrm{d}x &\leqslant \int \big(\|A_\nu^\star g\|_\infty \wedge (|x|\|(A_\nu^\star g)'\|_\infty)\big)\,\mathrm{d}\nu(x) \\ &\leqslant \|g\|_{C^1} \int (1 \wedge |x|)\,\mathrm{d}\nu(x).\end{aligned}$$

A similar estimate holds for $L^2(\nu)$. Therefore, $A_\nu^\star g \in L^1(\nu) \cap L^\infty(\mathbb{R}) \subseteq \dot{\Theta}_\nu$. Owing to $\|\varphi_\nu\|_\infty \leqslant 1$, it holds for any $s > 0$

$$\big\|(1+|u|^2)^{s/2}\,\mathcal{F}[A_\nu^\star g](u)\big\|_{L^2} = \big\|(1+|u|^2)^{s/2}\varphi_\nu(-u)\,\mathcal{F}\,g(u)\big\|_{L^2} \leqslant \|g\|_{H^s} < \infty.$$

We conclude $\operatorname{ran} A_\nu^\star|_{\mathcal{G}} \subseteq \mathcal{H}$.

Let us show now that the inverse adjoint score operator as given in (4.13) is well-defined on $\mathcal{H}$. Applying the assumption $|\varphi_\nu(u)| \gtrsim (1+|u|)^{-\beta}$, we obtain for all $b \in \mathcal{H}$ and $s > 0$

$$\begin{aligned}\big\|(1+|u|^2)^{s/2}\,\mathcal{F}[(A_\nu^\star)^{-1}b](u)\big\|_{L^2} &= \big\|(1+|u|^2)^{s/2}\,\mathcal{F}\,b(u)/\varphi_\nu(-u)\big\|_{L^2}\\ &\lesssim \big\|(1+|u|^2)^{(s+\beta)/2}\,\mathcal{F}\,b(u)\big\|_{L^2}\\ &\leqslant \|b\|_{H^{s+\beta}} < \infty.\end{aligned}$$

Therefore, $(A_\nu^\star)^{-1}b \in H^\infty(\mathbb{R})$ and the Sobolev embedding yields $\|(A_\nu^\star)^{-1}b\|_{L^2(P_\nu)} \leqslant \|(A_\nu^\star)^{-1}b\|_\infty \leqslant \|(A_\nu^\star)^{-1}b\|_{C^s} < \infty$. It remains to verify the condition $\int (A_\nu^\star)^{-1}b\,\mathrm{d}P_\nu = 0$. By construction

$$\begin{aligned}\int (A_\nu^\star)^{-1}b\,\mathrm{d}P_\nu &= \big(((A_\nu^\star)^{-1}b) * P_\nu\big)(0)\\ &= \big(\mathcal{F}^{-1}\Big[\mathcal{F}\,b/\varphi_\nu(-\bullet)\Big] * P_\nu\big)(0) = b(0)\end{aligned}$$

where the last equality is clear in distributional sense and can be shown via integration against test functions. Since $b(0) = 0$ for all $b \in \mathcal{H}$, we conclude $\operatorname{ran}(A_\nu^\star)^{-1}|_{\mathcal{H}} \subseteq \mathcal{G}$.

By construction $g = (A_\nu^\star)^{-1}A_\nu^\star g$ and $b = A_\nu^\star(A_\nu^\star)^{-1}b$ for all $g \in \mathcal{G}, b \in \mathcal{H}$ which proves that $A_\nu^\star|_{\mathcal{G}}$ is a bijection from $\mathcal{G}$ onto $\mathcal{H}$.

(ii) Let us show that $\mathcal{G}$ is dense in $L_0^2(P_\nu)$. Since the Borel σ-field is generated by $\mathcal{E} := \{[a,b] : -\infty < a < b < \infty\}$, the set of indicator functions $\{\mathbb{1}_E : E \in \mathcal{E}\}$ is dense in $L^2(P_\nu)$. Hence, it suffices to approximate in $L^2(P_\nu)$-sense the indicators $\mathbb{1}_E, E \in \mathcal{E}$, by smooth $L^2(\mathbb{R})$-integrable functions. Let $\varepsilon > 0$ be arbitrary and let us denote the distribution function of P_ν by $F(x) := P_\nu((-\infty, x]), x \in \mathbb{R}$. Since F is right continuous with left limits, for all $a < b$ there are $a' < a, b < b'$ such that

$$P_\nu\big((a',b'] \setminus [a,b]\big) = P_\nu\big((a',a)\big) + P_\nu\big((b,b']\big) = F(a-) - F(a') + F(b') - F(b) < \varepsilon.$$

