

# Low-rank diffusion matrix estimation for high-dimensional time-changed Lévy processes

Denis Belomestny<sup>a,b</sup> and Mathias Trabs<sup>c</sup>

<sup>a</sup>*Faculty of Mathematics, Duisburg-Essen University, Thea-Leymann-Str. 9, D-45127 Essen, Germany. E-mail: [denis.belomestny@uni-due.de](mailto:denis.belomestny@uni-due.de)*

<sup>b</sup>*National Research University Higher School of Economics, Shabolovka, 26, 119049 Moscow, Russia*

<sup>c</sup>*Faculty of Mathematics, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany. E-mail: [mathias.trabs@uni-hamburg.de](mailto:mathias.trabs@uni-hamburg.de)*

Received 15 June 2016; revised 3 April 2017; accepted 13 June 2017

---

**Abstract.** The estimation of the diffusion matrix  $\Sigma$  of a high-dimensional, possibly time-changed Lévy process is studied, based on discrete observations of the process with a fixed distance. A low-rank condition is imposed on  $\Sigma$ . Applying a spectral approach, we construct a weighted least-squares estimator with nuclear-norm-penalisation. We prove oracle inequalities and derive convergence rates for the diffusion matrix estimator. The convergence rates show a surprising dependency on the rank of  $\Sigma$  and are optimal in the minimax sense for fixed dimensions. Theoretical results are illustrated by a simulation study.

**Résumé.** Nous étudions le problème de l'estimation de la matrice de diffusion  $\Sigma$  d'un processus de Lévy en grande dimension, qui peut être changé de temps, en se basant sur des observations discrètes du processus à une distance fixée. Nous imposons une condition de faible rang sur  $\Sigma$ . À l'aide d'une méthode spectrale, nous construisons un estimateur pondéré des moindres carrés avec une pénalisation par une norme nucléaire. Nous prouvons des inégalités oracle et obtenons des vitesses de convergence pour l'estimateur de la matrice de diffusion. Nous constatons que ces vitesses dépendent du rang de  $\Sigma$  d'une façon surprenante, et qu'elles sont optimales au sens minimax pour une dimension fixée. Ces résultats théoriques sont illustrés par une étude de simulations.

*MSC:* Primary 62M05; secondary 60G51; 62G05; 62M15

*Keywords:* Volatility estimation; Lasso-type estimator; Minimax convergence rates; Nonlinear inverse problem; Oracle inequalities; Time-changed Lévy process

---

## 1. Introduction

Non-parametric statistical inference for Lévy-type processes has been attracting the attention of researchers for many years initiated by the works of Rubin and Tucker [35] as well as Basawa and Brockwell [5]. The popularity of Lévy processes is related to their simplicity on the one hand and the ability to reproduce many stylised patterns presented in economic and financial data or to model stochastic phenomena in biology or physics, on the other hand. While non-parametric inference for one dimensional Lévy processes is nowadays well understood (see e.g. the lecture notes [7]), there are surprisingly few results for multidimensional Lévy or related jump processes. Possible applications demand however for multi- or even high-dimensional methods. As a first contribution to statistics for jump processes in high dimensions, we investigate the optimal estimation of the diffusion matrix of a  $d$ -dimensional Lévy process where  $d$  may grow with the number of observations.

More general, we consider a larger class of time-changed Lévy process. Let  $X$  be a  $d$ -dimensional Lévy process and let  $\mathcal{T}$  be a non-negative, non-decreasing stochastic process with  $\mathcal{T}(0) = 0$ . Then the time-changed Lévy process is defined via  $Y_s = X_{\mathcal{T}(s)}$  for  $s \geq 0$ . For instance, in the finance context, the change of time is motivated by the fact

that some economical effects (as nervousness of the market which is indicated by volatility) can be better understood in terms of a “business time” which may run faster than the physical one in some periods (see, e.g. [41]). In view of large portfolios and a huge number of assets traded at the stock markets we have high-dimensional observations.

Monroe [28] has shown that even in the case of the Brownian motion  $X$ , the resulting class of time-changed Lévy processes is rather large and basically coincides, at least in the one-dimensional case, with the class of all semi-martingales. Nevertheless, a practical application of this fact for statistical modelling meets two major problems: first, the change of time  $\mathcal{T}$  can be highly intricate – for instance, if  $Y$  has discontinuous trajectories, cf. [4]; second, the dependence structure between  $X$  and  $\mathcal{T}$  can be also quite involved. In order to avoid these difficulties and to achieve identifiability, we allow  $X$  to be a general Lévy processes, but assume independence of  $X$  and  $\mathcal{T}$ .

A first natural question is which parameters of the underlying Lévy process  $X$  can be identified from discrete observations  $Y_0, Y_\Delta, \dots, Y_{n\Delta}$  for some  $\Delta > 0$  as  $n \rightarrow \infty$ . This question has been recently addressed in the literature (see [6] and references therein) and the answer turns out to crucially depend on the asymptotic behaviour of  $\Delta$  and on the degree of our knowledge about  $\mathcal{T}$ . We study the so-called low-frequency regime, meaning that the observation distance  $\Delta$  is fixed. In this case, we have severe identifiability problems which can be easily seen from the fact that subordinated Lévy processes are again Lévy processes. Assuming the time-change is known, all parameters of the underlying Lévy process  $X$  can be identified as  $n \rightarrow \infty$ . Since any Lévy process can be uniquely parametrised by the so-called Lévy triplet  $(\Sigma, \gamma, \nu)$  with a positive semi-definite diffusion (or volatility) matrix  $\Sigma \in \mathbb{R}^{d \times d}$ , a drift vector  $\gamma \in \mathbb{R}^d$  and a jump measure  $\nu$ , we face a semi-parametric estimation problem. Describing the covariance of the diffusion part of the Lévy process  $X$ , the matrix  $\Sigma$  is of particular interest. Aiming for a high dimensional setting, we study the estimation of  $\Sigma$  under the assumption that it has low rank. This low rank condition reflects the idea of a few principal components driving the whole process. We also discuss the case where the diffusion matrix can be decomposed into a low rank matrix and a sparse matrix as has been suggested by Fan et al. [17,18].

For discretely observed Lévy processes it has been recently shown that estimation methods coming from low-frequency observations attain also in the high-frequency case, where  $\Delta \rightarrow 0$ , optimal convergence results, cf. [21] for volatility estimation and [31] for the estimation of the jump measure. In contrast to estimators which are tailor-made for high-frequency observations, these low-frequency methods thus turn out to be robust with respect to the sampling frequency. This observation is a key motivation for considering a fixed  $\Delta$  in the present article. It is very common in the literature on statistics for stochastic processes, that the observations are supposed to be equidistant. In view of the empirical results by [2], this model assumption might however be often violated in financial applications. Our observation scheme can be equivalently interpreted as random observation times  $(T(\Delta j))_j$  of the Lévy process  $X$ . For one-dimensional Lévy processes general observation distances as been studied by Kappus [23].

The research area on statistical inference for discretely observed stochastic processes is rapidly growing. Let us only mention some closely related contributions while we refer to [7] for a recent overview of estimation for Lévy processes. Non-parametric estimation for time-changed Lévy models has been studied in [6,8,13]. In a two-dimensional setting the jump measure of a Lévy process has been estimated in [12]. For inference on the volatility matrix (in small dimensions) for continuous semi-martingales in a high-frequency regime we refer to [10,20] and references therein. Large sparse volatility matrix estimation for continuous Itô processes has been recently studied in [37,38,42].

Our statistical problem turns out to be closely related to a deconvolution problem for a multidimensional normal distribution with zero mean and covariance matrix  $\Sigma$  convolved with a nuisance distribution which is unknown except for some of its structural properties. Due to the time-change or random sampling, respectively, the normal vector is additionally multiplied with a non-negative random variable. Hence, we face a covariance estimation problem in a generalised mixture model. Since we have no direct access to a sample of the underlying normal distribution, our situation differs considerably from the inference problems which have been studied in the literature on high-dimensional covariance matrix estimation so far.

The analysis of many deconvolution and mixture models becomes more transparent in spectral domain. Our accordingly reformulated problem bears some similarity to the so-called trace regression problem and the related matrix estimation, which recently got a lot of attention in statistical literature (see, e.g. [1,24,29,34]). We face a non-linear analogue to the estimation of low-rank matrices based on noisy observations. Adapting some ideas, especially from [24], we construct a weighted least squares estimator with nuclear norm penalisation in the spectral domain.

We prove oracle inequalities for the estimator of  $\Sigma$  implying that the estimator adapts to the low-rank condition on  $\Sigma$  and that it also applies in the miss-specified case of only approximated sparsity. The resulting convergence rate fundamentally depends on the time-change while the dimension may grow polynomially in the sample size. Lower

bounds verify that the rates are optimal in the minimax sense for fixed dimensions. The influence of the rank of  $\Sigma$  on the convergence rate reveals a new phenomenon which was not observed in multi-dimensional non-parametric statistics before. Namely, in a certain regime the convergence rate in  $n$  is faster for a larger rank of  $\Sigma$ . To understand this surprising phase transition from a more abstract perspective, we briefly discuss a related regression model.

The paper is organised as follows. In Section 2, we introduce notations and formulate the statistical model. In Section 3 the estimator for the diffusion matrix is constructed and the oracle inequalities for this estimator are derived. Based on these oracle inequalities, we derive in Section 4 upper and lower bounds for the convergence rate. Section 5 is devoted to several extensions, including mixing time-changes, incorporating a sparse component of the diffusion matrix and the related trace-regression model. Some numerical examples can be found in Section 6. In Sections 7 and 8 the proofs of the oracle inequalities and of the minimax convergence rates, respectively, are collected. The Appendix contains some (uniform) concentration results for multivariate characteristic functions which are of independent interest.

## 2. Main setup

Recall that a random time change  $\{\mathcal{T}(t) : t \geq 0\}$  is an increasing right-continuous process with left limits such that  $\mathcal{T}(0) = 0$ . For each  $t$  the random variable  $\mathcal{T}(t)$  is a stopping time with respect to the underlying filtration. For a Lévy process  $X = \{X_t : t \geq 0\}$  the time-changed Lévy process is given by  $X_{\mathcal{T}(t)}, t \geq 0$ . We throughout assume that  $\mathcal{T}$  is independent of  $X$ . By reparametrising the time change we can assume without loss of generality that  $\Delta = 1$  and that the observations are thus given by the increments

$$Y_j := X_{\mathcal{T}(j)} - X_{\mathcal{T}(j-1)}, \quad j = 1, \dots, n,$$

for  $n \in \mathbb{N}$ . Note that in general the observations are not independent, in contrast to the special case of low-frequently observed Lévy processes. The estimation procedure relies on the following crucial insight: If the sequence

$$T_j := \mathcal{T}(j) - \mathcal{T}(j - 1), \quad j = 1, \dots, n,$$

is stationary and admits an invariant measure  $\pi$ , then the independence of  $\mathcal{T}$  and  $X$  together with the Lévy–Khintchine formula yield that the observations  $Y_j$  have a common characteristic function given by

$$\varphi(u) := \mathbb{E}[e^{i\langle Y_j, u \rangle}] = \int_0^\infty \mathbb{E}[e^{i\langle X_t, u \rangle}] \pi(dt) = \int_0^\infty e^{t\psi(u)} \pi(dt) = \mathcal{L}(-\psi(u)), \quad u \in \mathbb{R}^d,$$

where  $\mathcal{L}$  is the Laplace transform of  $\pi$  and where the characteristic exponent is given by

$$\begin{aligned} \psi(u) &= -\frac{1}{2}\langle u, \Sigma u \rangle + i\langle \gamma, u \rangle + \int_{\mathbb{R}^d} (e^{i\langle x, u \rangle} - 1 - i\langle x, u \rangle \mathbb{1}_{\{|x| \leq 1\}}(x)) \, d\nu(x) \\ &=: -\frac{1}{2}\langle u, \Sigma u \rangle + \Psi(u), \quad u \in \mathbb{R}^d, \end{aligned} \tag{2.1}$$

with the diffusion matrix  $\Sigma \in \mathbb{R}^{d \times d}$ , for a drift parameter  $\gamma \in \mathbb{R}^d$  and a jump measure  $\nu$ . If  $\gamma = 0$  and  $\nu = 0$ , we end up with the problem of estimating a covariance matrix. The function  $\Psi$  in (2.1) appears due to the presence of jumps and can be viewed as a nuisance parameter. Let us give some examples of typical time-changes and their Laplace transforms.

### Examples 2.1.

- (i) *Low-frequency observations* of  $X_t$  with observation distance  $\Delta > 0$ :

$$T_j \sim \pi = \delta_\Delta, \quad \mathcal{L}(z) = e^{-\Delta z}.$$

(ii) *Poisson process* time-change or exponential waiting times with intensity parameter  $\Delta > 0$ :

$$T_j \sim \pi = \text{Exp}(\Delta), \quad \mathcal{L}(z) = \frac{1}{1 + \Delta z}.$$

(iii) *Gamma process* time-change with parameters  $\Delta, \vartheta > 0$ :

$$T_j \sim \pi = \Gamma(\vartheta, \Delta), \quad \mathcal{L}(z) = (1 + \Delta z)^{-\vartheta}.$$

(iv) *Integrated CIR-process* for some  $\eta, \kappa, \xi > 0$  such that  $2\kappa\eta > \xi^2$  (which has  $\alpha$ -mixing increments, cf. [26]):

$$\begin{aligned} \mathcal{T}(t) &= \int_0^t Z_t \, dt \quad \text{with } dZ_t = \kappa(\eta - Z_t) \, dt + \xi \sqrt{Z_t} \, dW_t, \\ \mathcal{L}(z) &\sim \exp\left(-\frac{\sqrt{2z}}{\xi}(1 + \kappa\eta)\right) \quad \text{as } |z| \rightarrow \infty \text{ with } \text{Re } z \geq 0. \end{aligned}$$

If the time change  $\mathcal{T}$  is unknown, one faces severe identifiability problems. Indeed, even in the case of a multivariate time-changed Brownian motion  $\Sigma^\top W_{\mathcal{T}(t)}$  we can identify the matrix  $\Sigma$  and  $\mathcal{T}(t)$  only up to a multiplicative factor. More generally, if  $X$  is not restricted to be Brownian motion there is much more ambiguity as the following example illustrates.

**Example 2.2.** Consider two time changes  $\mathcal{T}(t)$  and  $\tilde{\mathcal{T}}(t)$ , where  $\mathcal{T}(t)$  is a Gamma subordinator with the unit parameters and  $\tilde{\mathcal{T}}(t) = t$  is deterministic. We have  $\mathcal{L}_t(z) = 1/(1 + z)^t$  and  $\tilde{\mathcal{L}}_t(z) = e^{-tz}$ . Note that  $\mathbb{E}[\mathcal{T}(t)] = \mathbb{E}[\tilde{\mathcal{T}}(t)] = t$ . Let  $(X_t)$  be a one-dimensional Lévy process with the characteristic exponent  $\psi(u)$  and let  $(\tilde{X}_t)$  be another Lévy process with the characteristic exponent  $\tilde{\psi}(u) = \log(1/(1 - \psi(u)))$  (the existence of such Lévy process follows from the well-known fact that  $\exp(\tilde{\psi}(u))$  is the characteristic function of some infinitely divisible distribution). Then we obviously have  $\mathcal{L}_t(-\psi(u)) \equiv \tilde{\mathcal{L}}_t(-\tilde{\psi}(u))$  for all  $t \geq 0$ . Hence, the corresponding time-changed Lévy processes  $Y_t = X_{\mathcal{T}(t)}$  and  $\tilde{Y}_t = \tilde{X}_{\tilde{\mathcal{T}}(t)}$  satisfy

$$Y_{k\Delta} - Y_{(k-1)\Delta} \stackrel{d}{=} \tilde{Y}_{k\Delta} - \tilde{Y}_{(k-1)\Delta}, \quad k \in \mathbb{N},$$

for any  $\Delta > 0$ . Moreover, the increments  $Y_{k\Delta} - Y_{(k-1)\Delta}$  are independent in this case.

The above example shows that even under the additional assumption  $\mathbb{E}[\mathcal{T}(t)] = t$ , we cannot, in general, consistently estimate the parameters of the one-dimensional time-changed Lévy process  $(Y_t)$  from the low-frequency observations. As shown by Belomestny [6] all parameters of a  $d$ -dimensional (with  $d > 1$ ) Lévy process  $X$  can be identified under additional assumption that all components of  $X$  are independent and  $\mathbb{E}[\mathcal{T}(t)] = t$ . However this assumption would imply that the matrix  $\Sigma$  is diagonal and is thus much too restrictive for our setting. Therefore, we throughout assume that the time-change and consequently its Laplace transform are known, see Section 5.1 for a discussion for unknown  $\mathcal{L}$ . Let us note that similar identifications problems appear even in high-frequency setting, see [19].

Before we construct the volatility matrix estimator in the next section, we should introduce some notation. Recall the Frobenius or trace inner product  $\langle A, B \rangle := \text{tr}(A^\top B)$  for matrices  $A, B \in \mathbb{R}^{d \times d}$ . For  $p \in (0, \infty]$ , the Schatten- $p$  norm of  $A$  is given by

$$\|A\|_p := \left( \sum_{j=1}^d \sigma_j(A)^p \right)^{1/p}$$

with  $\sigma_1(A) \geq \dots \geq \sigma_d(A)$  being the singular values of  $A$ . In particular,  $\|A\|_1$ ,  $\|A\|_2$  and  $\|A\|_\infty$  denote the nuclear, the Frobenius and the spectral norm of  $A$ , respectively. We will frequently apply the trace duality property

$$|\text{tr}(AB)| \leq \|A\|_1 \|B\|_\infty, \quad A, B \in \mathbb{R}^{d \times d}.$$

Moreover, we may consider the entry-wise norms  $|A|_p := (\sum_{i,j} |a_{i,j}|^p)^{1/p}$  for  $0 < p \leq \infty$  with the usual modification  $|A|_0$  denoting the number of non-zero entries of  $A = (a_{i,j})_{i,j}$ .

For any matrices  $A \in \mathbb{R}^{k \times k}$ ,  $B \in \mathbb{R}^{l \times l}$ ,  $k, l \in \mathbb{N}$ , we write

$$\text{diag}(A, B) := \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \in \mathbb{R}^{(k+l) \times (k+l)}.$$

For  $a, b \in \mathbb{R}$  we write  $a \lesssim b$  if there is a constant  $C$  independent of  $n, d$  and the involved parameters such that  $a \leq Cb$ .

### 3. The estimator and oracle inequalities

In view of the Lévy–Khintchine formula, we apply a spectral approach. The natural estimator of  $\varphi$  is given by the empirical characteristic function

$$\varphi_n(u) := \frac{1}{n} \sum_{j=1}^n e^{i\langle u, Y_j \rangle}, \quad u \in \mathbb{R}^d,$$

which is consistent whenever  $(Y_j)_{j \geq 0}$  is ergodic. Even more, it concentrates around the true characteristic function with parametric rate uniformly on compact sets, cf. Theorems A.2 and A.4 in the Appendix. A plug-in approach yields the estimator for the characteristic exponent given by

$$\widehat{\psi}_n(u) := -\mathcal{L}^{-1}(\varphi_n(u)), \quad (3.1)$$

where  $\mathcal{L}^{-1}$  denotes a continuously chosen inverse of the map  $\{\text{Re } z > 0\} \ni z \mapsto \mathcal{L}(v) \in \mathbb{C} \setminus \{0\}$ . Since  $\Sigma$  only appears in the real part of the characteristic exponent, we may use  $\text{Re}(\widehat{\psi}_n)$ .