Therefore, for any $A \in \mathcal{E}$ there is a bounded set $B \in \mathscr{B}(\mathbb{R})$ satisfying $A \subseteq B$, $P_\nu(B \setminus A) < \varepsilon$ and the distance between $x \in A$ and $\mathbb{R} \setminus B$ is strictly positive. Consequently, there is some non-negative $\psi \in C^\infty(\mathbb{R})$ with $\psi(x) = 1$ for $x \in A$, $\|\psi\|_\infty < 1$ and $\operatorname{supp}\psi \subseteq B$. Obviously, ψ is contained in $H^\infty(\mathbb{R}) \cap L^2(P_\nu)$ and $\|\mathbb{1}_A - \psi\|_{L^2(P_\nu)} < \sqrt{\varepsilon}$. Since $L_0^2(P_\nu) = (\operatorname{lin} 1)^\perp$ is a closed subspace of $L^2(P_\nu)$, we conclude that $\mathcal{G}$ is dense in $L_0^2(P_\nu)$.

Continuity of $A_\nu^\star$ follows from the continuity of A_ν which was shown in Proposition 4.1. Hence, $A_\nu^\star$ is uniquely given by the continuous extension of $A_\nu^\star|_{\mathcal{G}}$ to $L_0^2(P_\nu)$. □

6.5 Proof of Proposition 4.10

Taking the derivative of the Lévy-Khintchine formula (4.2), we obtain

$$\varphi_\nu'(u) = \varphi_\nu(u)\big(i\gamma + \mathcal{F}[x\nu](u)\big), \quad u \in \mathbb{R}\,.$$

In a first step we will show that the drift can be discarded, which was also the case for the upper bound by Nickl and Reiß (2012). Since Lemma 4.6 shows that

the inverse adjoint score operator is given by $\mathcal{F}^{-1}[1/\varphi_\nu(-\bullet)]$ on the smooth subset $\mathcal{G}$, we study the mapping properties of this deconvolution operator in Step 2. Finally in Step 3, we apply the characterization in Proposition 2.11 to prove that $Z^\beta(\mathbb{R}) \subseteq \operatorname{ran} A_\nu^\star$ and to determine $(A_\nu^\star)^{-1}$ on $Z^\beta(\mathbb{R})$.

Step 1: Let us show that $\gamma = 0$ can be assumed, meaning that the process L has no drift. For any $\gamma \in \mathbb{R}$ consider the infinitely divisible distribution $\widetilde{P}_\nu := P_\nu * \delta_{-\gamma}$. Then the following map is an isomorphism

$$\Phi: \quad L_0^2(P_\nu) \to L_0^2(\widetilde{P}_\nu), \quad g \mapsto g(\bullet + \gamma).$$

Lemma 4.6 determines the adjoint score operator $\widetilde{A}_\nu^\star$ which corresponds to $\widetilde{P}_\nu$. Also by Lemma 4.6 we see for $g \in \mathcal{G}$ that $A_\nu^\star g = P_\nu(-\bullet) * g = \widetilde{P}_\nu(-\bullet) * g(\bullet + \gamma)$. Therefore, $A_\nu^\star = \widetilde{A}_\nu^\star \circ \Phi$ which implies

$$\operatorname{ran} A_\nu^\star = \operatorname{ran}\left(\widetilde{A}_\nu^\star \circ \Phi\right) = \operatorname{ran} \widetilde{A}_\nu^\star.$$

Hence, for the rest of the proof suppose $\gamma = 0$.

Step 2: The aim of this step is to show for $\zeta = \zeta^s + \zeta^c$ and any $\varepsilon > 0$

$$\| \mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)\, \mathcal{F}\, \zeta]\|_{L^2(P_\nu)} \lesssim \left\|\zeta^s(x)\right\|_{H^\beta} + \left\|\tfrac{1}{x}\zeta^s(x)\right\|_{H^\beta} + \|\zeta^c\|_{C^{\beta+\varepsilon}}. \tag{6.5}$$