**Remark 3.1.** For high-frequency observations, that is  $\Delta = \Delta_n \rightarrow 0$ , the characteristic function of the observations  $Y_j$  can be written as  $\varphi_\Delta(u) = \mathcal{L}(-\Delta\psi(u))$  where  $\psi$  is the characteristic exponent as before and where  $\mathcal{L}$  is the Laplace transform of the rescaled time-change increments  $\frac{\mathcal{T}(\Delta j) - \mathcal{T}(\Delta(j-1))}{\Delta}$ , which may be assumed to be stationary. Under some regularity assumption on  $\mathcal{L}$ , we obtain the asymptotic expansion

$$\varphi_\Delta(u) = \mathcal{L}(-\Delta\psi(u)) = 1 - \mathcal{L}'(0)\Delta\psi(u) + \mathcal{O}(\Delta^2(|u|^4 \vee 1)) \quad \text{for } \Delta \rightarrow 0.$$

Instead of the inversion in (3.1) the estimator  $\overline{\psi}_n(u) = \frac{1 - \varphi_n(u)}{\Delta \mathcal{L}'(0)}$  can be used in this case, depending only on Laplace transform via  $\mathcal{L}'(0)$ . Note that the latter is given by the first moment of rescaled time-change increments. However, this asymptotic identification can only be applied in the high-frequency regime  $\Delta \rightarrow 0$ , while the estimator (3.1) is robust with respect to the sampling frequency.

Applying  $u^\top \Sigma u = \langle uu^\top, \Sigma \rangle$ , estimating the low-rank matrix  $\Sigma$  can thus be reformulated as the regression problem

$$\frac{\widehat{\psi}_n(u)}{|u|^2} = -\frac{1}{2} \langle \Theta(u), \Sigma \rangle + \frac{\Psi(u)}{|u|^2} + \frac{\widehat{\psi}_n(u) - \psi(u)}{|u|^2} \quad \text{with } \Theta(u) := \frac{uu^\top}{|u|^2} \in \mathbb{R}^{d \times d}, \quad (3.2)$$

where we have normalised the whole formula by the factor  $|u|^{-2}$ . The design matrix  $vv^\top$  for an arbitrary unit vector  $v \in \mathbb{R}^d$  is deterministic and degenerated. The second term in (3.2) is a deterministic error which will be small only for large  $u$ . The last term reflects the stochastic error. Due to the identification via the Laplace transform  $\mathcal{L}$ , the estimation problem is non-linear and turns out to be ill-posed: the stochastic error grows for large frequencies.

Inspired by the regression formula (3.2), we define the diffusion matrix estimator as a penalised least squares type estimator with regularisation parameter  $\lambda > 0$  and a spectral cut-off  $U > 1$ :

$$\widehat{\Sigma}_{n,\lambda} := \arg \min_{M \in \mathbb{M}} \left\{ \int_{\mathbb{R}^d} (2|u|^{-2} \text{Re} \widehat{\psi}_n(u) + \langle \Theta(u), M \rangle)^2 w_U(u) du + \lambda \|M\|_1 \right\}, \quad (3.3)$$

where  $\mathbb{M}$  is a subset of positive semi-definite  $d \times d$  matrices and  $w_U$  is a weight function. We impose the following standing assumption on the weight function which is chosen by the practitioner.

**Assumption A.** Let  $w : \mathbb{R}^d \rightarrow \mathbb{R}_+$  be a radial non-negative function which is supported on the annulus  $\{1/4 < |u| \leq 1/2\}$ . For any  $U > 1$  let  $w_U(u) = U^{-d}w(u/U)$ ,  $u \in \mathbb{R}^d$ .

In the special case of a Lévy process with a finite jump activity  $\alpha := \nu(\mathbb{R}^d) \in (0, \infty)$ , we can write remainder from (2.1) as

$$\Psi(u) = i\langle \gamma_0, u \rangle + \mathcal{F}\nu(u) - \alpha, \quad u \in \mathbb{R}^d,$$

for  $\gamma_0 := \gamma - \int_{|u| \leq 1} x \nu(dx)$ . Since the Fourier transform  $\mathcal{F}\nu(u)$  converges to zero as  $|u| \rightarrow \infty$  under a mild regularity assumption on the finite measure  $\nu$ , we can reduce the bias of our estimator by the following modification

$$\begin{aligned} (\tilde{\Sigma}_{n,\lambda}, \tilde{\alpha}_{n,\lambda}) &:= \arg \min_{M \in \mathbb{M}, a \in I} \left\{ \int_{\mathbb{R}^d} \left( \frac{2}{|u|^2} \operatorname{Re}(\widehat{\psi}_n(u) + a) + \langle \Theta(u), M \rangle \right)^2 w_U(u) \, du + \lambda \left( \|M\|_1 + \frac{a}{U^2} \right) \right\} \\ &= \arg \min_{M \in \mathbb{M}, a \in I} \left\{ \int_{\mathbb{R}^d} \left( \frac{2}{|u|^2} \operatorname{Re} \widehat{\psi}_n(u) + \langle \tilde{\Theta}(u), \operatorname{diag} \left( M, \frac{a}{U^2} \right) \rangle \right)^2 w_U(u) \, du \right. \\ &\quad \left. + \lambda \left( \|M\|_1 + \frac{a}{U^2} \right) \right\} \end{aligned} \tag{3.4}$$

with

$$\tilde{\Theta}(u) := \tilde{\Theta}_U(u) = \operatorname{diag} \left( \Theta(u), \frac{2U^2}{|u|^2} \right), \quad u \in \mathbb{R}^d,$$

and some interval  $I \subseteq \mathbb{R}_+$ . The most interesting cases are  $I = \{0\}$ , where  $\widehat{\Sigma}_{n,\lambda}$  and  $\tilde{\Sigma}_{n,\lambda}$  coincide, and  $I = [0, \infty)$ . The factor  $U^{-2}$  in front of  $a$  is natural, since the ill-posedness of the estimation problem for the jump activity is two degrees larger than for estimating the volatility, cf. [9]. As a side product,  $\tilde{\alpha}_{n,\lambda}$  is an estimator for the jump intensity. By penalising large  $a$ , the estimator  $\tilde{\alpha}_{n,\lambda}$  is pushed back to zero if the least squares part cannot profit from a finite  $a$ . It thus coincides with the convention to set  $\alpha = 0$  in the infinite intensity case.

Our estimators  $\widehat{\Sigma}_{n,\lambda}$  and  $\tilde{\Sigma}_{n,\lambda}$  are related to the weighted scalar product

$$\begin{aligned} \langle (A, a), (B, b) \rangle_w &:= \int_{\mathbb{R}^d} \langle \tilde{\Theta}(u), \operatorname{diag}(A, a) \rangle \langle \tilde{\Theta}(u), \operatorname{diag}(B, b) \rangle w_U(u) \, du \\ &= \int_{\mathbb{R}^d} \left( \langle \Theta(u), A \rangle + \frac{2U^2}{|u|^2} a \right) \left( \langle \Theta(u), B \rangle + \frac{2U^2}{|u|^2} b \right) w_U(u) \, du \end{aligned}$$

with the usual notation  $\|(A, a)\|_w^2 := \langle (A, a), (A, a) \rangle_w$ . For convenience we abbreviate  $\|A\|_w := \|(A, 0)\|_w$ . The scalar product  $\langle \cdot, \cdot \rangle_w$  is the counterpart to the empirical scalar product in the matrix estimation literature. As the following lemma reveals, the weighted norm  $\|(A, a)\|_w$  and the Frobenius norm  $\|\operatorname{diag}(A, a)\|_2 = (\|A\|_2^2 + a^2)^{1/2}$  are equivalent. As the isometry is not restricted to any sparse sub-class of matrices, it especially implies the often imposed restricted isometry property. The proof is postponed to Section 7.1.

**Lemma 3.2.** *For any positive semi-definite matrix  $A \in \mathbb{R}^{d \times d}$  and any  $a \geq 0$ , it holds*

$$\underline{\kappa}_w \|\operatorname{diag}(A, a)\|_2 \leq \|(A, a)\|_w \leq \overline{\kappa}_w \|\operatorname{diag}(A, a)\|_2, \tag{3.5}$$

where  $\underline{\kappa}_w^2 := \int_{\mathbb{R}^d} v_1^4 / (|v|^4) w(v) \, dv$  and  $\overline{\kappa}_w^2 := 8 \int_{\mathbb{R}^d} |v|^{-4} w(v) \, dv$ .

Using well-known calculations from the Lasso literature, we obtain the following elementary oracle inequality, which is proven in Section 7.2. The condition  $\alpha \in I$  is trivially satisfied for  $I = \mathbb{R}_+$ .

**Proposition 3.3.** Let  $\mathbb{M} \subseteq \mathbb{R}^{d \times d}$  be an arbitrary subset of matrices and define

$$\mathcal{R}_n := 2 \int_{\mathbb{R}^d} \left( \frac{2}{|u|^2} \operatorname{Re} \widehat{\psi}_n(u) - \left\langle \widetilde{\Theta}(u), \operatorname{diag} \left( \Sigma, \frac{\alpha}{U^2} \right) \right\rangle \right) \widetilde{\Theta}(u) w_U(u) \, du \in \mathbb{R}^{(d+1) \times (d+1)}, \quad (3.6)$$

where we set  $\alpha = 0$  if  $\nu(\mathbb{R}^d) = \infty$ . Suppose  $\alpha \in I$ . On the event  $\{\|\mathcal{R}_n\|_\infty \leq \lambda\}$  for some  $\lambda > 0$  we have

$$\|(\widetilde{\Sigma}_{n,\lambda} - \Sigma, U^{-2}(\widetilde{\alpha}_{n,\lambda} - \alpha))\|_w^2 \leq \inf_{M \in \mathbb{M}} \{ \|M - \Sigma\|_w^2 + 2\lambda(\|M\|_1 + U^{-2}\alpha) \}. \quad (3.7)$$

If the set  $\mathbb{M}$  is convex, then sub-differential calculus can be used to refine the inequality (3.7). The proof is inspired by Koltchinskii et al. [24, Theorem 1] and postponed to Section 7.3. Note that this second oracle inequality improves (3.7) with respect to two aspects. Instead of  $\lambda$  we have  $\lambda^2$  in the second term on the right-hand side and the nuclear-norm of  $M$  is replaced by its rank.

**Theorem 3.4.** Suppose that  $\mathbb{M} \subseteq \mathbb{R}^{d \times d}$  is a convex subset of the positive semi-definite matrices and let  $\alpha \in I$ . On the event  $\{\|\mathcal{R}_n\|_\infty \leq \lambda\}$  for some  $\lambda > 0$  and for  $\mathcal{R}_n$  from (3.6) the estimators  $\widetilde{\Sigma}_{n,\lambda}$  and  $\widetilde{\alpha}_{n,\lambda}$  from (3.4) satisfy

$$\|(\widetilde{\Sigma}_{n,\lambda} - \Sigma, U^{-2}(\widetilde{\alpha}_{n,\lambda} - \alpha))\|_w^2 \leq \inf_{M \in \mathbb{M}} \left\{ \|M - \Sigma\|_w^2 + \left( \frac{1 + \sqrt{2}}{2\underline{z}_w} \right)^2 \lambda^2 (\operatorname{rank}(M) + \mathbb{1}_{\alpha \neq 0}) \right\},$$

where  $\underline{z}_w^2 := \int_{\mathbb{R}^d} v_1^4 / (2|v|^4) w(v) \, dv$ .

This oracle inequality is purely non-asymptotic and sharp in the sense that we have the constant one in front of  $\|M - \Sigma\|_w^2$  on the right-hand side. Combining Lemma 3.2 and Theorem 3.4 immediately yields an oracle inequality for the error in the Frobenius norm.

**Corollary 3.5.** Let  $\mathbb{M} \subseteq \mathbb{R}^{d \times d}$  be a convex subset of the positive semi-definite matrices and let  $\alpha \in I$ . On the event  $\{\|\mathcal{R}_n\|_\infty \leq \lambda\}$  for some  $\lambda > 0$  and for  $\mathcal{R}_n$  from (3.6) we have

$$\|\widetilde{\Sigma}_{n,\lambda} - \Sigma\|_2^2 + U^{-4} |\widetilde{\alpha}_{n,\lambda} - \alpha|^2 \leq C_w \inf_{M \in \mathbb{M}} \{ \|M - \Sigma\|_2^2 + \lambda^2 (\operatorname{rank}(M) + \mathbb{1}_{\alpha \neq 0}) \}$$

for a constant  $C_w$  depending only on  $\underline{z}_w$  and  $\overline{z}_w$ .

The remaining question is how the parameters  $U$  and  $\lambda$  should be chosen in order to control the event  $\{\|\mathcal{R}_n\|_\infty \leq \lambda\}$ . The answer will be given in the next section.

#### 4. Convergence rates

The above oracle inequalities hold true for any Lévy-process  $X$  and any independent time-change  $\mathcal{T}$  with stationary increments. However, to find some  $\lambda$  and  $U$  such that probability of the event  $\{\|\mathcal{R}_n\|_\infty \leq \lambda\}$  with the error term  $\mathcal{R}_n$  from (3.6) is indeed large, we naturally need to impose some assumptions. For simplicity we will concentrate on the well-specified case  $\Sigma \in \mathbb{M}$  noting that, thanks to Corollary 3.5, the results carry over to the miss-specified situation.

**Assumption B.** Let the jump measure  $\nu$  of  $X$  fulfil:

- (i) for some  $s \in (-2, \infty)$  and for some constant  $C_\nu > 0$  let

$$\alpha := 0 \quad \text{and} \quad \sup_{|h|=1} \int_{\mathbb{R}^d} |\langle x, h \rangle|^{|s|} \nu(dx) \leq C_\nu, \quad \text{if } s < 0, \quad (4.1)$$

$$\alpha := \nu(\mathbb{R}^d) \in (0, \infty) \quad \text{and} \quad |\mathcal{F}\nu(u)|^2 \leq C_\nu (1 + |u|^2)^{-s}, \quad u \in \mathbb{R}^d, \quad \text{if } s \geq 0.$$

- (ii)  $\int_{\mathbb{R}^d} |x|^p \nu(dx) < \infty$  for some  $p > 0$ .

**Assumption C.** Let the time-change  $\mathcal{T}$  satisfy:

- (i)  $\mathbb{E}[\mathcal{T}^p(1)] < \infty$  for some  $p > 0$ .
- (ii) The sequence  $T_j = \mathcal{T}(j) - \mathcal{T}(j - 1)$ ,  $j \in \mathbb{N}$ , is mutually independent and identically distributed with some law  $\pi$  on  $\mathbb{R}_+$ .
- (iii) The derivatives of the Laplace transform  $\mathcal{L}(z) = \int_0^\infty e^{-tz} \pi(dt)$  satisfy  $|\mathcal{L}''(z)/\mathcal{L}'(z)| \leq C_L$  for some  $C_L > 0$  for all  $z \in \mathbb{C}$  with  $\text{Re}(z) > 0$ .

If the Laplace transform decays polynomially, we may impose the stronger assumption

- (iv)  $|\mathcal{L}''(z)/\mathcal{L}'(z)| \leq C_L(1 + |z|)^{-1}$  for some  $C_L > 0$  for all  $z \in \mathbb{C}$  with  $\text{Re}(z) > 0$ .

**Remarks 4.1.**

- (i) For  $s \in (-2, 0)$  Assumption B(i) allows for Lévy processes with infinite jump activity. In that case  $|s|$  in (4.1) is an upper bound for the Blumenthal–Gettoor index of the process. A more pronounced singularity of the jump measure  $\nu$  at zero corresponds to larger values of  $|s|$ . The lower bound  $s > -2$  is natural, since any jump measure satisfies  $\int_{\mathbb{R}^d} (|x|^2 \wedge 1) \nu(dx)$ . On the contrary,  $s = 0$  in (4.1) implies that  $\nu$  is a finite measure. In that case we could further profit from its regularity which we measure by the decay of its Fourier transform. Altogether, the parameter  $s$  will determine the approximation error that is due to the term  $|u|^{-2}\Psi(u)$  in (3.2).
- (ii) Assumption B(ii) implies that  $\mathbb{E}[|X_t|^p] < \infty$  for all  $t > 0$  and together with the moment condition Assumption C(i), we conclude that  $\mathbb{E}[|Y_k|^p] < \infty$ , cf. Lemma 8.1. Assumption C(ii) implies that the increments  $Y_j = X_{\mathcal{T}(j)} - X_{\mathcal{T}(j-1)}$  are independent and identically distributed. Note that  $(T_j)$  being stationary with invariant measure  $\pi$  is necessary to construct the estimator of  $\Sigma$ . The independence can, however, be relaxed to an  $\alpha$ -mixing condition as discussed in Section 5.2. Finally, Assumptions C(iii) and (iv) are used to linearise the stochastic error.

In the sequel we denote by  $B_U^d := \{u \in \mathbb{R}^d : |u| \leq U\}$  the  $d$ -dimensional ball of radius  $U > 0$ . We can now state the first main result of this section.

**Theorem 4.2.** *Grant Assumptions B and C(i)–(iii). If also Assumption C(iv) is fulfilled, we set  $q = 1$  and otherwise let  $q = 0$ . Then for any  $n, d \in \mathbb{N}$  and  $U \geq 1, \kappa > 0$  satisfying*

$$\sqrt{d \log(d + 1)}(\log U) \leq \kappa \leq \sqrt{n}U^{-d/2} \|\varphi\|_{L^1(B_U^d)}^{1/2} \frac{\inf_{|u| \leq U} |\psi(u)|^q |\mathcal{L}'(-\psi(u))|^2}{\inf_{|u| \leq U} |\mathcal{L}'(-\psi(u))|},$$

the matrix  $\mathcal{R}_n$  from (3.6) satisfies for some constants  $c, D > 0$  depending only on  $w, C_\nu$  and  $C_L$

$$\mathbb{P}\left(\|\mathcal{R}_n\|_\infty \geq \frac{\kappa \|\varphi\|_{L^1(B_U^d)}^{1/2}}{\sqrt{n}U^{2+d/2} \inf_{|u| \leq U} |\mathcal{L}'(-\psi(u))|} + DU^{-(s+2)}\right) \leq 2(d + 1)e^{-c\kappa^2}.$$

In particular,  $\mathbb{P}(\|\mathcal{R}_n\|_\infty > \lambda) \leq 2(d + 1)e^{-c\kappa^2}$  if

$$\lambda \geq \frac{\kappa \|\varphi\|_{L^1(B_U^d)}^{1/2}}{\sqrt{n}U^{2+d/2} \inf_{|u| \leq U} |\mathcal{L}'(-\psi(u))|} + DU^{-(s+2)}.$$

In order to prove this theorem, we decompose  $\mathcal{R}_n$  into a stochastic error term and a deterministic approximation error. More precisely, the regression formula (3.2) and  $\|\tilde{\Theta}(u)\|_\infty = \max\{1, 2U^2|u|^{-2}\} \leq 1$  for any  $u \in \text{supp } w_U$  yield

$$\begin{aligned} \|\mathcal{R}_n\|_\infty &\leq 4 \left\| \int_{\mathbb{R}^d} |u|^{-2} \text{Re}(\mathcal{L}^{-1}(\varphi_n(u)) - \mathcal{L}^{-1}(\varphi(u))) \tilde{\Theta}(u) w_U(u) du \right\|_\infty \\ &\quad + 4 \int_{\mathbb{R}^d} \frac{|\text{Re } \Psi(u) + \alpha|}{|u|^2} w_U(u) du. \end{aligned} \tag{4.2}$$



The order of the deterministic error is  $\mathcal{O}(U^{-s-2})$ , cf. Lemma 8.3, which decays as  $U \rightarrow \infty$ . The rate deteriorates for higher jump activities. This is reasonable since even for high frequency observations it is very difficult to distinguish between small jumps and fluctuations due to the diffusion component, cf. [21]. The stochastic error is dominated by its linearisation

$$\left\| \int_{\mathbb{R}^d} \frac{1}{|u|^2} \operatorname{Re} \left( \frac{\varphi_n(u) - \varphi(u)}{\mathcal{L}'(-\psi(u))} \right) \tilde{\Theta}(u) w_U(u) \, du \right\|_{\infty},$$

which is increasing in  $U$  due to the denominator  $|u|^2 \mathcal{L}'(-\psi(u)) \rightarrow 0$  as  $|u| \rightarrow \infty$ . To obtain a sharp oracle inequality for the spectral norm of the linearised stochastic error, we use the noncommutative Bernstein inequality by Recht [32]. To bound the remainder, we apply a concentration result (Theorem A.4) for the empirical characteristic function around  $\varphi$ , uniformly on  $B_U^d$ .