To this end note that Assumption B yields, due to $\gamma = 0$,

$$|\varphi_\nu^{-1}(u)| \lesssim (1+|u|)^\beta \quad \text{and} \quad |(\varphi_\nu^{-1})'(u)| \lesssim (1+|u|)^{\beta-1}$$

and thus Lemma 4(c) by Nickl and Reiß (2012) or alternatively Lemma 5(i) by Söhl and Trabs (2012) shows that for all $s \in \mathbb{R}, p, q \in [1,\infty]$ the linear map

$$B_{p,q}^{s+\beta}(\mathbb{R}) \to B_{p,q}^s(\mathbb{R}), \quad f \mapsto \mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)\, \mathcal{F}\, f] \tag{6.6}$$

is bounded. This yields for any $\varepsilon > 0$ and $\zeta^c \in C^{\beta+\varepsilon}(\mathbb{R})$

$$\| \mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)\, \mathcal{F}\, \zeta^c]\|_{L^2(P_\nu)} \lesssim \|F^{-1}[\varphi_\nu^{-1}(-\bullet)\, \mathcal{F}\, \zeta^c]\|_\infty \lesssim \|\zeta^c\|_{B^\beta_{\infty,1}} \lesssim \|\zeta^c\|_{C^{\beta+\varepsilon}}. \tag{6.7}$$

For the singular part we apply a similar decomposition as Nickl and Reiß (2012). Integration by parts yields

$$\begin{aligned}\mathcal{F}^{-1}\left[\frac{\mathcal{F}\,\zeta^s}{\varphi_\nu(-\bullet)}\right] &= \mathcal{F}^{-1}\left[\frac{(\mathcal{F}[\frac{1}{ix}\zeta^s(x)])'}{\varphi_\nu(-\bullet)}\right] \\ &= ix\,\mathcal{F}^{-1}\left[\frac{\mathcal{F}[\frac{1}{ix}\zeta^s(x)]}{\varphi_\nu(-\bullet)}\right] + \mathcal{F}^{-1}\left[\mathcal{F}\left[\tfrac{1}{ix}\zeta^s(x)\right](\varphi_\nu^{-1})'(-\bullet)\right]\end{aligned} \tag{6.8}$$

Note that $1/\varphi_\nu$ is a Fourier multiplier from H^β into $H^0 = L^2(\mathbb{R})$ on the assumption $|\varphi_\nu(u)| \gtrsim (1+|u|)^{-\beta}$. Similarly, $(\varphi_\nu^{-1})'$ is a Fourier multiplier from H^β into H^1. Hence, for $\frac{1}{i\bullet}\zeta^s \in H^\beta$

$$\begin{aligned}\mathcal{F}^{-1}\left[\frac{\mathcal{F}[\frac{1}{ix}\zeta^s(x)]}{\varphi_\nu(-\bullet)}\right] &\in L^2(\mathbb{R}), \\ \mathcal{F}^{-1}\left[\mathcal{F}\left[\tfrac{1}{ix}\zeta^s(x)\right](\varphi_\nu^{-1})'(-\bullet)\right] &\in H^1(\mathbb{R}) \subseteq C^0(\mathbb{R}),\end{aligned}$$

where the last inclusion holds by the Sobolev embedding. Moreover,

$$ix\,\mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\frac{1}{ix}\zeta^s(x)]}{\varphi_\nu(-\bullet)}\Big] = \mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\zeta^s(x)]}{\varphi_\nu(-\bullet)}\Big] - \mathcal{F}^{-1}\Big[\mathcal{F}\big[\tfrac{1}{ix}\zeta^s(x)\big](\varphi_\nu^{-1})'(-\bullet)\Big]$$

which is an $L^2(\mathbb{R})$-function. Applying Lemma 4(a) by Nickl and Reiß (2012), the distribution $xP_\nu(\,\mathrm{d}x)$ has a bounded Lebesgue density and which yields together with $|x|^2 \lesssim |x||1+ix|^2$ and the continuous embeddings above

$$\begin{aligned}
\|\,\mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\,\zeta^s]\|_{L^2(P_\nu)} \leqslant& \int_{\mathbb{R}} \Big|ix\,\mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\frac{1}{ix}\zeta^s(x)]}{\varphi_\nu(-\bullet)}\Big](x)\Big|^2\,\mathrm{d}P_\nu(x)\\
&+\int_{\mathbb{R}} \Big|\,\mathcal{F}^{-1}\Big[\mathcal{F}\big[\tfrac{1}{ix}\zeta^s(x)\big](\varphi_\nu^{-1})'(-\bullet)\Big](x)\Big|^2\,\mathrm{d}P_\nu(x)\\
\lesssim& \Big\|(1+ix)\,\mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\frac{1}{ix}\zeta^s(x)]}{\varphi_\nu(-\bullet)}\Big](x)\Big\|_{L^2}\\
&+\Big\|\,\mathcal{F}^{-1}\Big[\mathcal{F}\big[\tfrac{1}{ix}\zeta^s(x)\big](\varphi_\nu^{-1})'(-\bullet)\Big](x)\Big\|_\infty\\
\lesssim& \big\|\tfrac{1}{x}\zeta^s(x)\big\|_{H^\beta} + \big\|\zeta^s\big\|_{H^\beta}.
\end{aligned}$$

Combining with (6.7), we get (6.5).

Step 3: Define the sets

$$\mathcal{G}' := C^\infty(\mathbb{R}) \cap L^2_0(P_\nu) \quad \text{and} \quad \mathcal{H}' := \big\{b \in C^\infty(\mathbb{R}) | b(0) = 0\big\} \cap \dot{\Theta}_\nu$$

which are larger than $\mathcal{G}$ an $\mathcal{H}$ from above. Using the Fourier multiplier property on Besov spaces (6.6) and an analogous result for the Fourier multiplier $\varphi_\nu(-\bullet)$, we obtain

$$\|P_\nu(-\bullet) * f\|_{C^{s'}} \lesssim \|f\|_{C^{s'}} \quad \text{and} \quad \|F^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\,f]\|_{C^s} \lesssim \|f\|_{C^{s'}}$$

for any $s > 0$ and $f \in C^{s'}$ for $s' > s + \beta$. Therefore, following the lines of the proof of Lemma 4.6(i), we see that $A^\star_\nu|_{\mathcal{G}'}$ is given by $A^\star_\nu g = P_\nu(-\bullet) * g$ for $g \in \mathcal{G}'$ and that $A^\star_\nu|_{\mathcal{G}'}$ is a bijection from $\mathcal{G}'$ onto $\mathcal{H}'$ with inverse $(A^\star_\nu|_{\mathcal{G}'})^{-1}b = F^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\,b]$ for $b \in \mathcal{H}'$.

By Proposition 2.11 for any ζ a necessary and sufficient condition to be in the range of $A^\star_\nu$ is the existence of a sequence $(\chi_m)_{m\in\mathbb{N}} \subseteq \mathcal{H}'$ such that $\chi_m \to \zeta$ in $L^2(\nu)$ and $(A^\star_\nu)^{-1}\chi_m$ converges in $L^2(P_\nu)$. Now, for any $\zeta \in Z^\beta \cap L^1(\nu) \cap L^2(\nu)$ we find $\chi_m = \chi^s_m + \chi^c_m$ with $\chi^s_n, \chi^c_n \in \mathcal{H}'$ satisfying $\chi^s_m \to \zeta^s$ and $\chi^c_m \to \zeta^c$ in $L^2(\nu)$ as well as

$$\begin{aligned}
\Big\|\,\mathcal{F}^{-1}\Big[\frac{\mathcal{F}[\zeta-\chi_n]}{\varphi_\nu(-\bullet)}\Big]\Big\|_{L^2(P_\nu)} \lesssim& \big\|(\zeta^s - \chi^s_n)(x)\big\|_{H^\beta}\\
&+\big\|\tfrac{1}{x}(\zeta^s-\chi^s_n)(x)\big\|_{H^\beta} + \|\zeta^c - \chi^c_n\|_{C^{\beta+\varepsilon}} \to 0
\end{aligned}$$

for $m \to \infty$ owing to (6.5). Hence, $\zeta \in \operatorname{ran} A^\star_\nu$ with $(A^\star_\nu)^{-1}\zeta = \mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)\,\mathcal{F}\,\zeta] = \mathcal{F}^{-1}[\varphi_\nu^{-1}(-\bullet)] * \zeta$. □

References

Belomestny, D. and M. Reiß (2006). Spectral calibration of exponential Lévy models. *Finance Stoch. 10*(4), 449–474.

Bickel, P. J., C. A. Klaassen, Y. Ritov, and J. A. Wellner (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York.

Bissantz, N., T. Hohage, and A. Munk (2004). Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise. *Inverse Problems 20*(6), 1773–1789.

Brown, L. D. and M. G. Low (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist. 24*(6), 2384–2398.

Buchmann, B. and R. Grübel (2003). Decompounding: an estimation problem for Poisson random sums. *Ann. Statist. 31*(4), 1054–1074.

Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems 24*(3), 034004, 19.

Cavalier, L., G. K. Golubev, D. Picard, and A. B. Tsybakov (2002). Oracle inequalities for inverse problems. *Ann. Statist. 30*(3), 843–874. Dedicated to the memory of Lucien Le Cam.

Clément, E., S. Delattre, and A. Gloter (2013). An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility. *Stochastic Process. Appl. 123*(7), 2500–2521.