The lower bound on  $\kappa$  reflects the typical dependence on the dimension  $d$  that also appear is in theory on matrix completion, cf. Corollary 2 by Koltchinskii et al. [24]. Our upper bound on  $\kappa$  ensures that the remainder term in the stochastic error is negligible.

The choice of  $U$  is determined by the trade-off between approximation error and stochastic error. Since neither  $U$  nor  $\lambda$  depend on the rank of  $\Sigma$ , the theorem verifies that the estimator  $\tilde{\Sigma}_{n,\lambda}$  is adaptive with respect to  $\operatorname{rank}(\Sigma)$ . Supposing lower bounds for  $\mathcal{L}'(-\psi(u))$  that depend only on the radius  $|u|$ , the only term that depends on the dimension is  $E_U := U^{-d} \|\varphi\|_{L^1(B_U^d)}$ . Owing to  $\|\varphi\|_{\infty} \leq 1$ , it is uniformly bounded by the volume of the unit ball  $B_1^d$  in  $\mathbb{R}^d$  which in turn is uniformly bounded in  $d$  (in fact it is decreasing as  $d \rightarrow \infty$ ). If  $\varphi \in L^1(\mathbb{R}^d)$ , we even have  $E_U \lesssim U^{-d}$ . We discuss this quite surprising factor further after Corollary 4.4 and, from a more abstract perspective, in Section 5.5.

For specific decay behaviours of  $\mathcal{L}(z)$  we can now conclude convergence rates for the diffusion matrix estimator. In contrast to the nonparametric estimation of the coefficients of a diffusion process, as studied by Chorowski and Trabs [15], the convergence rates depend enormously on the sampling distribution (resp. time-change). We start with an exponential decay of the Laplace transform of  $\pi$ . This is especially the case if the Lévy process  $X$  is observed at equidistant time points  $\Delta j$  for some  $\Delta > 0$  and thus  $\mathcal{L}(v) = e^{\Delta v}$ .

**Corollary 4.3.** *Grant Assumptions B and C(i)–(iii) and let  $d = o(n/(2 \log \log n))$ . Suppose that  $0 \neq \Sigma \in \mathbb{M}$ , where  $\mathbb{M}$  is a convex subset of positive semi-definite  $d \times d$  matrices and let  $\alpha \in I$ . If  $|\mathcal{L}'(z)| \gtrsim \exp(-a|z|^\eta)$  for  $z \in \mathbb{C}$  with  $\operatorname{Re}(z) > 0$  and for some  $a, \eta > 0$ , we set*

$$U = \left( \frac{\tau}{a(\|\Sigma\|_{\infty} + C_v + \alpha)} \log \left( \frac{n}{d \log(d+1)} \right) \right)^{1/(2\eta)} \vee 2 \quad \text{and}$$

$$\lambda = C (\|\Sigma\|_{\infty} + C_v + \alpha)^{(s+2)/(2\eta)} \left( \log \frac{n}{d \log(d+1)} \right)^{-(s+2)/(2\eta)}$$

for some  $\tau < 1/2$  and a constant  $C > 0$  which can be chosen independently of  $\gamma$ ,  $\Sigma$  and  $v$ . Then we have for sufficiently large  $n$

$$\|\tilde{\Sigma}_{n,\lambda} - \Sigma\|_2 \leq C (\|\Sigma\|_{\infty} + C_v + \alpha)^{(s+2)/(2\eta)} \sqrt{\operatorname{rank}(\Sigma)} \left( \log \frac{n}{d \log(d+1)} \right)^{-(s+2)/(2\eta)}$$

with probability larger than  $1 - 2(d+1)e^{-cd(\log \log n)^2}$  for some  $c > 0$  depending only on  $w$ ,  $C_v$  and  $C_L$ .

The assertion follows from Corollary 3.5 and Theorem 4.2 (setting  $\kappa = 2\sqrt{d \log(d+1)}(\log \log n)$ ), taking into account that due to (2.1) (and (8.1)) we have

$$\sup_{|u| \leq U} |\psi(u)| \leq \|\Sigma\|_{\infty} U^2/2 + 2C_v U^{-(s \wedge 0)} + \alpha \leq \frac{U^2}{2} (\|\Sigma\|_{\infty} + C_v + \alpha).$$

This corollary shows that the exponential decay of  $\mathcal{L}'$  leads to a severely ill-posed problem and, as a consequence, the rate is only logarithmic. Therein, the interplay between sample size  $n$  and dimension  $d$ , i.e. the term  $\frac{n}{d \log(d+1)}$  has

been also observed for matrix completion [24, Corollary 2]. Whenever there is some  $\rho \in (0, 1)$  such that  $d \lesssim n^\rho$  the logarithmic term in the upper bound simplifies to  $(\log n)^{-(s+2)/(2\eta)}$ . The slow rate implies that  $\tilde{\Sigma}_{n,\lambda}$  is consistent in absolute error only if  $\text{rank } \Sigma = o((\log n)^{(s+2)/\eta})$ . Considering the relative error  $\|\tilde{\Sigma}_{n,\lambda} - \Sigma\|_2 / \sqrt{\text{rank}(\Sigma)}$ , this restriction vanishes. Note also that in a high-dimensional principal component analysis, the spectral norm  $\|\Sigma\|_\infty$  may grow in  $d$ , too, cf. the discussion for the factor model by Fan et al. [18].

If  $\mathcal{L}'$  decays only polynomially, as for instance for the gamma subordinator, cf. Examples 2.1, the estimation problem is only mildly ill-posed and the convergence rates are polynomial in  $n$ .

**Corollary 4.4.** *Grant Assumptions B and C(i)–(iv) and assume  $d \log(d + 1) = o(n(\log n)^{-2})$ . For a convex subset  $\mathbb{M} \subseteq \mathbb{R}^{d \times d}$  of positive semi-definite matrices, let  $0 \neq \Sigma \in \mathbb{M}$ ,  $\text{rank } \Sigma = k$  and  $\alpha \in I$ . Suppose  $|\mathcal{L}(z)| \lesssim (1 + |z|)^{-\eta}$  and  $|\mathcal{L}'(z)| \gtrsim |z|^{-\eta-1}$  for  $z \in \mathbb{C}$  with  $\text{Re}(z) > 0$  and for some  $\eta > 0$  such that  $s + 2 > (2\eta \wedge k)$ . Denoting the smallest strictly positive eigenvalue of  $\Sigma$  by  $\lambda_{\min}(\Sigma)$ , we set*

$$U = \left( \frac{n}{(\log n)^2 d \log(d + 1)} \right)^{1/(2s+4+4\eta-(2\eta \wedge k))} \quad \text{and}$$

$$\lambda = C(\|\Sigma\|_\infty + C_v + \alpha)^{\eta+1} \lambda_{\min}(\Sigma)^{-(2\eta \wedge k)/4} \left( \frac{n}{(\log n)^2 d \log(d + 1)} \right)^{-(s+2)/(2s+4+4\eta-(2\eta \wedge k))}$$

for a constant  $C > 0$  independent of  $\gamma$ ,  $\Sigma$  and  $v$ . Then we have for sufficiently large  $n$

$$\|\tilde{\Sigma}_{n,\lambda} - \Sigma\|_2 \leq C \frac{(\|\Sigma\|_\infty + C_v + \alpha)^{\eta+1}}{\lambda_{\min}(\Sigma)^{(2\eta \wedge k)/4}} \sqrt{\text{rank}(\Sigma)} \left( \frac{n}{(\log n)^2 d \log(d + 1)} \right)^{-(s+2)/(2s+4+4\eta-(2\eta \wedge k))}$$

with probability larger than  $1 - 2(d + 1)e^{-cd(\log n)^2}$  for some  $c > 0$  depending only on  $w$ ,  $C_v$  and  $C_L$ .

**Remarks 4.5.**

- (i) The convergence rate reflects the regularity  $s + 2$  and the degree of ill-posedness  $2\eta = 2(\eta + 1) - 2$  of the statistical inverse problem, where  $2(\eta + 1)$  is the decay rate of the characteristic function and we gain two degrees since  $\langle u, \Sigma u \rangle$  grows like  $|u|^2$ . The term  $-(2\eta) \wedge k$  appearing in the denominator is very surprising since the rate becomes faster as the rank of  $\Sigma$  increases up to some critical threshold value  $2\eta$ . To see this, it remains to note that the assumption  $\text{rank } \Sigma = k$  yields

$$\begin{aligned} U^{-d} \|\varphi\|_{L^1(B_{U^d}^d)} &= U^{-d} \int_{|u| \leq U} |\mathcal{L}(-\psi(u))| \, du \\ &\lesssim U^{-d} \int_{|u| \leq U} (1 + |\psi(u)|)^{-\eta} \, du \\ &\lesssim U^{-k} \int_{u \in \mathbb{R}^k: |u| \leq U} (1 + \lambda_{\min}(\Sigma)|u|^2)^{-\eta} \, du \lesssim \frac{(\sqrt{\lambda_{\min}(\Sigma)}U)^{-2\eta+k} \vee 1}{(\sqrt{\lambda_{\min}(\Sigma)}U)^k}. \end{aligned}$$

In order to shed some light on this phenomenon, we consider a related regression problem in Section 5.5.

- (ii) It is interesting to note that for very small  $\eta$  we almost attain the parametric rate. Having the example of gamma-distributed increments  $T_j$  in mind, a small  $\eta$  corresponds to measures  $\pi$  which are highly concentrated at the origin. In that case we expect many quite small increments  $X_{\mathcal{T}(j)} - X_{\mathcal{T}(j-1)}$  where the jump component has only a small effect. On the other hand, for the remaining few rather large increments, the estimator is not worse than in a purely low-frequency setting. Hence,  $|\mathcal{L}(z)| \sim (1 + |z|)^{-\eta}$  heuristically corresponds to an interpolation between high- and low-frequency observations, cf. also [23].
- (iii) If the condition  $s + 2 \geq 2\eta \wedge \text{rank } \Sigma$  is not satisfied, the linearised stochastic error appears to be smaller than the remainder of the linearisation. In that case we only obtain the presumably suboptimal rate  $n^{-(s+2)/(s+2+4\eta)}$  (for fixed  $d$ ). It is still an open problem whether this condition is only an artefact of our proofs or if there is some intrinsic reason.

Let us now investigate whether the rates are optimal in the minimax sense. The dependence on the the rank of  $\Sigma$  is the same as usual in matrix estimation problems and seems natural. The optimality of the influence of the dimension  $d$  is less clear. In lower bounds for matrix completion, cf. [24, Theorem 6], a similar dependence appears except for the fact that we face non-parametric rates. In the following we prove lower bounds for the rate of convergence in the number of observations  $n$  for a fixed  $d$ . More general lower bounds in a high-dimensional setting where  $d$  may grow with  $n$  are highly intricate since usually lower bounds for high-dimensional statistical problems are based on Kullback–Leibler divergences in Gaussian models – an approach that cannot be applied in our model. This problem is left open for future research.

Let us introduce the class  $\mathfrak{S}(s, p, C_\nu)$  of all Lévy measures satisfying Assumption B with  $s \in (-2, \infty)$ ,  $p > 2$  and a constant  $C_\nu > 0$ . In order to prove sharp lower bounds, we need the following stronger assumptions on the time change:

**Assumption D.** Grant (i) and (ii) from Assumption C. For  $C_L > 0$  and  $L \in \mathbb{N}$  let the Laplace transform  $\mathcal{L}(z)$  satisfy

(m) for some  $\eta > 0$  and all  $z > 0$

$$|\mathcal{L}'(z)| \leq C_L(1 + |z|)^{-\eta-1}, \quad |\mathcal{L}^{(l+1)}(z)/\mathcal{L}^{(l)}(z)| \leq C_L(1 + |z|)^{-1}, \quad l = 1, \dots, L,$$

or

(s) for some  $a, \eta > 0$  and all  $z \in \mathbb{C}$  with  $\text{Re}(z) > 0$

$$|\mathcal{L}'(z)| \leq C_L e^{-a|z|^\eta}, \quad |\mathcal{L}^{(l+1)}(z)/\mathcal{L}^{(l)}(z)| \leq C_L, \quad l = 1, \dots, L.$$

**Theorem 4.6.** Fix  $d \geq 1$  and let  $s \in (-2, \infty)$ ,  $p > 2$ ,  $C_\nu > 0$  and  $k \in \{1, \dots, d\}$ .

(i) Suppose  $2\eta > k$  and  $k \leq s$ . Under Assumption D(m) with  $L > \frac{k+p\nu(-s)}{2} \vee k$ , it holds for any  $\varepsilon > 0$  that

$$\liminf_{n \rightarrow \infty} \inf_{\widehat{\Sigma}} \sup_{\substack{\text{rank}(\Sigma)=k \\ v \in \mathfrak{S}(s, p, C_\nu)}} \mathbb{P}_{(\Sigma, v, \mathcal{T})}^{\otimes n} (\|\widehat{\Sigma} - \Sigma\|_2 > \varepsilon n^{-(s+2)/(2(s+2)+4\eta-k)}) > 0.$$

Moreover, if  $\eta > 1$ , we have under Assumption D(m) with  $L > \frac{1+p\nu(-s)}{2} \vee 1$  for any  $\varepsilon > 0$

$$\liminf_{n \rightarrow \infty} \inf_{\widehat{\Sigma}} \sup_{\substack{2\eta \leq \text{rank}(\Sigma) \leq d \\ v \in \mathfrak{S}(s, p, C_\nu)}} \mathbb{P}_{(\Sigma, v, \mathcal{T})}^{\otimes n} (\|\widehat{\Sigma} - \Sigma\|_2 > \varepsilon n^{-(s+2)/(2(s+2)+2\eta)}) > 0.$$

(ii) Under the Assumption D(s) with  $L > \frac{1+p\nu(-s)}{2}$  it holds for any  $\varepsilon > 0$  that

$$\liminf_{n \rightarrow \infty} \inf_{\widehat{\Sigma}} \sup_{\substack{0 < \text{rank}(\Sigma) \leq d \\ v \in \mathfrak{S}(s, p, C_\nu)}} \mathbb{P}_{(\Sigma, v, \mathcal{T})}^{\otimes n} (\|\widehat{\Sigma} - \Sigma\|_2 > \varepsilon (\log n)^{-(s+2)/(2\eta)}) > 0.$$

Note that the infima are taken over all estimators of  $\Sigma$  based on  $n$  independent observations of the random variable  $Y_1$  whose law we denoted by  $\mathbb{P}_{(\Sigma, v, \mathcal{T})}$ .

Up to a logarithmic factors in the mildly ill-posed case, the upper and lower bounds coincide and thus our convergence rates are minimax optimal for fixed  $d$ .

### 5. Discussions and extensions

In this section we discuss several generalisations of the previously developed theory.

5.1. *Estimating the time-change*

In order to define the diffusion matrix estimator we have assumed that the Laplace transform of the time-change increments is known. In the context of random observation times it is reasonable to suppose that we additionally observe  $\mathcal{T}(1), \dots, \mathcal{T}(m)$  for some sample size  $m$ . Using empirical process theory, it is not difficult to show that

$$\mathcal{L}_m(z) := \frac{1}{m} \sum_{j=1}^m e^{-z(\mathcal{T}(j) - \mathcal{T}(j-1))}$$

converges uniformly to  $\mathcal{L}(z)$  for  $z \in \mathbb{R}_+$  with  $\sqrt{n}$ -rate as exploited by Chorowski and Trabs [15]. This result could be generalised to appropriate arguments on the complex plane and may allow for replacing  $\mathcal{L}$  by  $\mathcal{L}_n$  for the estimator of the characteristic exponent (3.1).

More precisely, we need to invert the function  $\mathcal{L}_m$ . This is not a trivial problem, since the inverse function  $\mathcal{L}_m^{-1}$  should be well defined on a complex plane (we are going to plug in a complex-valued function  $\varphi_n(u)$ ). We use the Lagrange inversion formula to compute the derivatives of  $\mathcal{L}_m^{-1}(z)$  at point  $z = 1$  and then expand this function into Taylor series around  $z = 1$ . This approximation turns out to be very stable and can even improve the quality of estimating  $\Sigma$ , see Section 6.2. We have a formal expansion

$$\mathcal{L}_m(z) = \sum_{k=1}^{\infty} M_k \frac{(-z)^k}{k!}, \quad M_k := M_k(m) = \frac{1}{m} \sum_{j=1}^m (\mathcal{T}(j) - \mathcal{T}(j-1))^k.$$

In the case of a continuous distribution of  $\mathcal{T}(1)$ ,  $\mathcal{L}'_m(z) > 0$  for any  $z \in \mathbb{R}$  with probability one and the inverse function  $\mathcal{L}_m^{-1}(z)$  exists and is unique. In some vicinity of the point  $z = 1$  in the complex plan, we obtain the expansion

$$\mathcal{L}_m^{-1}(z) = \sum_{j=1}^{\infty} H_j \frac{(z-1)^j}{j!} \tag{5.1}$$

with

$$H_1 = \frac{1}{M_1}, \quad H_j = \frac{1}{M_1^j} \sum_{k=1}^{j-1} (-1)^k j^{(k)} B_{j-1,k}(\widehat{M}_1, \widehat{M}_2, \dots, \widehat{M}_{j-k}), \quad j \geq 2,$$

$$\widehat{M}_k := \frac{M_{k+1}}{(k+1)M_1} \quad \text{and} \quad j^{(k)} := j(j+1) \dots (j+k-1),$$

where  $B_{j-1,k}$  are partial Bell polynomials. Formula (5.1) can then be used to obtain another estimator for the characteristic exponent which adapts to the unknown Laplace transform:

$$\tilde{\psi}_{n,m}(u) = -\mathcal{L}_m^{-1}(\varphi_n(u)).$$

We investigate the numerical performance of the this approach in Section 6.2.

5.2. *Mixing time-change*

Motivated by the integrated CIR process from Example 2.1(iv), for instance used by Carr et al. [14] to model stock price processes, we will generalise the results from the previous section to time-changes  $\mathcal{T}$ , whose increments are not i.i.d., but form a strictly stationary  $\alpha$ -mixing sequence. Recall that the strong mixing coefficients of the sequence  $(T_j)$  are defined by

$$\alpha_T(n) := \sup_{k \geq 1} \alpha(\mathcal{M}_k, \mathcal{G}_{k+n}), \quad \alpha(\mathcal{M}_k, \mathcal{G}_\ell) := \sup_{A \in \mathcal{M}_k, B \in \mathcal{G}_\ell} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

for  $\mathcal{M}_k := \sigma(T_j : j \leq k)$  and  $\mathcal{G}_k := \sigma(T_j : j \geq k)$ . We replace Assumption C by the following

**Assumption E.** Let the time-change  $\mathcal{T}$  satisfy:

- (i)  $\mathbb{E}[\mathcal{T}^p] < \infty$  for some  $p > 2$ .
- (ii) The sequence  $T_j = \mathcal{T}(j) - \mathcal{T}(j - 1)$ ,  $j \in \mathbb{N}$ , is strictly stationary with invariant measure  $\pi$  and  $\alpha$ -mixing with

$$\alpha_{\mathcal{T}}(j) \leq \bar{\alpha}_0 \exp(-\bar{\alpha}_1 j), \quad j \in \mathbb{N},$$

for some positive constants  $\bar{\alpha}_0$  and  $\bar{\alpha}_1$ .

- (iii) The Laplace transform  $\mathcal{L}$  satisfies  $|\mathcal{L}''(z)/\mathcal{L}'(z)| \leq C_L$  for some  $C_L > 0$  for all  $z \in \mathbb{C}$  with  $\text{Re}(z) > 0$ .

If  $T_j$  is  $\alpha$ -mixing, Lemma 7.1 by Belomestny [6] shows that the sequence  $(Y_j)$  inherits the mixing-property from the sequence  $(T_j)$ . In combination with the finite moments  $\mathbb{E}[|Y_j|^p] < \infty$  for  $p > 2$  we can apply a concentration inequality on the empirical characteristic function  $\varphi_n$ , see Theorem A.4 below, which follows from the results by Merlevède et al. [27].

In the  $\alpha$ -mixing case, a noncommutative Bernstein inequality is not known (at least to the authors' knowledge) and thus we cannot hope for a concentration inequality for  $\|\mathcal{R}_n\|_\infty$  analogous to Theorem 4.2. Possible workarounds are either to estimate

$$\|\mathcal{R}_n\|_\infty \leq 2 \int_{\mathbb{R}^d} \left| \frac{2}{|u|^2} \text{Re} \widehat{\psi}_n(u) - \langle \tilde{\Theta}(u), \Sigma \rangle \right| \|\tilde{\Theta}(u)\|_\infty w_U(u) du, \tag{5.2}$$

where  $\|\tilde{\Theta}(u)\|_\infty = 1$ , or to bound  $\|\mathcal{R}_n\|_\infty \leq (d + 1)\|\mathcal{R}_n\|_{\max}$  for the maximum entry norm  $\|A\|_{\max} = \max_{ij} |A_{ij}|$  for  $A \in \mathbb{R}^{(d+1) \times (d+1)}$ . While the former estimate leads to suboptimal rates for polynomially decaying  $\mathcal{L}$ , we loose a factor  $d$  in the latter bound which is critical in a high-dimensional setting.

Having in mind that the Laplace transform of the integrated CIR process decays exponentially, we pursue the first idea and obtain the following concentration result:

**Theorem 5.1.** *Grant Assumptions B and E and let  $\rho > 1/2$ . There are constants  $\underline{\xi}, \bar{\xi} > 0$  such that for any  $n \in \mathbb{N}$  and  $U, \kappa > 0$  satisfying*

$$\underline{\xi} \sqrt{d \log n} < \kappa < \bar{\xi} (\log U)^{-\rho} \inf_{|u| \leq U} |\mathcal{L}'(-\psi(u))| (\log n)^{-1/2} \sqrt{n}$$

*the matrix  $\mathcal{R}_n$  from (3.6) satisfies for some constants  $c, C, D > 0$  depending only on  $w, C_v$  and  $C_L$  that*

$$\mathbb{P}\left(\|\mathcal{R}_n\|_\infty \geq \frac{\kappa (\log U)^\rho}{\sqrt{n} U^2 \inf_{|u| \leq U} |\mathcal{L}'(-\psi(u))|} + D U^{-s-2}\right) \leq C e^{-c\kappa^2} + C n^{-\rho/2}.$$

*In particular, we have  $\mathbb{P}(\|\mathcal{R}_n\|_\infty > \lambda) \leq C e^{-c\kappa^2} + C n^{-\rho/2}$  if*

$$\lambda \geq \frac{\kappa (\log U)^\rho}{\sqrt{n} U^2 \inf_{|u| \leq U} |\mathcal{L}'(-\psi(u))|} + D U^{-s-2}.$$

The suboptimal term  $\sqrt{\log n}$  in lower bound of  $\kappa$  comes from the estimate (5.2). The term  $(\log U)^\rho$  could be omitted with a more precise estimate of the linearised stochastic error term (similar to the proof of Theorem 4.2), but let us keep the proof of Theorem 5.1 simple, see Section 8.3. For exponentially decaying Laplace transforms the resulting rate is already sharp and coincides with our results for the independent case as soon as  $d \leq n^\tau$  for some  $\tau \in [0, 1)$ .

**Corollary 5.2.** *Grant Assumptions B and E, assume  $d = o(n/\log n)$  and let  $\mathbb{M}$  be a convex subset of positive semi-definite  $d \times d$  matrices. Suppose that  $0 \neq \Sigma \in \mathbb{M}$  and  $\alpha \in I$ . Assume  $|\mathcal{L}'(z)| \gtrsim \exp(-a|z|^\eta)$  for  $z \in \mathbb{C}$  with  $\text{Re}(z) > 0$  and for  $a, \eta > 0$ . Choosing*

$$U = \left( \frac{\tau \log(n/(d \log n))}{a(\|\Sigma\|_\infty + C_v + \alpha)} \right)^{1/(2\eta)} \vee 2 \quad \text{and} \quad \lambda = C_1 (\|\Sigma\|_\infty + C_v + \alpha)^{(s+2)/(2\eta)} \left( \log \frac{n}{d \log n} \right)^{-(s+2)/(2\eta)},$$

for some  $\tau < 1/2$  and a constant  $C_1 > 0$ , we have

$$\|\tilde{\Sigma}_{n,\lambda} - \Sigma\|_2 \leq C_1 (\|\Sigma\|_\infty + C_v + \alpha)^{(s+2)/(2\eta)} \sqrt{\text{rank}(\Sigma)} \left( \log \frac{n}{d \log n} \right)^{-(s+2)/(2\eta)}$$

with probability larger than  $1 - C_2(e^{-c\sqrt{d \log n}} + n^{-p/2})$  and for some  $C_2, c > 0$ . The constants  $C_1, C_2, c$  depend only on  $w, C_L, C_v$ .

### 5.3. Low rank plus sparse matrix

In view of Fan et al. [17,18] it might be interesting in applications to relax the low rank assumption on the diffusion matrix  $\Sigma$  to case where  $\Sigma = \Sigma_r + \Sigma_s$  for a low-rank matrix  $\Sigma_r \in \mathbb{R}^{d \times d}$  and a sparse matrix  $\Sigma_s \in \mathbb{R}^{d \times d}$ . The underlying idea is that  $\Sigma_r$  reflects a low dimensional factor model while  $\Sigma_s$  represents a sparse error covariance matrix. Sparsity in  $\Sigma_s$  means here that most entries of  $\Sigma_s$  are zero or very small, see for instance [11] for this approach to sparsity. More precisely, let us assume that the entry-wise  $\ell_q$ -norm  $|\Sigma_s|_q$  for some  $q \in [0, 2)$  is small. For the sake of clarity, we focus on extending the estimator  $\hat{\Sigma}_{n,\lambda}$  from (3.3) while it can be easily seen how to modify the following for the bias corrected estimator  $\tilde{\Sigma}_{n,\lambda}$ .

In terms of the vectorisation operator  $\text{vec}(A) = (a_{11}, \dots, a_{d1}, a_{21}, \dots, a_{dd})^\top$  for any  $A = (a_{i,j}) \in \mathbb{R}^{d \times d}$  we rewrite

$$\langle \Theta(u), \Sigma_s \rangle = \text{tr}(\Theta(u)^\top \Sigma_s) = \text{vec}(\Theta(u))^\top \text{vec}(\Sigma_s) = \langle \text{diag}(\text{vec}(\Theta(u))), \text{diag}(\text{vec}(\Sigma_s)) \rangle.$$

Hence, defining the set  $\mathbb{D} := \{\text{diag}(A, \text{diag}(a)) : A \in \mathbb{R}^{d \times d}, a \in \mathbb{R}^{d^2}\} \subseteq \mathbb{R}^{(d+d^2) \times (d+d^2)}$  and the matrices

$$\overline{\Theta}(u) := \text{diag}(\Theta(u), \text{diag}(\text{vec}(\Theta(u)))) \in \mathbb{D} \quad \text{and} \quad \overline{\Sigma} := \text{diag}(\Sigma_r, \text{diag}(\text{vec}(\Sigma_s))) \in \mathbb{D},$$

we obtain the representation

$$\langle \Theta(u), \Sigma \rangle = \langle \Theta(u), \Sigma_r \rangle + \langle \Theta(u), \Sigma_s \rangle = \langle \overline{\Theta}(u), \overline{\Sigma} \rangle.$$

Motivated by this reformulation, we introduce the estimator  $\overline{\Sigma}_{n,\lambda} := \overline{\Sigma}_r + \overline{\Sigma}_s$  where

$$\text{diag}(\overline{\Sigma}_r, \text{diag}(\text{vec}(\overline{\Sigma}_s))) := \arg \min_{M \in \overline{\mathbb{M}}} \left\{ \int_{\mathbb{R}^d} (2|u|^{-2} \text{Re} \widehat{\psi}_n(u) + \langle \overline{\Theta}(u), M \rangle)^2 w_U(u) du + \lambda \|M\|_1 \right\} \tag{5.3}$$

for a subset  $\overline{\mathbb{M}} \subseteq \mathbb{D}$ . Since  $\text{rank}(\overline{\Sigma}) = \text{rank}(\Sigma_r) + |\Sigma_s|_0$ , the nuclear norm penalisation in (5.3) will enforce a low rank structure in  $\overline{\Sigma}_r$  and a sparse structure in  $\overline{\Sigma}_s$ . In order to carry over the oracle inequalities, we first need verify the isometry property for the modified weighted norm

$$\langle A, B \rangle_{\overline{w}} := \int_{\mathbb{R}^d} \langle \overline{\Theta}(u), A \rangle \langle \overline{\Theta}(u), B \rangle w_U(u) du, \quad A, B \in \mathbb{D}.$$

Indeed, we easily deduce from Lemma 3.2 and  $|\Theta(u)|_2 = 1$ :

**Lemma 5.3.** *For any symmetric matrix  $A \in \mathbb{D}$  we have*

$$\underline{\kappa}_{\overline{w}} \|A\|_2 \leq \|A\|_{\overline{w}} \leq \overline{\kappa}_{\overline{w}} \|A\|_2,$$

where  $\underline{\kappa}_{\overline{w}}^2 := \min_{i,j=1,2} \int_{\mathbb{R}^d} v_i^2 v_j^2 / (|v|^4) w(v) dv$  and  $\overline{\kappa}_{\overline{w}}^2 := 2 \int_{\mathbb{R}^d} w(v) dv$ .

We recover the exact structure for which Theorem 3.4 has been proven and thus conclude:

**Proposition 5.4.** *Suppose that  $\overline{\mathbb{M}} \subseteq \mathbb{D}$  is a convex subset of the symmetric matrices. Define*

$$\overline{\mathcal{R}}_n := 2 \int_{\mathbb{R}^d} \left( \frac{2}{|u|^2} \text{Re} \widehat{\psi}_n(u) - \langle \overline{\Theta}(u), \overline{\Sigma} \rangle \right) \overline{\Theta}(u) w_U(u) du \in \mathbb{D}. \tag{5.4}$$

On the event  $\{\|\overline{\mathcal{R}}_n\|_\infty \leq \lambda\}$  for some  $\lambda > 0$  the estimator  $\overline{\Sigma}_{n,\lambda} = \overline{\Sigma}_r + \overline{\Sigma}_s$  given in (5.3) satisfies

$$\|\text{diag}(\overline{\Sigma}_r, \text{diag}(\text{vec}(\overline{\Sigma}_s))) - \overline{\Sigma}\|_w^2 \leq \inf_{M \in \overline{\mathbb{M}}} \left\{ \|M - \overline{\Sigma}\|_w^2 + \left( \frac{1 + \sqrt{2}}{2\underline{\mathcal{Z}}_w} \right)^2 \lambda^2 \text{rank}(M) \right\}$$

with  $\underline{\mathcal{Z}}_w$  from Lemma 5.3.

Noting  $\|\overline{\Sigma}\|_2^2 = \|\Sigma_r\|_2^2 + \|\Sigma_s\|_2^2 \geq \frac{1}{2}\|\Sigma\|_2^2$  and using Lemma 5.3, we obtain an oracle inequality for the estimation error of  $\overline{\Sigma}_{n,\lambda}$  in the Frobenius norm. The extension from  $q = 0$  to  $q \in (0, 2)$  can be proven as Corollary 2 by Rigollet and Tsybakov [33].

**Theorem 5.5.** *Let  $\overline{\mathbb{M}} \subseteq \mathbb{D}$  is a convex subset of the symmetric matrices. On the event  $\{\|\overline{\mathcal{R}}_n\|_\infty \leq \lambda\}$  for some  $\lambda > 0$  and for  $\overline{\mathcal{R}}_n$  from (5.4) we have for a constant  $C_w$  depending only on  $\underline{\mathcal{Z}}_w$  and  $\overline{\mathcal{Z}}_w$ :*

$$\begin{aligned} \|\overline{\Sigma}_{n,\lambda} - \Sigma\|_2^2 &\leq 2(\|\overline{\Sigma}_r - \Sigma_r\|_2^2 + \|\overline{\Sigma}_s - \Sigma_s\|_2^2) \\ &\leq C_w \inf_{M_r, M_s} \left\{ \|M_r - \Sigma_r\|_2^2 + \|M_s - \Sigma_s\|_2^2 + \lambda^2(\text{rank}(M_r) + |M_s|_0) \right\}, \end{aligned}$$

where the infimum is taken over all  $M_r, M_s$  such that  $\text{diag}(M_r, \text{diag}(\text{vec}(M_s))) \in \overline{\mathbb{M}}$ . Moreover, for another constant  $C'_w$  and any  $q \in (0, 2)$  we have

$$\begin{aligned} \|\overline{\Sigma}_{n,\lambda} - \Sigma\|_2^2 &\leq 2(\|\overline{\Sigma}_r - \Sigma_r\|_2^2 + \|\overline{\Sigma}_s - \Sigma_s\|_2^2) \\ &\leq C'_w \inf_{M_r, M_s} \left\{ \|M_r - \Sigma_r\|_2^2 + \|M_s - \Sigma_s\|_2^2 + \lambda^2 \text{rank}(M_r) + \lambda^{2-q} |M_s|_q^q \right\}. \end{aligned}$$

The analysis of the error term  $\overline{\mathcal{R}}_n$  can be done as in Section 4 noting that  $\overline{\Theta}(u)$  has very similar properties as  $\Theta(u)$ , especially  $\|\overline{\Theta}(u)\|_\infty \leq \|\Theta(u)\|_\infty + |\Theta(u)|_\infty \leq 2$  and  $\|\overline{\Theta}(u)\|_2 = 2\|\Theta(u)\|_2$ . We omit the details.

#### 5.4. Incorporating positive definiteness constraints

Consider the optimisation problem

$$M_n := \arg \inf_{M \geq 0, M \in \mathbb{R}^{d \times d}} Q_n(M) \quad \text{with} \quad Q_n(M) = \int_{\mathbb{R}^d} \left( \frac{2}{|u|^2} \text{Re} \widehat{\psi}_n(u) - \langle \Theta(u), M \rangle \right)^2 w_U(u) du. \quad (5.5)$$

In general, the solution of (5.5) has to be searched in a space of dimension  $O(d^2)$ . Solving such a problem becomes rapidly intractable for large  $d$ . In order to solve (5.5) at a reduced computational cost, we assume that  $\text{rank}(\Sigma) \leq k \ll d$ . In order to handle the positive definiteness constraint, we let  $M = LL^\top$  with  $L \in \mathbb{R}^{k \times d}$  and rewrite the problem (5.5) as

$$L_n := \arg \inf_{L \in \mathbb{R}^{k \times d}} Q_n(LL^\top). \quad (5.6)$$

In fact any local minima of (5.6) leads to a local minima of (5.5). Since any local minima is a global minima for the convex minimization problem (5.5), any local minima of (5.6) is a global minima.

#### 5.5. A related regression problem

In order to understand the dimension effect that we have observed in the convergence rates in Corollary 4.4, we will take a more abstract point of view considering a related regression-type problem in this section. Motivated from the regression formula (3.2), we study the estimation of a possibly low-rank matrix  $\Sigma \in \mathbb{R}^{d \times d}$ ,  $\Sigma > 0$ , based on the observations

$$X_i(u) = \langle u, \Sigma u \rangle + \Psi(u) + \varepsilon_i(u), \quad i = 1, \dots, n, u \in \mathbb{R}^d, \quad (5.7)$$

where  $\Psi : \mathbb{R}^d \rightarrow \mathbb{R}$  is an unknown deterministic nuisance function satisfying  $|\Psi(u)| \rightarrow 0$  as  $|u| \rightarrow \infty$  and  $\varepsilon_i = (\varepsilon_i(u))_{u \in \mathbb{R}^d}$  are centred, bounded random variables such that:

$$\varepsilon_1, \dots, \varepsilon_n \text{ are i.i.d. with } \mathbb{E}[\varepsilon_i(u)\varepsilon_i(v)] = (|u|^\beta |v|^\beta \vee 1)\varphi(u - v), \quad u, v \in \mathbb{R}^d, i \in \{1, \dots, n\}, \tag{5.8}$$

for some symmetric function  $\varphi \in L^1(\mathbb{R}^d)$  and  $\beta \geq 0$ . The covariance structure of  $\varepsilon_i(u)$  includes on the one hand that the variance of  $\varepsilon_i(u)$  increases (polynomially) as  $|u| \rightarrow \infty$ , implying some ill-posedness, and on the other hand the random variables  $\varepsilon_i(u)$  and  $\varepsilon_i(v)$  decorrelate as the distance of  $u$  and  $v$  increases (supposing that  $\varphi$  is a reasonable regular function).

Following our weighted Lasso approach, we define the statistic

$$T_n(u) := \frac{1}{n} \sum_{i=1}^n X_i(u), \quad u \in \mathbb{R}^d,$$

and introduce the weighted least squares estimator

$$\widehat{\Sigma}_{n,\lambda} := \arg \min_{M \in \mathbb{M}} \int_{\mathbb{R}^d} (T_n(u) - \langle u, Mu \rangle)^2 w_U(u) \, du + \lambda \|M\|_1 \tag{5.9}$$

for a convex subset  $\mathbb{M} \subseteq \{A \in \mathbb{R}^{d \times d} : A > 0\}$ , a weight function  $w_U(u) := w(u/U)$ ,  $u \in \mathbb{R}^d$ , for some radial function  $w : \mathbb{R}^d \rightarrow \mathbb{R}_+$  with support  $\text{supp } w \subseteq \{\frac{1}{2} \leq |u| \leq 1\}$  and a penalisation parameter  $\lambda > 0$ .