Da Prato, G. and J. Zabczyk (1992). *Stochastic equations in infinite dimensions*, Volume 44 of *Encyclopedia of Mathematics and its Applications.* Cambridge: Cambridge University Press.

Engl, H. W., M. Hanke, and A. Neubauer (1996). *Regularization of inverse problems*, Volume 375 of *Mathematics and its Applications.* Dordrecht: Kluwer Academic Publishers Group.

Fan, J. (1991). On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *The Annals of Statistics 19*(3), 1257–1272.

Giné, E. and R. Nickl (2008). Uniform central limit theorems for kernel density estimators. *Probab. Theory Related Fields 141*(3-4), 333–387.

Goldenshluger, A. and S. V. Pereverzev (2000). Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory Related Fields 118*(2), 169–186.

Goldenshluger, A. and S. V. Pereverzev (2003). On adaptive inverse estimation of linear functionals in Hilbert scales. *Bernoulli 9*(5), 783–807.

Jacod, J. (1989). Une application de la topologie d'Emery: le processus information d'un modèle statistique filtré. In *Séminaire de Probabilités, XXIII*, Volume 1372 of *Lecture Notes in Math.*, pp. 448–474. Berlin: Springer.

Jacod, J. (1990). Regularity, partial regularity, partial information process, for a filtered statistical model. *Probab. Theory Related Fields 86*(3), 305–335.

Jacod, J. and A. N. Shiryaev (2003). *Limit theorems for stochastic processes* (Second ed.), Volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Berlin: Springer-Verlag.

Khoujmane, A., F. Ruymgaart, and M. Shubov (2007). The convolution theorem for estimating linear functionals in indirect nonparametric regression models. *J. Statist. Plann. Inference 137*(3), 811–820.

Klaassen, C. A. J., E.-J. Lee, and F. H. Ruymgaart (2001). On efficiency of indirect estimation of nonparametric regression functions. In *Algebraic methods in statistics and probability (Notre Dame, IN, 2000)*, Volume 287 of *Contemp. Math.*, pp. 173–184. Providence, RI: Amer. Math. Soc.

Klaassen, C. A. J., E.-J. Lee, and F. H. Ruymgaart (2005). Asymptotically efficient estimation of linear functionals in inverse regression models. *J. Nonparametr. Stat. 17*(7), 819–831.

Kuo, H. H. (1975). *Gaussian measures in Banach spaces.* Lecture Notes in Mathematics, Vol. 463. Berlin: Springer-Verlag.

Le Cam, L. (1972). Limits of experiments. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. I: Theory of statistics*, Berkeley, Calif., pp. 245–261. Univ. California Press.

Liese, F. (1987). Estimates of Hellinger integrals of infinitely divisible distributions. *Kybernetika 23*(3), 227–238.

Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics.* Lecture notes in statistics. Springer, Berlin Heidelberg.

Natterer, F. (1984). Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Anal. 18*(1-2), 29–37.

Neumann, M. H. and M. Reiß (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli 15*(1), 223–248.

Nickl, R. and M. Reiß (2012). A Donsker theorem for Lévy measures. *J. Funct. Anal. 263*(10), 3306–3332.

Nussbaum, M. (1996). Asymptotic equivalence of density estimation and gaussian white noise. *Ann. Statist. 24*(6), 2399–2430.

Reiß, M. (2013). Testing the characteristics of a Lévy process. *Stochastic Process. Appl. 123*(7), 2808–2828. Special Issue International Year of Statistics.

Sato, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press.

Söhl, J. and M. Trabs (2012). A uniform central limit theorem and efficiency for deconvolution estimators. *Electron. J. Statist. 6*, 2486–2518.

Trabs, M. (2013). Calibration of self–decomposable Lévy models. *Bernoulli*, to appear.

van der Vaart, A. W. (1988). *Statistical estimation in large parameter spaces*, Volume 44 of *CWI Tract.* Amsterdam: Stichting Mathematisch Centrum Centrum voor Wiskunde en Informatica.

van der Vaart, A. W. (1991). On differentiable functionals. *Ann. Statist. 19*(1), 178–204.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes.* Springer Series in Statistics. Springer, New York.

van Rooij, A. C. M., F. H. Ruymgaart, and W. R. van Zwet (1999). Asymptotic efficiency of inverse estimators. *Teor. Veroyatnost. i Primenen. 44*(4), 826–844. translation in *Theory Probab. Appl.* 44(4), 722–738.

Witting, H. (1985). *Mathematische Statistik. I.* Stuttgart: B. G. Teubner. Parametrische Verfahren bei festem Stichprobenumfang. [Parametric methods for fixed sample size].