Our oracle inequalities easily carry over to the regression setting. We define the weighted scalar product and corresponding (semi-)norm

$$\langle A, B \rangle_U := \int_{\mathbb{R}^d} \langle u, Au \rangle \langle u, Bu \rangle w_U(u) \, du \quad \text{and} \quad \|A\|_U^2 := \langle A, A \rangle_U$$

as well as the error matrix

$$\mathcal{R}_n := \int_{\mathbb{R}^d} (T_n(u) - \langle u, \Sigma u \rangle) u u^\top w_U(u) \, du \in \mathbb{R}^{d \times d}.$$

Along the lines of Proposition 3.3 we obtain

$$\|\widehat{\Sigma}_{n,\lambda} - \Sigma\|_U^2 \leq \inf_{M \in \mathbb{M}} \{ \|M - \Sigma\|_U^2 + 2\lambda \|M\|_1 \} \quad \text{on the event } \{ \|\mathcal{R}_n\|_\infty \leq \lambda \}.$$

It is now crucial to compare the weighted norm  $\|\cdot\|_U$  with the Frobenius norm  $\|\cdot\|_2$ . Analogously to Lemma 3.2, we have for any positive definite matrix  $A = QDQ^\top \in \mathbb{R}^{d \times d}$  with diagonal matrix  $D$  and orthogonal matrix  $Q$  that

$$\begin{aligned} \|A\|_U^2 &= \int_{\mathbb{R}^d} \langle Q^\top u, DQ^\top u \rangle^2 w_U(u) \, du \\ &= U^{d+4} \int_{\mathbb{R}^d} \langle v, Dv \rangle^2 w(v) \, dv \geq U^{d+4} \|A\|_2^2 \underbrace{\int v_1^4 w(v) \, dv}_{=: \varkappa^2}. \end{aligned} \tag{5.10}$$

Hence, compared to the Frobenius norm the weighted norm becomes stronger as  $U$  and  $d$  increase. In the well specified case  $\Sigma \in \mathbb{M}$  we conclude

$$\|\widehat{\Sigma}_{n,\lambda} - \Sigma\|_2 \leq U^{-d/2-2} \varkappa \sqrt{2\lambda \|\Sigma\|_1} \quad \text{on the event } \{ \|\mathcal{R}_n\|_\infty \leq \lambda \}.$$

Theorem 3.4 corresponds to the following stronger inequality which can be proved analogously. Because  $w$  is not normalised in the sense of inequality (5.10), we have to be a bit careful in the very last step in Section 7.3 in order not to oversee a factor  $U^{-d/2-2}$ .



**Theorem 5.6.** *In the regression model (5.7) with  $\Sigma > 0$  the estimator  $\widehat{\Sigma}_{n,\lambda}$  from (5.9) satisfies*

$$\|\widehat{\Sigma}_{n,\lambda} - \Sigma\|_2 \leq CU^{-d-4}\lambda\sqrt{\text{rank } \Sigma} \quad \text{on the event } \{\|\mathcal{R}_n\|_\infty \leq \lambda\} \tag{5.11}$$

for some numerical constant  $C > 0$ .

Let us now estimate the spectral norm of the error matrix

$$\mathcal{R}_n = \frac{1}{n} \sum_{i=1}^n \underbrace{\int_{\mathbb{R}^d} \varepsilon_i(u)uu^\top w_U(u) \, du}_{=:Z_i} + \int_{\mathbb{R}^d} \Psi(u)uu^\top w_U(u) \, du. \tag{5.12}$$

The second term is a deterministic error term which is bounded by

$$\begin{aligned} \left\| \int_{\mathbb{R}^d} \Psi(u)uu^\top w_U(u) \, du \right\|_\infty &\leq \int_{\mathbb{R}^d} |\Psi(u)| |u|^2 w_U(u) \, du \\ &\leq \sup_{|u| \geq U/2} |\Psi(u)| U^{d+2} \int_{\mathbb{R}^d} |v|^2 w(v) \, dv, \end{aligned}$$

using  $\|uu^\top\|_\infty = |u|^2, u \in \mathbb{R}^d$ . The dimension occurs here as a consequence of how the regularity (decay of  $\Psi$ ) is measured in this problem. For the first term in (5.12) we apply, for instance, non-commutative Bernstein inequality noting that  $Z_1, \dots, Z_n$  are i.i.d., bounded and centred. The Cauchy–Schwarz inequality, or more precisely the estimate (8.4), yields

$$\begin{aligned} \|\mathbb{E}[Z_1 Z_1^\top]\|_\infty &= \|\mathbb{E}[Z_1 Z_1^\top]\|_\infty = \left\| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (|u|^\beta |v|^\beta \vee 1) \varphi(u-v)(uu^\top)(vv^\top) w_U(u) w_U(v) \, du \, dv \right\|_\infty \\ &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\varphi(u-v)| |u|^{2+\beta} |v|^{2+\beta} w_U(u) w_U(v) \, du \, dv \\ &\leq \int_{\mathbb{R}^d} |u|^{4+2\beta} w_U^2(u) \int_{\mathbb{R}^d} |\varphi(u-v)| \, dv \, du \\ &= \|\varphi\|_{L^1} U^{4+2\beta+d} \int_{\mathbb{R}^d} |v|^{4+2\beta} w^2(v) \, dv. \end{aligned}$$

Note that the dependence of the variance on  $d$  and  $\beta$  is the usual non-parametric behaviour. Therefore,

$$\|\mathcal{R}_n\|_\infty = \mathcal{O}_P\left(n^{-1/2}U^{2+\beta+d/2} + \sup_{|u| \geq U/2} |\Psi(u)|U^{d+2}\right).$$

We thus choose  $\lambda \sim n^{-1/2}U^{2+\beta+d/2} + \sup_{|u| \geq U/2} |\Psi(u)|U^{d+2}$  and conclude from (5.11)

$$\|\widehat{\Sigma}_{n,\lambda} - \Sigma\|_2 \lesssim \left(n^{-1/2}U^{\beta-2-d/2} + \sup_{|u| \geq U/2} |\Psi(u)|U^{-2}\right)\sqrt{\text{rank } \Sigma}. \tag{5.13}$$

If  $\beta > 2 + d/2$ , then we face a bias-variance trade-off. Supposing  $|\Psi(u)| \lesssim |u|^{-\alpha}$  for some  $\alpha > 0$ , we may choose the optimal  $U^* = n^{1/(2\alpha+2\beta-d)}$  and obtain the rate

$$\|\widehat{\Sigma}_{n,\lambda} - \Sigma\|_2 \lesssim n^{-(\alpha+2)/(2\alpha+2\beta-d)}.$$

This rate coincides with our findings in Corollary 4.4. Note that the second regime in the mildly ill-posed case comes from the auto-deconvolution structure of the Lévy process setting which is not represented in the regression model (5.7).

From this analysis we see that root for the surprising dimension dependence is the factor  $U^d$  in the isometry property (5.10). Intuitively, it reflects that we “observe” more frequencies in the annulus  $\{u \in \mathbb{R}^d : \frac{U}{2} \leq |u| \leq U\}$  as  $d$  increases and, due to the function  $\varphi$  in (5.8), we can profit from these since the observations decorrelate if the frequencies have a large distance from each other.

## 6. Simulations

### 6.1. Known time change

Let us analyse a model based on a time-changed normal inverse Gaussian (NIG) Lévy process. The NIG Lévy processes, characterised by their increments being NIG distributed, have been introduced in [3] as a model for log returns of stock prices. Barndorff-Nielsen [3] considered classes of normal variance-mean mixtures and defined the NIG distribution as the case when the mixing distribution is inverse Gaussian. Shortly after its introduction it was shown that the NIG distribution fits very well the log returns on German stock market data, making the NIG Lévy processes of great interest for practioneers.

An NIG distribution has in general the four parameters  $\alpha \in \mathbb{R}_+$ ,  $\beta \in \mathbb{R}$ ,  $\delta \in \mathbb{R}_+$  and  $\mu \in \mathbb{R}$  with  $|\beta| < \alpha$  having different effects on the shape of the distribution:  $\alpha$  is responsible for the tail heaviness of steepness,  $\beta$  affects symmetry,  $\delta$  scales the distribution and  $\mu$  determines its mean value. The NIG distribution is infinitely divisible with the characteristic function

$$\varphi(u) = \exp\left\{\delta\left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + iu)^2 + i\mu u}\right)\right\}.$$

Therefore one can define the NIG Lévy process  $(X_t)_{t \geq 0}$  which starts at zero and has independent and stationary increments such that each increment  $X_{t+\Delta} - X_t$  is NIG( $\alpha, \beta, \Delta\delta, \Delta\mu$ ) distributed. The NIG process has no diffusion component making it a pure jump process with the Lévy density

$$\nu(x) = \frac{2\alpha\delta \exp(\beta x) K_1(\alpha|x|)}{\pi |x|}, \quad (6.1)$$

where  $K_1(z)$  is the modified Bessel function of the third kind. Taking into account the asymptotic relations

$$K_1(z) \asymp 2/z, \quad z \rightarrow +0 \quad \text{and} \quad K_1(z) \asymp \sqrt{\frac{\pi}{2z}} e^{-z}, \quad z \rightarrow +\infty,$$

we conclude that Assumption B is fulfilled for  $s = -1$  and any  $p > 0$ .

A Gamma process is used for the time change  $\mathcal{T}$  which is a Lévy process such that its increments have a Gamma distribution, so that  $\mathcal{T}$  is a pure-jump increasing Lévy process with Lévy density

$$\nu_{\mathcal{T}}(x) = \vartheta x^{-1} \exp(-\eta x), \quad x \geq 0,$$

where the parameter  $\vartheta$  controls the rate of jump arrivals and the scaling parameter  $\eta$  inversely controls the jump size. The Laplace transform of  $\mathcal{T}$  is of the form

$$\mathcal{L}_{\mathcal{T}}(z) = \mathbb{E}[\exp(-z\mathcal{T}(t))] = (1 + z/\eta)^{-\vartheta t}, \quad \operatorname{Re} z \geq 0.$$

It follows from the properties of the Gamma distribution that Assumption C (with (iv)) is fulfilled for the Gamma process  $\mathcal{T}$  for any  $p > 0$ . Introduce a time-changed Lévy process

$$Y_t = \Sigma^{1/2} W_{\mathcal{T}(t)} + L_{\mathcal{T}(t)}, \quad t \geq 0,$$

where  $W_t$  is a 10-dimensional Wiener process,  $\Sigma$  is a  $10 \times 10$  positive semi-definite matrix,  $L_t = (L_t^1, \dots, L_t^{10})$ ,  $t \geq 0$ , is a 10-dimensional Lévy process with independent NIG components and  $\mathcal{T}$  is a Gamma process. Note that

the process  $Y_t$  is a multi-dimensional Lévy process since  $\mathcal{T}$  was itself a Lévy process (subordinator). Let us be more specific and set

$$\Sigma = O^\top \Lambda O,$$

where  $\Lambda := \text{diag}(1, 0.5, 0.1, 0, \dots, 0) \in \mathbb{R}^{10} \times \mathbb{R}^{10}$  and  $O$  is a randomly generated  $10 \times 10$  orthogonal matrix. We choose  $\Delta = 1$  and  $\text{NIG}(1, -0.1, 1, -0.1)$  distributed increments of the coordinate Lévy processes  $L^j, j = 1, \dots, 10$ . Take also  $\vartheta = 1$  and  $\eta = 1$  for the parameters of the Gamma process  $\mathcal{T}$ .

Simulating a discretised trajectory  $Y_0, Y_1, \dots, Y_n$  of the process  $(Y_t)$  of the length  $n$ , we estimate the covariance matrix  $\Sigma$  with  $\widehat{\Sigma}_{n,\lambda}$  from (3.3). We solve the convex optimisation problem

$$\widehat{\Sigma}_{n,\lambda} := \arg \min_{M \in \{A \in \mathbb{R}^{10 \times 10} : A = A^\top\}} \left\{ \int_{\mathbb{R}^d} (2|u|^{-2} \text{Re} \widehat{\psi}_n(u) + \langle \Theta(u), M \rangle)^2 w_U(u) du + \lambda \|M\|_1 \right\}, \tag{6.2}$$

where  $\widehat{\psi}_n(u) := -\mathcal{L}^{-1}(\varphi_n(u))$  and  $\varphi_n(u) = \frac{1}{n} \sum_{j=1}^n e^{i \langle u, Y_j \rangle}$ . The integral in (6.2) is approximated by a Monte Carlo algorithm using 70 independent draws from the uniform distribution on  $[-U, U]^d$ . In order to ensure that the estimate  $\widehat{\Sigma}_{n,\lambda}$  is a positive semi-definite matrix, we compute the nearest positive definite matrix  $\widetilde{\Sigma}_{n,\lambda}$  which approximates  $\widehat{\Sigma}_{n,\lambda}$ . To find such an approximation we use R package Matrix (function nearPD).

We choose the tuning parameters according to the rule of thumb  $U_n = 0.7 \cdot n^{1/4}$ . A data-driven selection procedure adapting to the unknown quantities  $s$  and  $\eta$  would be interesting, but is behind the scope of this paper. In view of the decomposition (4.2) and our concentration inequalities for  $\varphi_n$ , a model selection procedure or a Lepski method could be developed, cf. [22] or, in the very similar deconvolution setting with unknown error distribution, [16].

Figure 1 shows box plots of the relative estimation error  $\|\widetilde{\Sigma}_{n,\lambda} - \Sigma\|_2 / \|\Sigma\|_2$  for sample sizes  $n \in \{1000, 5000, 10,000, 50,000\}$  based on 100 simulation iterations. As one can see, the nuclear norm penalisation stabilises the estimates especially for smaller sample sizes. Without penalisation the approximated mean squared error is 5 to 10 times larger than the estimation error with penalisation choosing  $\lambda = 10^3/n$ .

Next we look at the ranks of estimated matrices  $\widetilde{\Sigma}_{n,\lambda}$ . The corresponding histograms (obtained over 100 simulation runs each consisting of 10,000 observations) are presented in Figure 2, where the ranks were computed using the function rankMatrix from the package Matrix with tolerance for testing of “practically zero” equal to  $10^{-6}$ . As expected

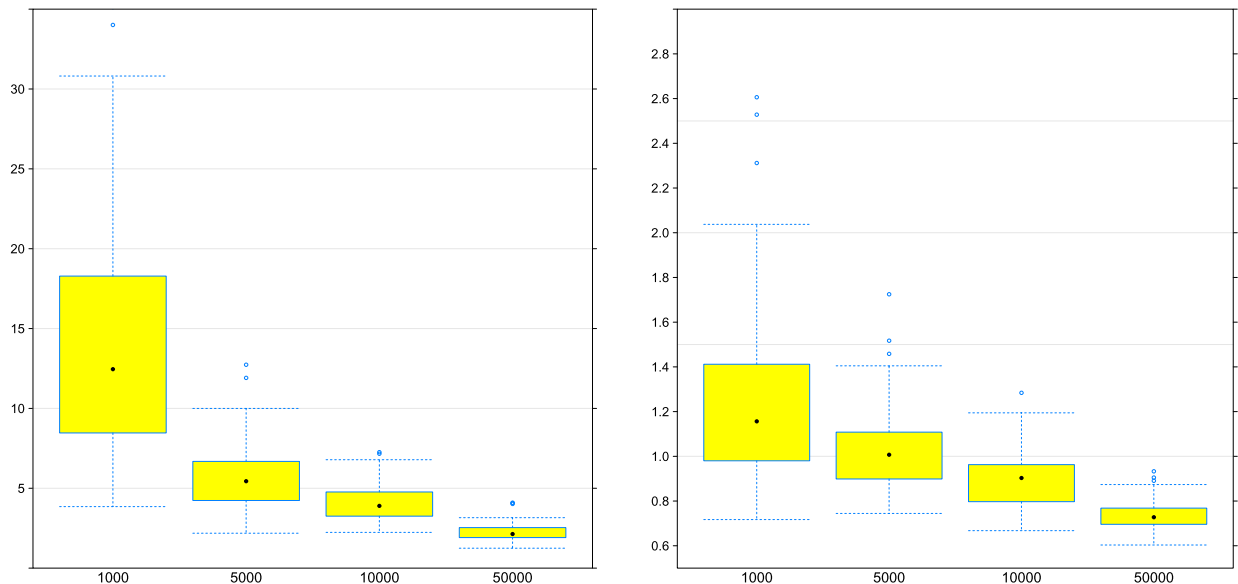


Fig. 1. The relative estimation error  $\|\widetilde{\Sigma}_{n,\lambda} - \Sigma\|_2 / \|\Sigma\|_2$  in dependence on  $n \in \{1000, 5000, 10,000, 50,000\}$  for  $\lambda \equiv 0$  (left) and  $\lambda_n = 10^3/n$  (right).

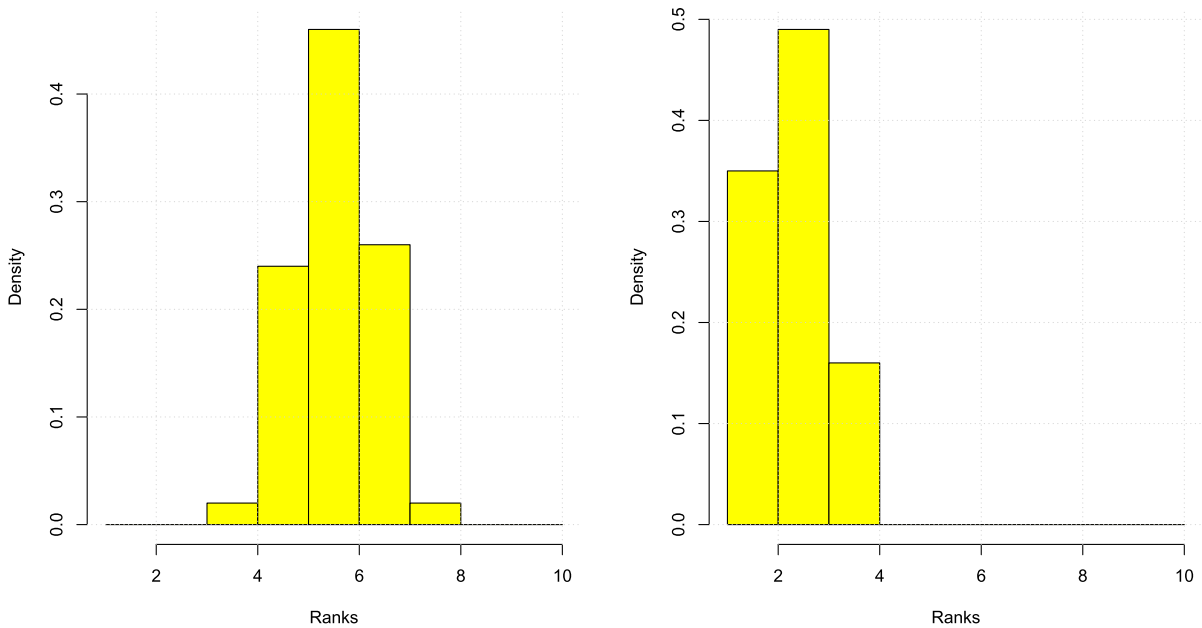


Fig. 2. The histograms (over 100 simulation runs) of ranks of the estimates  $\tilde{\Sigma}_{n,\lambda}$  for  $\lambda \equiv 0$  (left) and  $\lambda_n = 0.2$  (right) in dimension  $d = 10$  (the rank of  $\Sigma$  is 3).

the ranks of the estimated matrices  $\tilde{\Sigma}_{n,\lambda}$  are significantly lower in the case of nuclear norm penalization ( $\lambda > 0$ ) and concentrate around the true rank.

Let us now increase the dimension of the matrix  $\Sigma$  to 100 by adding zeros to eigenvalues, i.e., we set  $\Sigma = O^\top \Lambda O$  with  $\Lambda := \text{diag}(1, 0.5, 0.1, 0, \dots, 0) \in \mathbb{R}^{100} \times \mathbb{R}^{100}$  and randomly generated orthogonal matrix  $O \in \mathbb{R}^{100} \times \mathbb{R}^{100}$ . We take 100-dimensional time-changed Lévy process with independent NIG components and run our estimation algorithm with  $n = 20,000$  and  $\lambda = 5$ . The Figure 3 shows the histograms of ranks based on 100 repetitions of the estimation procedure. The relative estimation error under an optimal choice of  $\lambda$  is of order 0.7.

Let us now illustrate the importance of including the intercept  $a$  in the optimisation problem (3.4) for Lévy processes with  $\alpha = \nu(\mathbb{R}^d) < \infty$ . Consider a model of the type:

$$Y_t = \Sigma^{1/2} W_{\mathcal{T}(t)} + L_{\mathcal{T}(t)}, \quad t \geq 0,$$

where  $(W_t)$  is a 10-dimensional Wiener process,  $\Sigma$  is a  $10 \times 10$  positive semi-definite matrix as before,  $L_t = (L_t^1, \dots, L_t^{10})$ ,  $t \geq 0$ , is a 10-dimensional Lévy process of compound Poisson type and  $\mathcal{T}$  is a Gamma process with parameters as before. In particular, we take each Lévy process  $L_t^k$ ,  $k = 1, \dots, 10$ , to be of the form  $\sum_{i=1}^{N_t} \xi_i$ , where  $(N_t)$  is a Poisson process with intensity 1 and  $\xi_1, \xi_2, \dots$  are i.i.d. standard normals. We compute and compare the estimates  $\hat{\Sigma}$  and  $\tilde{\Sigma}$  for different values of  $\lambda$  ( $n = 10,000$ ). As can be seen from Figure 4, the estimate  $\tilde{\Sigma}$  has much better performance than  $\hat{\Sigma}$ . Moreover, larger values of  $\lambda$  (stronger penalisation) can make the difference between  $\hat{\Sigma}$  and  $\tilde{\Sigma}$  less severe.

### 6.2. Estimated time change

Let us now illustrate the performance of our estimation algorithm in the case of unknown distribution of the time change where the Laplace tranform  $\mathcal{L}$  is estimated according to Section 5.1. In practice we need to cut the expansion of  $\mathcal{L}_m^{-1}$  from (5.1). We thus consider an approximation  $\mathcal{L}_{m,J}^{-1}(z)$  of the form

$$\mathcal{L}_{m,J}^{-1}(z) := \sum_{j=1}^J H_j \frac{(z-1)^j}{j!} \tag{6.3}$$

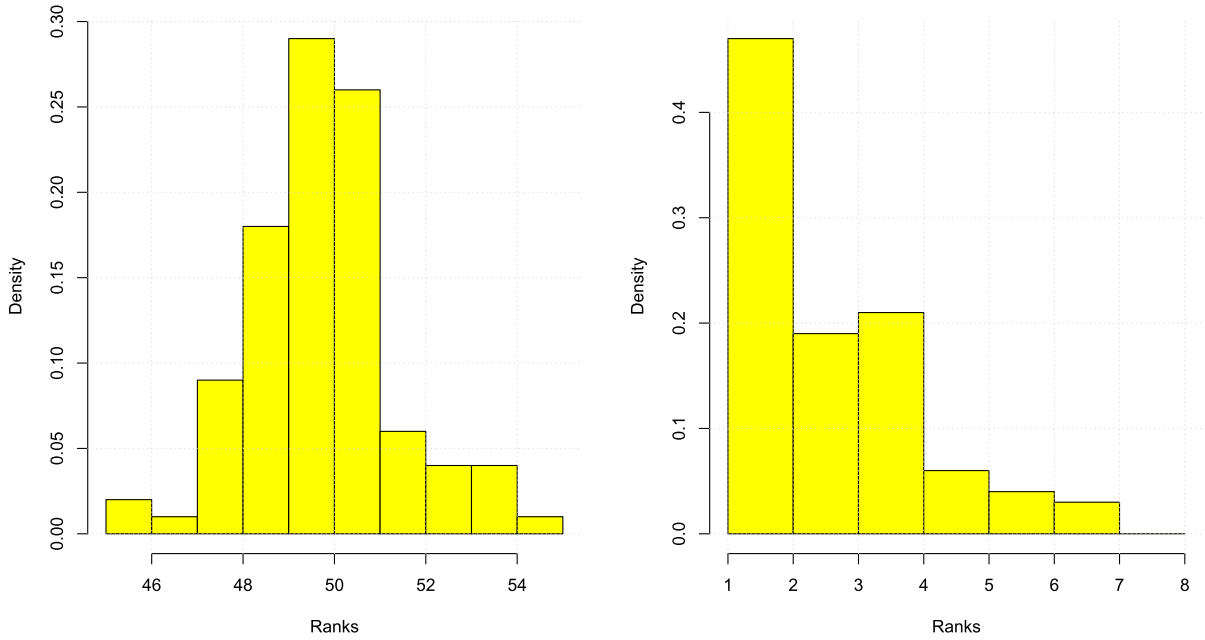


Fig. 3. The histograms (over 100 simulation runs) of ranks of the estimates  $\tilde{\Sigma}_{n,\lambda}$  for  $\lambda \equiv 0$  (left) and  $\lambda_n = 10$  (right) in dimension  $d = 100$  (the rank of  $\Sigma$  is 3).

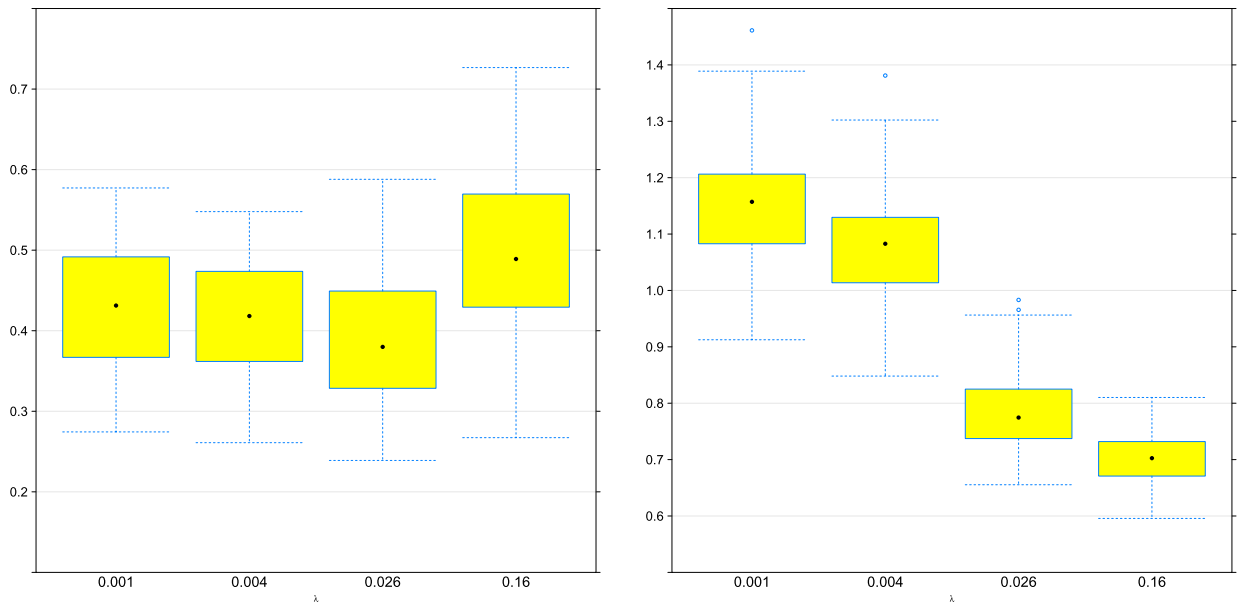


Fig. 4. The relative estimation errors  $\|\tilde{\Sigma}_{n,\lambda} - \Sigma\|_2 / \|\Sigma\|_2$  (left) and  $\|\hat{\Sigma}_{n,\lambda} - \Sigma\|_2 / \|\Sigma\|_2$  (right) in dependence on  $\lambda$  in the case  $\nu(\mathbb{R}^d) < \infty$ .

for some integer  $J > 1$ . Figure 5 (left) shows a typical realisation of the difference  $\mathcal{L}_{m,J}^{-1}(z) - \mathcal{L}^{-1}(z)$ ,  $z = x + iy$ , for  $m = 100$  and  $J = 20$  in the case of a Gamma subordinator with parameters as before. On the right hand side of Figure 5, one can see box plots of the error  $\|\hat{\Sigma}_{n,\lambda} - \Sigma\|_2 / \|\Sigma\|_2$  when using  $\mathcal{L}_{m,J}^{-1}(z)$  instead of  $\mathcal{L}^{-1}(z)$  for different values of  $m$  and  $J = 20$ , other parameters being the same as in the 10-dimensional example of Section 6.1. As can be seen, already relatively small values of  $m$  lead to reasonable estimates of  $\Sigma$ . Moreover by comparing Figure 5 with

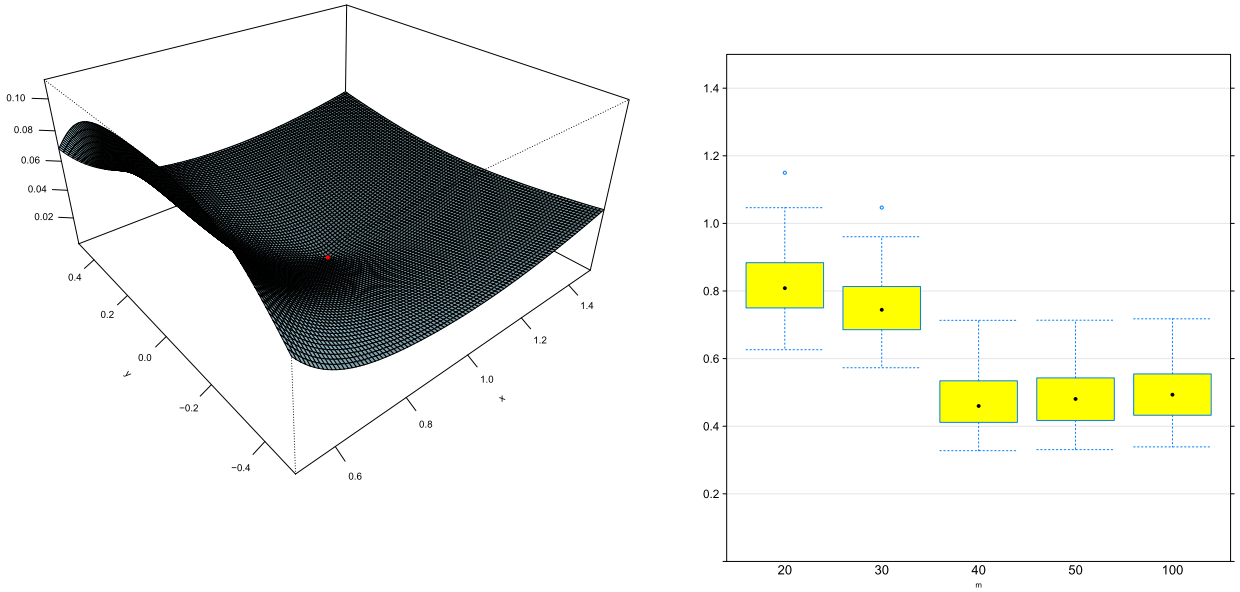


Fig. 5. The difference between  $\mathcal{L}_{m,J}^{-1}(z)$  and  $\mathcal{L}^{-1}(z)$  on the complex plane  $z = x + iy$  in the case of Gamma subordinator (right) and box plots of the error  $\|\widehat{\Sigma}_{n,\lambda} - \Sigma\|_2 / \|\Sigma\|_2$  when using  $\mathcal{L}_{m,J}^{-1}(z)$  instead of  $\mathcal{L}^{-1}(z)$  for different values of  $m$  and  $J = 20$ .

Figure 1, one can even observe a slight improvement of the error rate for  $\mathcal{L}_{m,J}^{-1}(z)$ . This can be due to a regularising effect of the series truncation in (6.3).

**7. Proofs of the oracle inequalities**

*7.1. Proof of the isometry property: Lemma 3.2*

Let  $A$  has rank  $k$  and admit the singular value decomposition  $A = QDQ^T$  for the diagonal matrix  $D = \text{diag}(\lambda_1, \dots, \lambda_k, 0, \dots, 0)$  and an orthogonal matrix  $Q$ . Noting that  $\langle \Theta(u), A \rangle \geq 0$  for all  $u \in \mathbb{R}^d$ , we have

$$\begin{aligned} \|(A, a)\|_w^2 &= \int_{\mathbb{R}^d} (\langle \Theta(u), A \rangle + 2U^2|u|^{-2}a)^2 w_U(u) \, du \\ &\geq \int_{\mathbb{R}^d} (\langle Q^T u, DQ^T u \rangle^2 + 4U^4 a^2) \frac{w_U(u)}{|u|^4} \, du \\ &= \int_{\mathbb{R}^d} ((\lambda_1 v_1^2 + \dots + \lambda_k v_k^2)^2 + 4a^2) \frac{w(v)}{|v|^4} \, dv \\ &\geq \int_{\mathbb{R}^d} (\lambda_1^2 v_1^4 + \dots + \lambda_k^2 v_k^4 + 4a^2) \frac{w(v)}{|v|^4} \, dv \\ &= \varkappa_1 \|A\|_2^2 + \varkappa_2 a^2, \end{aligned}$$

with  $\varkappa_1 := \int_{\mathbb{R}^d} (v_1^4/|v|^4) w(v) \, dv$  and  $\varkappa_2 := 4 \int_{\mathbb{R}^d} |v|^{-4} w(v) \, dv \geq \varkappa_1$ . On the other hand the Cauchy–Schwarz inequality yields

$$\begin{aligned} \|(A, a)\|_w^2 &\leq 2 \int_{\mathbb{R}^d} (\langle \Theta(u), A \rangle^2 + 4U^4|u|^{-4}a^2) w_U(u) \, du \\ &\leq 2\|A\|_2^2 \int_{\mathbb{R}^d} \|\Theta(u)\|_2^2 w_U(u) \, du + 8a^2 \int_{\mathbb{R}^d} |v|^{-4} w(v) \, dv = 2\|w\|_{L^1} \|A\|_2^2 + 2\varkappa_2 a^2, \end{aligned}$$

where the last equality follow from  $\|\Theta(u)\|_2 = 1$  for any  $u \neq 1$  since  $\Theta(u) = |u|^{-2}uu^\top$  has the spectrum  $\{1, 0, \dots, 0\}$ .  $\square$

7.2. Proof of the first oracle inequality: Proposition 3.3

For convenience define  $\mathcal{Y}_n := 2|u|^{-2} \operatorname{Re} \widehat{\psi}_n(u)$ . It follows from the definition of  $(\widehat{\Sigma}_{n,\lambda}, \widehat{\alpha}_{n,\lambda})$  that for all  $M \in \mathbb{M}$  and  $a \geq 0$

$$\begin{aligned} & \left\| \left( \widehat{\Sigma}_{n,\lambda}, \frac{\widehat{\alpha}_{n,\lambda}}{U^2} \right) \right\|_w^2 - 2 \int_{\mathbb{R}^d} \mathcal{Y}_n(u) \left\langle \widetilde{\Theta}(u), \operatorname{diag} \left( \widehat{\Sigma}_{n,\lambda}, \frac{\widehat{\alpha}_{n,\lambda}}{U^2} \right) \right\rangle w_U(u) \, du + \lambda \left( \|\widehat{\Sigma}_{n,\lambda}\|_1 + \frac{\widehat{\alpha}_{n,\lambda}}{U^2} \right) \\ & \leq \left\| \left( M, \frac{a}{U^2} \right) \right\|_w^2 - 2 \int_{\mathbb{R}^d} \mathcal{Y}_n(u) \left\langle \widetilde{\Theta}(u), \operatorname{diag} \left( M, \frac{a}{U^2} \right) \right\rangle w_U(u) \, du + \lambda \left( \|M\|_1 + \frac{a}{U^2} \right). \end{aligned}$$

Hence,

$$\begin{aligned} & \left\| \left( \widehat{\Sigma}_{n,\lambda}, \frac{\widehat{\alpha}_{n,\lambda}}{U^2} \right) \right\|_w^2 - 2 \left\langle \left( \Sigma, \frac{\alpha}{U^2} \right), \left( \widehat{\Sigma}_{n,\lambda}, \frac{\widehat{\alpha}_{n,\lambda}}{U^2} \right) \right\rangle_w + \lambda \left( \|\widehat{\Sigma}_{n,\lambda}\|_1 + \frac{\widehat{\alpha}_{n,\lambda}}{U^2} \right) \\ & \leq \left\| \left( M, \frac{a}{U^2} \right) \right\|_w^2 - 2 \left\langle \left( \Sigma, \frac{\alpha}{U^2} \right), \left( M, \frac{a}{U^2} \right) \right\rangle_w + \lambda \left( \|M\|_1 + \frac{a}{U^2} \right) \\ & \quad + \left\langle \mathcal{R}_n, \operatorname{diag} \left( \widehat{\Sigma}_{n,\lambda} - M, \frac{\widehat{\alpha}_{n,\lambda} - a}{U^2} \right) \right\rangle. \end{aligned}$$

Due to the trace duality we have on the good event that  $|\langle \mathcal{R}_n, \operatorname{diag}(\widehat{\Sigma}_{n,\lambda} - M, U^{-2}(\widehat{\alpha}_{n,\lambda} - a)) \rangle| \leq \lambda(\|\widehat{\Sigma}_{n,\lambda} - M\|_1 + U^{-2}|\widehat{\alpha}_{n,\lambda} - a|)$  and as a result

$$\begin{aligned} & \left\| \left( \widehat{\Sigma}_{n,\lambda} - \Sigma, U^{-2}(\widehat{\alpha}_{n,\lambda} - \alpha) \right) \right\|_w^2 \\ & \leq \left\| \left( M - \Sigma, U^{-2}(a - \alpha) \right) \right\|_w^2 + \lambda(\|\widehat{\Sigma}_{n,\lambda} - M\|_1 + \|M\|_1 - \|\widehat{\Sigma}_{n,\lambda}\|_1 + U^{-2}(|\widehat{\alpha}_{n,\lambda} - a| + a - \widehat{\alpha}_{n,\lambda})) \\ & \leq \left\| \left( M - \Sigma, U^{-2}(a - \alpha) \right) \right\|_w^2 + 2\lambda(\|M\|_1 + U^{-2}a). \end{aligned}$$

Consequently, choosing  $a = \alpha$  yields

$$\left\| \left( \widehat{\Sigma}_{n,\lambda} - \Sigma, U^{-2}(\widehat{\alpha}_{n,\lambda} - \alpha) \right) \right\|_w^2 \leq \inf_{M \in \mathbb{M}} \{ \|M - \Sigma\|_w^2 + 2\lambda(\|M\|_1 + U^{-2}\alpha) \}. \quad \square$$

7.3. Proof of the second oracle inequality: Theorem 3.4

We first introduce some abbreviations. Define  $\overline{\mathbb{M}} := \{\operatorname{diag}(M, a) : M \in \mathbb{M}, a \in I\}$  with elements  $\overline{M} = \operatorname{diag}(M, a) \in \overline{\mathbb{M}}$  and the convention  $\|\overline{M}\|_w := \|(M, a)\|_w$ . Note that convexity of  $\mathbb{M}$  and  $I$  implies convexity of  $\overline{\mathbb{M}}$ . Moreover, we write  $\overline{\Sigma}_{n,\lambda} := \operatorname{diag}(\widehat{\Sigma}_{n,\lambda}, U^{-2}\widehat{\alpha}_{n,\lambda})$  and  $\overline{\Sigma} = \operatorname{diag}(\Sigma, U^{-2}\alpha)$ . As above we use  $\mathcal{Y}_n := 2|u|^{-2} \operatorname{Re} \widehat{\psi}_n(u)$ .

The proof borrows some ideas and notation from the proof of Theorem 1 in [24]. First note that a necessary condition of extremum in (3.4) implies that there is  $V_n \in \partial \|\overline{\Sigma}_{n,\lambda}\|_1$ , i.e.  $V_n$  is an element of the subgradient of the nuclear norm, such that the matrix

$$A = 2 \int_{\mathbb{R}^d} \widetilde{\Theta}(u) [\mathcal{Y}_n(u) - \langle \widetilde{\Theta}(u), \overline{\Sigma}_{n,\lambda} \rangle] w_U(u) \, du - \lambda V_n$$

belongs to the normal cone at the point  $\overline{\Sigma}_{n,\lambda}$  i.e.  $\langle A, \overline{\Sigma}_{n,\lambda} - \overline{M} \rangle \geq 0$  for all  $\overline{M} \in \overline{\mathbb{M}}$ . Hence

$$2 \int_{\mathbb{R}^d} \langle \widetilde{\Theta}(u), \overline{\Sigma}_{n,\lambda} - \overline{M} \rangle [\mathcal{Y}_n(u) - \langle \widetilde{\Theta}(u), \overline{\Sigma}_{n,\lambda} \rangle] w_U(u) \, du - \lambda \langle V_n, \overline{\Sigma}_{n,\lambda} - \overline{M} \rangle \geq 0$$

or equivalently

$$2 \int_{\mathbb{R}^d} \langle \tilde{\Theta}(u), \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle \mathcal{Y}_n(u) w_U(u) \, du - 2 \langle \bar{\Sigma}_{n,\lambda}, \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle_w - \lambda \langle V_n, \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle \geq 0.$$

Furthermore

$$\begin{aligned} & 2 \int_{\mathbb{R}^d} \langle \tilde{\Theta}(u), \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle [\mathcal{Y}_n(u) - \langle \tilde{\Theta}(u), \bar{\Sigma} \rangle] w_U(u) \, du - 2 \langle \bar{\Sigma}_{n,\lambda} - \bar{\Sigma}, \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle_w - \lambda \langle V_n - V, \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle \\ & \geq \lambda \langle V, \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle \end{aligned}$$

for any  $V \in \partial \|\bar{M}\|_1$ . Fix some  $\bar{M} \in \bar{\mathbb{M}}$  of rank  $r$  with the spectral representation

$$\bar{M} = \sum_{j=1}^r \sigma_j u_j v_j^\top,$$

where  $\sigma_1 \geq \dots \geq \sigma_r > 0$  are singular values of  $\bar{M}$ . Due to the representation for the subdifferential of the mapping  $\bar{M} \rightarrow \|\bar{M}\|_1$  (see [43])

$$\partial \|\bar{M}\|_1 = \left\{ \sum_{j=1}^r u_j v_j^\top + \Pi_{S_1^\top} \Lambda \Pi_{S_2^\top} : \|\Lambda\|_\infty \leq 1 \right\},$$

where  $(S_1, S_2) = (\text{span}(u_1, \dots, u_r), \text{span}(v_1, \dots, v_r))$  is the support of  $\bar{M}$ ,  $\Pi_S$  is a projector on the linear vector subspace  $S$ , we get

$$V = \sum_{j=1}^r u_j v_j^\top + \Pi_{S_1^\top} \Lambda \Pi_{S_2^\top}, \quad \text{for some } \Lambda \text{ with } \|\Lambda\|_\infty \leq 1.$$

By the trace duality, there is  $\Lambda$  such that

$$\langle \Pi_{S_1^\top} \Lambda \Pi_{S_2^\top}, \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle = \langle \Pi_{S_1^\top} \Lambda \Pi_{S_2^\top}, \bar{\Sigma}_{n,\lambda} \rangle = \langle \Lambda, \Pi_{S_1^\top} \bar{\Sigma}_{n,\lambda} \Pi_{S_2^\top} \rangle = \|\Pi_{S_1^\top} \bar{\Sigma}_{n,\lambda} \Pi_{S_2^\top}\|_1$$

and for this  $\Lambda$

$$\begin{aligned} & 2 \int_{\mathbb{R}^d} \langle \tilde{\Theta}(u), \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle [\mathcal{Y}_n(u) - \langle \tilde{\Theta}(u), \bar{\Sigma} \rangle] w_U(u) \, du - 2 \langle \bar{\Sigma}_{n,\lambda} - \bar{\Sigma}, \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle_w \\ & \geq \lambda \left\langle \sum_{j=1}^r u_j v_j^\top, \bar{\Sigma}_{n,\lambda} - \bar{M} \right\rangle + \lambda \|\Pi_{S_1^\top} \bar{\Sigma}_{n,\lambda} \Pi_{S_2^\top}\|_1 \end{aligned}$$

since  $\langle V_n - V, \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle = \langle V_n, \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle + \langle V, \bar{M} - \bar{\Sigma}_{n,\lambda} \rangle \geq 0$  ( $V \in \partial \|\bar{M}\|_1$  and  $V_n \in \partial \|\bar{\Sigma}_{n,\lambda}\|_1$ ). Using the identities

$$\left\| \sum_{j=1}^r u_j v_j^\top \right\|_\infty = 1, \quad \left\langle \sum_{j=1}^r u_j v_j^\top, \bar{\Sigma}_{n,\lambda} - \bar{M} \right\rangle = \left\langle \sum_{j=1}^r u_j v_j^\top, \Pi_{S_1} (\bar{\Sigma}_{n,\lambda} - \bar{M}) \Pi_{S_2} \right\rangle,$$

we deduce

$$\begin{aligned} & \|\bar{\Sigma}_{n,\lambda} - \bar{\Sigma}\|_w^2 + \|\bar{\Sigma}_{n,\lambda} - \bar{M}\|_w^2 + \lambda \|\Pi_{S_1^\top} \bar{\Sigma}_{n,\lambda} \Pi_{S_2^\top}\|_1 \\ & \leq \|\bar{M} - \bar{\Sigma}\|_w^2 + \lambda \|\Pi_{S_1} (\bar{\Sigma}_{n,\lambda} - \bar{M}) \Pi_{S_2}\|_1 - \langle \mathcal{R}_n, \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle. \end{aligned}$$



On the good event the trace duality yields

$$\begin{aligned} |\langle \mathcal{R}_n, \bar{\Sigma}_{n,\lambda} - \bar{M} \rangle| &\leq |\langle \mathcal{R}_n, \pi_{\bar{M}}(\bar{\Sigma}_{n,\lambda} - \bar{M}) \rangle| + |\langle \mathcal{R}_n, \Pi_{S_1^\top} \bar{\Sigma}_{n,\lambda} \Pi_{S_2^\top} \rangle| \\ &\leq \lambda \|\pi_{\bar{M}}(\bar{\Sigma}_{n,\lambda} - \bar{M})\|_1 + \lambda \|\Pi_{S_1^\top} \bar{\Sigma}_{n,\lambda} \Pi_{S_2^\top}\|_1 \end{aligned}$$

with  $\pi_{\bar{M}}(A) = A - \Pi_{S_1^\top} A \Pi_{S_2^\top}$ . Next by the Cauchy–Schwarz inequality,

$$\|\Pi_{S_1}(\bar{\Sigma}_{n,\lambda} - \bar{M})\Pi_{S_2}\|_1 \leq \sqrt{\text{rank}(\bar{M})} \|\Pi_{S_1}(\bar{\Sigma}_{n,\lambda} - \bar{M})\Pi_{S_2}\|_2 \leq \sqrt{\text{rank}(\bar{M})} \|\bar{\Sigma}_{n,\lambda} - \bar{M}\|_2,$$

as well as

$$\begin{aligned} \|\pi_{\bar{M}}(\bar{\Sigma}_{n,\lambda} - \bar{M})\|_1 &\leq \sqrt{\text{rank}(\pi_{\bar{M}}(\bar{\Sigma}_{n,\lambda} - \bar{M}))} \|\pi_{\bar{M}}(\bar{\Sigma}_{n,\lambda} - \bar{M})\|_2 \\ &\leq \sqrt{2 \text{rank}(\bar{M})} \|\bar{\Sigma}_{n,\lambda} - \bar{M}\|_2, \end{aligned}$$

since  $\pi_{\bar{M}}(A) = \Pi_{S_1^\top} A \Pi_{S_2} + \Pi_{S_1} A$ . Combining the above inequalities, we get

$$\begin{aligned} \|\bar{\Sigma}_{n,\lambda} - \bar{\Sigma}\|_w^2 + \|\bar{\Sigma}_{n,\lambda} - \bar{M}\|_w^2 + \lambda \|\Pi_{S_1^\top} \bar{\Sigma}_{n,\lambda} \Pi_{S_2^\top}\|_1 \\ \leq \|\bar{M} - \bar{\Sigma}\|_w^2 + \lambda \sqrt{\text{rank}(\bar{M})} \|\bar{\Sigma}_{n,\lambda} - \bar{M}\|_2 + \lambda \sqrt{2 \text{rank}(\bar{M})} \|\bar{\Sigma}_{n,\lambda} - \bar{M}\|_2 + \lambda \|\Pi_{S_1^\top} \bar{\Sigma}_{n,\lambda} \Pi_{S_2^\top}\|_1 \end{aligned}$$

which is equivalent to

$$\|\bar{\Sigma}_{n,\lambda} - \bar{\Sigma}\|_w^2 + \|\bar{\Sigma}_{n,\lambda} - \bar{M}\|_w^2 \leq \|\bar{M} - \bar{\Sigma}\|_w^2 + (1 + \sqrt{2})\lambda \sqrt{\text{rank}(\bar{M})} \|\bar{\Sigma}_{n,\lambda} - \bar{M}\|_2.$$

Finally, we choose  $\bar{M} = \text{diag}(M, U^{-2}\alpha)$  and apply Lemma 3.2 as well as the standard estimate  $ab \leq (a/2)^2 + b^2$  for any  $a, b \in \mathbb{R}$  to obtain on the good event

$$\|\bar{\Sigma}_{n,\lambda} - \bar{\Sigma}\|_w^2 \leq \|M - \Sigma\|_w^2 + \left(\frac{1 + \sqrt{2}}{2}\right)^2 \underline{\kappa}_w^{-1} \lambda^2 (\text{rank}(M) + \mathbb{1}_{\alpha \neq 0}). \quad \square$$

## 8. Proof of the convergence rates

### 8.1. Proof of the upper bound: Theorem 4.2

We start with an auxiliary lemma:

**Lemma 8.1.** *Let  $L = \{L_t : t \geq 0\}$  be a  $d$ -dimensional Lévy process with characteristic triplet  $(A, c, \nu)$ . If  $\int_{|x| \geq 1} |x|^p \nu(dx)$  for  $p \geq 1$ , then  $\mathbb{E}[|L_t|^p] = \mathcal{O}(t^{p/n} \vee t^p)$  for any  $t > 0$  where  $n \in \mathbb{N}$  is the smallest even natural number satisfying  $n \geq p$ .*

**Proof.** We decompose the Lévy process  $L = M + N$  into two independent Lévy processes  $M$  and  $N$  with characteristics  $(A, c, \nu \mathbb{1}_{\{|x| \leq 1\}})$  and  $(0, 0, \nu \mathbb{1}_{\{|x| > 1\}})$ , respectively. The triangle inequality yields

$$\mathbb{E}[|L_t|^p] \lesssim \mathbb{E}[|M_t|^p] + \mathbb{E}[|N_t|^p].$$

Let  $n \geq p$  be an even integer. Due to the compactly supported jump measure of  $M$ , the  $n$ th moment of  $M_t$  is finite and can be written as a polynomial of degree  $n$  of its first  $n$  cumulants. Since all cumulants of  $M_t$  are linear in  $t$ , we conclude with Jensen's inequality

$$\mathbb{E}[|M_t|^p] \leq \mathbb{E}[M_t^n]^{p/n} \lesssim (t + t^n)^{p/n} \lesssim t^{p/n} \vee t^p.$$

Since  $N$  is a compound Poisson process, it can be represented as  $N_t = \sum_{k=1}^{P_t} Z_k$  for a Poisson process  $P_t$  with intensity  $\lambda = \nu(\{|x| > 1\}) < \infty$  and a sequence of i.i.d. random variables  $(Z_k)_{k \geq 1}$  which is independent of  $P$ . Hence,

$$\mathbb{E}[|N_t|^p] \leq \mathbb{E}\left[\sum_{k=1}^{P_t} |Z_k|^p\right] \leq \mathbb{E}[P_t] \mathbb{E}[|Z_1|^p] = t\lambda \mathbb{E}[|Z_1|^p].$$

Note that  $\mathbb{E}[|Z_1|^p]$  is finite owing to the assumption  $\int_{|x| \geq 1} |x|^p \nu(dx) < \infty$ . □

**Remark 8.2.** While the upper bound  $t^p$  is natural, the order  $t^{p/n}$  is sharp too in the sense that for a Brownian motion  $L_t = W_t$  and  $p = 1$  we have  $\mathbb{E}[|W_t|] = t^{1/2} \mathbb{E}[|W_1|]$ .

To show the upper bound, we start with bounding the approximation error term in the decomposition (4.2).

**Lemma 8.3.** *If the jump measure  $\nu$  of  $Z$  satisfies Assumption B(i) for some  $s \in (-2, \infty)$ ,  $C_\nu > 0$ , then*

$$\int_{\mathbb{R}^d} \frac{|\operatorname{Re} \Psi(u) + \alpha|}{|u|^2} w_U(u) \, du \leq C_s |U_n|^{-s-2}, \tag{8.1}$$

where  $C_s := 2C_\nu \int_{\{1/4 \leq |v| \leq 1/2\}} |v|^{-s-2} w(v) \, dv$ .

**Proof.** Let us start with the case  $s \in (-2, 0)$  and  $\alpha = 0$ . For all  $u \in \mathbb{R}^d \setminus \{0\}$  we have

$$\begin{aligned} |\Psi(u)| &\leq \int_{\mathbb{R}^d} |e^{i\langle x, u \rangle} - 1 - i\langle x, u \rangle \mathbb{1}_{\{|x| \leq 1\}}(x)| \nu(dx) \leq \int_{\mathbb{R}^d} \left(\frac{\langle x, u \rangle^2}{2} \wedge 2\right) \nu(dx) \\ &\leq 2^{1+s} \int_{\mathbb{R}^d} |\langle x, u \rangle|^{s+1} \nu(dx) \leq 2^{1+s} C_\nu |u|^{-s}. \end{aligned}$$

Hence, we obtain

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{|\operatorname{Re} \Psi(u)|}{|u|^2} w_U(u) \, du &\leq 2^{1+s} C_\nu \int_{\mathbb{R}^d} |u|^{-s-2} w_U(u) \, du \\ &\leq U^{-s-2} 2^{1+s} C_\nu \int_{\{1/4 \leq |v| \leq 1/2\}} |v|^{-s-2} w(v) \, dv. \end{aligned}$$

In the case  $s \geq 0$  and  $\alpha = \nu(\mathbb{R}^d) < \infty$ , we have  $\operatorname{Re} \Psi(u) + \alpha = \operatorname{Re}(\mathcal{F}\nu)(u)$ ,  $u \in \mathbb{R}^d$ , such that

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{|\operatorname{Re} \Psi(u) + \alpha|}{|u|^2} w_U(u) \, du &\leq C_\nu \int_{\mathbb{R}^d} |u|^{-2} (1 + |u|^2)^{-s/2} w_U(u) \, du \\ &\leq U^{-s-2} C_\nu \int_{\{1/4 \leq |u| \leq 1/2\}} |v|^{-s-2} w_U(u) \, du. \end{aligned} \tag{□}$$

In order to bound the stochastic error term in (4.2), we apply the following linearisation lemma. We denote throughout  $\|f\|_U := \sup_{|u| \leq U} |f(u)|$ .

**Lemma 8.4.** *Grant Assumption C(iii) with  $C_L > 0$ . For all  $n \in \mathbb{N}$  and  $U > 0$  we have on the event*

$$\mathcal{H}_{n,U} := \left\{ \|\varphi_n - \varphi\|_U \leq \frac{2}{C_L} \inf_{|u| \leq U} \mathcal{L}'(-\psi(u)) \right\}$$

that for all  $u \in \mathbb{R}^d$  with  $|u| \leq U$  it holds

$$\left| \mathcal{L}^{-1}(\varphi_n(u)) - \mathcal{L}^{-1}(\varphi(u)) - \frac{\varphi_n(u) - \varphi(u)}{\mathcal{L}'(-\psi(u))} \right| \leq 4C_L \frac{|\varphi_n(u) - \varphi(u)|^2}{|\mathcal{L}'(-\psi(u))|^2 |\psi(u)|^q},$$

where  $q = 1$  if Assumption C(iv) is satisfied and  $q = 0$  otherwise.

**Proof.** First note that

$$(\mathcal{L}^{-1})'(z) = \frac{1}{\mathcal{L}'(\mathcal{L}^{-1}(z))} \quad \text{and} \quad (\mathcal{L}^{-1})''(z) = -\frac{\mathcal{L}''(\mathcal{L}^{-1}(z))}{\mathcal{L}'(\mathcal{L}^{-1}(z))^3}$$

and in particular  $(\mathcal{L}^{-1})'(\varphi(u)) = 1/\mathcal{L}'(-\psi(u))$ . Hence, the Taylor formula yields

$$\mathcal{L}^{-1}(\varphi_n(u)) - \mathcal{L}^{-1}(\varphi(u)) = \frac{\varphi_n(u) - \varphi(u)}{\mathcal{L}'(-\psi(u))} + R(u)$$

with

$$\begin{aligned} |R(u)| &\leq |\varphi_n(u) - \varphi(u)|^2 |(\mathcal{L}^{-1})''(\varphi(u) + \xi_1(\varphi_n(u) - \varphi(u)))| \\ &\leq \frac{C_L |\varphi_n(u) - \varphi(u)|^2}{|\mathcal{L}'(\mathcal{L}^{-1}(\varphi(u) + \xi_1(\varphi_n(u) - \varphi(u))))|^2} \end{aligned} \tag{8.2}$$

for some intermediate point  $\xi_1 \in [0, 1]$  depending on  $u$ . For another intermediate point  $\xi_2 \in [0, 1]$  we estimate

$$\begin{aligned} &|\mathcal{L}'(\mathcal{L}^{-1}(\varphi(u) + \xi_1(\varphi_n(u) - \varphi(u))) - \mathcal{L}'(-\psi(u))| \\ &\leq |\varphi_n(u) - \varphi(u)| |(\mathcal{L}' \circ \mathcal{L}^{-1})'(\varphi(u) + \xi_2(\varphi_n(u) - \varphi(u)))| \\ &= |\varphi_n(u) - \varphi(u)| \left| \left( \frac{\mathcal{L}'' \circ \mathcal{L}^{-1}}{\mathcal{L}' \circ \mathcal{L}^{-1}} \right) (\varphi(u) + \xi_2(\varphi_n(u) - \varphi(u))) \right| \\ &\leq C_L |\varphi_n(u) - \varphi(u)|. \end{aligned} \tag{8.3}$$

Therefore, we have on the event  $\mathcal{H}_n$  for any  $u$  in the support of  $w_U$

$$|R(u)| \leq C_L |\varphi_n(u) - \varphi(u)|^2 (|\mathcal{L}'(-\psi(u))| - C_L |\varphi_n(u) - \varphi(u)|)^{-2} \leq 4C_L \frac{|\varphi_n(u) - \varphi(u)|^2}{|\mathcal{L}'(-\psi(u))|^2}.$$

Under Assumption C(iv) we can obtain a sharper estimate. More precisely, (8.2) and (8.3) imply together with the faster decay that

$$|R(u)| \leq \frac{4C_L |\varphi_n(u) - \varphi(u)|^2}{|\mathcal{L}'(-\psi(u))|^2 (1 + |\mathcal{L}^{-1}(\varphi(u) + \xi_1(\varphi_n(u) - \varphi(u)))|)}.$$

Using  $\varphi = \mathcal{L}(-\psi)$  and again (8.3), we have on  $\mathcal{H}_n$

$$\begin{aligned} &|\psi(u) + \mathcal{L}^{-1}(\varphi(u) + \xi_1(\varphi_n(u) - \varphi(u)))| \\ &\leq |\varphi_n(u) - \varphi(u)| |(\mathcal{L}^{-1})'(\varphi(u) + \xi_2(\varphi_n(u) - \varphi(u)))| \\ &\leq \frac{|\varphi_n(u) - \varphi(u)|}{|\mathcal{L}'(\mathcal{L}^{-1}(\varphi(u) + \xi_2(\varphi_n(u) - \varphi(u))))|} \leq \frac{2|\varphi_n(u) - \varphi(u)|}{|\mathcal{L}'(-\psi(u))|} \leq \frac{4}{C_L}. \end{aligned}$$

We conclude

$$|R(u)| \leq \frac{4C_L |\varphi_n(u) - \varphi(u)|^2}{|\mathcal{L}'(-\psi(u))|^2 |\psi(u)|}. \quad \square$$

Denoting the linearised stochastic error term by

$$L_n := \int_{\mathbb{R}^d} |u|^{-2} \operatorname{Re} \left( \frac{\varphi_n(u) - \varphi(u)}{\mathcal{L}'(-\psi(u))} \right) \Theta(u) w_U(u) \, du \in \mathbb{R}^{d \times d},$$

we obtain the following concentration inequality

**Lemma 8.5.** *Define  $\xi_U := U^2 \inf_{|u| \leq U} \mathcal{L}'(-\psi(u))$ . There is some  $c > 0$  depending only on  $w$  such that for any  $n \geq 1$  and any  $\kappa \in (0, (nU^d / \|\varphi\|_{L^1(B_U^d)})^{1/2})$*

$$\mathbb{P}\left(\|L_n\|_\infty \geq \frac{\kappa}{\sqrt{n}\xi_U} \sqrt{U^{-d} \|\varphi\|_{L^1(B_U^d)}}\right) \leq 2de^{-c\kappa^2}.$$

**Proof.** We write

$$L_n = \frac{1}{n} \sum_{j=1}^n C_j \quad \text{with } C_j := \int_{\mathbb{R}^d} \operatorname{Re}\left(\frac{e^{i\langle u, Y_j \rangle} - \varphi(u)}{|u|^2 \mathcal{L}'(-\psi(u))}\right) \Theta(u) w_U(u) \, du,$$

where  $C_j$  are independent, centred and symmetric matrices in  $\mathbb{R}^{d \times d}$ . In order to apply the noncommutative Bernstein inequality in [32, Theorem 4], we need to bound  $\|C_j\|_\infty$  and  $\|\mathbb{E}[C_j^2]\|_\infty$ . Since  $\|\Theta(u)\|_\infty = 1$ , we have

$$\|C_j\|_\infty \leq \int_{\mathbb{R}^d} \frac{2}{|u|^2 |\mathcal{L}'(-\psi(u))|} w_U(u) \, du \leq 2\xi_U^{-1} \int_{\mathbb{R}^d} |v|^{-2} w(v) \, dv.$$

Using that  $\operatorname{Re}(z_1) \operatorname{Re}(z_2) = \frac{1}{2}(\operatorname{Re}(z_1 z_2) + \operatorname{Re}(z_1 \bar{z}_2))$  for  $z_1, z_2 \in \mathbb{C}$  and symmetry in  $v$ , the Variance of  $C_j$  is bounded as follows:

$$\begin{aligned} \mathbb{E}[C_j^2] &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{E}\left[\operatorname{Re}\left(\frac{e^{i\langle u, Y_1 \rangle} - \varphi(u)}{\mathcal{L}'(-\psi(u))}\right) \operatorname{Re}\left(\frac{e^{i\langle v, Y_1 \rangle} - \varphi(v)}{\mathcal{L}'(-\psi(v))}\right)\right] \frac{\Theta(u)}{|u|^2} \frac{\Theta(v)}{|v|^2} w_U(u) w_U(v) \, du \, dv \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \operatorname{Re}\left(\frac{\varphi(u+v) - \varphi(u)\varphi(v)}{\mathcal{L}'(-\psi(u))\mathcal{L}'(-\psi(v))}\right) \frac{\Theta(u)}{|u|^2} \frac{\Theta(v)}{|v|^2} w_U(u) w_U(v) \, du \, dv. \end{aligned}$$

To estimate  $\|\mathbb{E}[C_j^2]\|_\infty$  we bound the spectral norm of the integral by the integral over the spectral norms (Minkowski inequality). Moreover, we use that for any functions  $f : \mathbb{R}^d \rightarrow \mathbb{C}$  and  $g : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{C}$  with  $|g(u, v)| = |g(v, u)|$  the Cauchy–Schwarz inequality and Fubini’s theorem yield

$$\begin{aligned} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(u) f(v) g(u, v)| \, du \, dv &\leq \|f(u) \sqrt{|g(u, v)|}\|_{L^2(\mathbb{R}^{2d})} \|f(v) \sqrt{|g(u, v)|}\|_{L^2(\mathbb{R}^{2d})} \\ &= \int_{\mathbb{R}^d} |f(u)|^2 \int_{\mathbb{R}^d} |g(u, v)| \, dv \, du. \end{aligned} \tag{8.4}$$

Taking into account the compact support of  $w_U$  and applying the previous estimate to the functions  $f(u) = w_U(u) / (|u|^2 \mathcal{L}'(-\psi(u)))$  and  $g(u, v) = (\varphi(u+v) - \varphi(u)\varphi(v)) \mathbb{1}_{\{|u| \leq U/2\}} \mathbb{1}_{\{|v| \leq U/2\}}$ , we obtain

$$\begin{aligned} \|\mathbb{E}[C_j^2]\|_\infty &\leq \int_{\mathbb{R}^d} \frac{w_U(u)^2}{|u|^4 |\mathcal{L}'(-\psi(u))|^2} \int_{|v| \leq U/2} (|\varphi(u+v)| + |\varphi(u)\varphi(-v)|) \, dv \, du \\ &\leq 2\|\varphi\|_{L^1(B_U^d)} \int_{\mathbb{R}^d} \frac{w_U(u)^2}{|u|^4 \mathcal{L}'(-\psi(u))^2} \, du. \end{aligned}$$

Using  $w_U(u) = U^{-d} w(u/U)$ , we conclude

$$\begin{aligned} \|\mathbb{E}[C_j^2]\|_\infty &\leq 2\|\varphi\|_{L^1(B_U^d)} \left(\inf_{|u| \leq U} \mathcal{L}'(-\psi(u))\right)^{-2} \int_{\mathbb{R}^d} |u|^{-4} w_U(u)^2 \, du \\ &\leq 2\xi_U^{-2} U^{-d} \|\varphi\|_{L^1(B_U^d)} \int_{\mathbb{R}^d} |v|^{-4} w(v)^2 \, dv. \end{aligned}$$

Consequently, Theorem 4 from [32] yields

$$\mathbb{P}\left(\|L_n\|_\infty \geq \frac{\kappa}{\sqrt{n}\xi_U} \frac{\|\varphi\|_{L^1(B_U^d)}^{1/2}}{U^{d/2}}\right) \leq 2d \exp\left(-\frac{c\kappa^2}{1 + \kappa(\|\varphi\|_{L^1(B_U^d)})^{1/2}}\right)$$

for some constant  $c > 0$  depending only on  $w$ .  $\square$

**Proof of Theorem 4.2.** We write again  $q = 1$  if Assumption C(iv) is satisfied and  $q = 0$  otherwise. Applying Lemmas 8.3 and 8.4 we deduce from (4.2) on the event  $\mathcal{H}_{n,U}$ , defined the linearisation lemma,

$$\begin{aligned} \|\mathcal{R}_n\|_\infty &\leq 4 \left\| \int_{\mathbb{R}^d} \frac{1}{|u|^2} \operatorname{Re}(\mathcal{L}^{-1}(\varphi_n(u)) - \mathcal{L}^{-1}(\varphi(u))) \Theta(u) w_U(u) du \right\|_\infty \\ &\quad + 4 \int_{\mathbb{R}^d} \frac{|\operatorname{Re} \Psi(u) + \alpha|}{|u|^2} w_U(u) du \\ &\leq 4\|L_n\|_\infty + 16C_L \int_{\mathbb{R}^d} \frac{|\varphi_n(u) - \varphi(u)|^2}{|u|^2 |\psi(u)|^q |\mathcal{L}'(-\psi(u))|^2} w_U(u) du + DC_\nu U^{-s-2} \\ &\leq 4\|L_n\|_\infty + 16C_L \frac{\|\varphi_n - \varphi\|_U^2}{\inf_{|u| \leq U} |\psi(u)|^q |\mathcal{L}'(-\psi(u))|^2} \int_{\mathbb{R}^d} |u|^{-2} w_U(u) du + DC_\nu U^{-s-2} \\ &\leq 4\|L_n\|_\infty + DC_L \frac{\|\varphi_n - \varphi\|_U^2}{U^2 \inf_{|u| \leq U} |\psi(u)|^q |\mathcal{L}'(-\psi(u))|^2} + DC_\nu U^{-s-2} \end{aligned}$$

for some constant  $D > 0$  depending only on  $w$  and  $s$ . Writing again  $\xi_U := U^2 \inf_{|u| \leq U} |\mathcal{L}'(-\psi(u))|$  and defining  $\zeta_U := U^2 \inf_{|u| \leq U} |\psi(u)|^q |\mathcal{L}'(-\psi(u))|^2$ , we obtain

$$\begin{aligned} &\mathbb{P}\left(\|\mathcal{R}_n\|_\infty \geq \frac{\kappa}{\sqrt{n}\xi_U} \sqrt{U^{-d} \|\varphi\|_{L^1(B_U^d)} + DC_\nu U^{-s-2}}\right) \\ &\leq \mathbb{P}\left(\|L_n\|_\infty \geq \frac{\kappa}{8} \frac{\|\varphi\|_{L^1(B_U^d)}^{1/2}}{\sqrt{n}U^{d/2}\xi_U}\right) + \mathbb{P}\left(\|\varphi_n - \varphi\|_U^2 \geq \frac{\kappa}{2DC_L} \frac{\|\varphi\|_{L^1(B_U^d)}^{1/2} \zeta_U}{U^{d/2} \sqrt{n}\xi_U}\right) + \mathbb{P}(\mathcal{H}_{n,U}^c). \end{aligned} \quad (8.5)$$

The first probability is bounded by Lemma 8.5. Defining

$$\delta_n := \frac{\sqrt{n} \|\varphi\|_{L^1(B_U^d)}^{1/2} \zeta_U}{U^{d/2} \xi_U} = \frac{\sqrt{n} \|\varphi\|_{L^1(B_U^d)}^{1/2} \inf_{|u| \leq U} |\psi(u)|^q |\mathcal{L}'(-\psi(u))|^2}{U^{d/2} \inf_{|u| \leq U} |\mathcal{L}'(-\psi(u))|}$$

and using that  $\kappa \leq \delta_n$  by assumption, we can bound the second probability in (8.5) by Theorem A.2:

$$\mathbb{P}\left((\sqrt{n} \|\varphi_n - \varphi\|_U)^2 \geq \frac{\kappa \delta_n}{2DC_L}\right) \leq \mathbb{P}\left((\sqrt{n} \|\varphi_n - \varphi\|_U)^2 \geq \frac{\kappa^2}{2DC_L}\right) \leq 2e^{-c\kappa^2}$$

for some numerical constant  $c > 0$ , provided  $\kappa \geq \sqrt{d \log(d+1)} (\log U)^\rho$  for some  $\rho > 1/2$  and  $\kappa \leq \sqrt{n}$ . The probability of the complement of  $\mathcal{H}_{n,U}$  can be similarly estimated by

$$\mathbb{P}(\mathcal{H}_n^c) \leq \mathbb{P}\left(\sqrt{n} \|\varphi_n - \varphi\|_U \geq \frac{C}{2} \sqrt{n} U^{-2} \xi_U\right)$$

which yields the claimed bound owing to Theorem A.2 and  $\kappa^2 \leq n \xi_U^2 U^{-4}$ .  $\square$

### 8.2. Proof of the lower bounds: Theorem 4.6

We follow the standard strategy to prove lower bounds adapting some ideas by Belomestny et al. [7, Chapter 1]. We start with the proof of (i) which is divided into several steps.

*Step 1:* We need to construct two alternatives of Lévy triplets. Let  $K : \mathbb{R}^k \rightarrow \mathbb{R}$  be a kernel given via its Fourier transform

$$\mathcal{F}K(u) = \begin{cases} 1, & |u| \leq 1, \\ \exp\left(-\frac{e^{-1/(|u|-1)}}{2-|u|}\right), & 1 < |u| < 2, \\ 0, & |u| \geq 2, \end{cases} \quad u \in \mathbb{R}^k.$$

Since  $\mathcal{F}K$  is real and even,  $K$  is indeed a real valued function. For each  $n$  we define two jump measures  $\nu_0$  and  $\nu_n$  on  $\mathbb{R}^d$  via their Lebesgue density, likewise denoted by  $\nu_0$  and  $\nu_n$ , respectively. Slightly abusing notation we define the densities on  $\mathbb{R}^k$  and set the remaining  $d - k$  coordinates equal to zero. Denoting the Laplace operator by  $\Delta := \sum_{j=1}^k \partial_j^2$ , we set

$$\nu_0(x) := \left(1 + \sum_{j=1}^k |x_j|^{2L}\right)^{-1}, \quad x = (x_1, \dots, x_k)^\top \in \mathbb{R}^k,$$

$$\nu_n(x) := \nu_0(x) + a\delta_n^{s-k}(\Delta K)(x/\delta_n), \quad x \in \mathbb{R}^k,$$

for  $L \in \mathbb{N}$  such that  $2L > k + p \vee (-s)$  and  $L > k$ , some positive sequence  $\delta_n \rightarrow 0$ , to be chosen later and a sufficiently small constant  $a > 0$ . Since for any  $l \in \mathbb{N}$  it holds  $|x|^l |\Delta K(x)| \leq C_l$  uniformly and due to the assumption  $k \leq s$ ,  $\nu_0$  and  $\nu_n$  are non-negative finite measures. In particular, they are Lévy measures.

By construction  $\nu_0 \in \mathfrak{S}(s, p, C_\nu)$  for any  $s > -2$ ,  $p > 0$  and some  $C_\nu > 0$  (by rescaling  $C_\nu$  can be arbitrary). To verify that  $\nu_n \in \mathfrak{S}(s, p, C_\nu)$  holds for some sufficiently small  $a > 0$  and for all  $n \in \mathbb{N}$ , we first note that  $\int_{\mathbb{R}^k} |x|^{-s} \delta_n^{s-k} |\Delta K|(x/\delta_n) dx = \int_{\mathbb{R}^k} |y|^{-s} |\Delta K|(y) dy$  for  $s \in (-2, 0]$ . In the case  $s > 0$  we use

$$(1 + |u|^2)^{s/2} |\mathcal{F}[\delta_n^{s-k} \Delta K(x/\delta_n)](u)| = \delta_n^{s+2} (1 + |u|^2)^{s/2} |u|^2 |\mathcal{F}K(\delta_n u)|$$

$$\leq \delta_n^{s+2} (1 + \delta_n^{-2})^{s/2} \delta_n^{-2} \lesssim 1,$$

owing to the compact support of  $\mathcal{F}K$ .

Now define the rank  $k$  diagonal matrix  $\Sigma_0 = \text{diag}(1, \dots, 1, 0, \dots, 0)$  (i.e.,  $k$  ones followed by  $d - k$  zeros) and its perturbation  $\Sigma_n := (1 + 2a\delta_n^{2+s})\Sigma_0$ . Finally define

$$Y_t^{(0)} = \Sigma_0 W_t + Z_t^{(0)} \quad \text{and} \quad Y_t^{(n)} = \Sigma_n W_t + Z_t^{(n)}$$

with a Brownian motion  $W_t$  and with  $Z_t^{(0)}$  and  $Z_t^{(n)}$  being compound Poisson processes independent of  $W_t$ , with jump measures  $\nu_0$  and  $\nu_n$ , respectively.

*Step 2:* We now bound the  $\chi^2$  distance of the observation laws  $\mathbb{P}_0^{\otimes n} := \mathbb{P}_{(\Sigma_0, \nu_0, \mathcal{T})}^{\otimes n}$  and  $\mathbb{P}_n^{\otimes n} := \mathbb{P}_{(\Sigma_n, \nu_n, \mathcal{T})}^{\otimes n}$ . First we observe that both laws are equal on the last  $d - k$  coordinates, namely being a Dirac measure in zero. Owing to the diffusion component, the marginals  $\mathbb{P}_0$  and  $\mathbb{P}_n$  admit Lebesgue densities on  $\mathbb{R}^k$  denoted by  $f_0$  and  $f_n$ , respectively (cf. [36, Theorem 27.7]). Since the observations are i.i.d.,  $\chi^2(\mathbb{P}_n^{\otimes n}, \mathbb{P}_0^{\otimes n})$  is uniformly bounded in  $n$ , if

$$n\chi^2(\mathbb{P}_n, \mathbb{P}_0) = n \int_{\mathbb{R}^k} \frac{|f_n(x) - f_0(x)|^2}{f_0(x)} dx < c$$

for some constant  $c > 0$ . The density  $f_0$  is given by  $f_0(x) = \int_{\mathbb{R}_+} p_t(x) \pi(dt)$ , where  $p_t$  denotes the density of  $Y_t^{(0)}$ . Since  $Y^0$  is of compound Poisson type, its marginal density is given by the convolution exponential

$$p_t(x) = \mu_{0,t\Sigma_0} * \left( e^{-t\nu_0(\mathbb{R}^k)} \sum_{j=0}^{\infty} \frac{t^j \nu_0^{*j}}{j!} \right) (x) \geq t e^{-t\nu_0(\mathbb{R}^k)} (\mu_{0,t\Sigma_0} * \nu_0)(x)$$

with the density  $\mu_{0,t\Sigma_0}$  of the  $\mathcal{N}(0, t\Sigma_0)$ -distribution. Using that there is some interval  $[r, s] \subseteq (0, \infty)$  with  $\pi([r, s]) > 0$  and that  $\nu_0$  is independent of  $t$ , we obtain

$$f_0(x) \gtrsim (\mu_{0,r\Sigma_0} * \nu_0)(x) \gtrsim \left(1 + \sum_{j=1}^k |x_j|^{2L}\right)^{-1}, \quad x = (x_1, \dots, x_k)^\top \in \mathbb{R}^k.$$

By Plancherel's identity we thus have

$$\begin{aligned} \chi^2(\mathbb{P}_n, \mathbb{P}_0) &\leq \int_{\mathbb{R}^k} \left(1 + \sum_{j=1}^k |x_j|^{2L}\right) |f_n(x) - f_0(x)|^2 dx \\ &\lesssim \|\varphi_n - \varphi_0\|_{L^2(\mathbb{R}^k)}^2 + \sum_{j=1}^k \|\partial_j^L(\varphi_n - \varphi_0)\|_{L^2(\mathbb{R}^k)}^2, \end{aligned}$$

where  $\varphi_0$  and  $\varphi_n$  denote the characteristic functions of  $\mathbb{P}_0$  and  $\mathbb{P}_n$ , respectively.

*Step 3:* We have to estimate the distance of the characteristic functions. Let us denote the characteristic exponents of the Lévy processes  $Y_t^{(0)}$  and  $Y_t^{(n)}$  (restricted on the first  $k$  coordinates) by  $\psi_0$  and  $\psi_n$ , respectively. Then,

$$\psi_m(u) = -\frac{1}{2} \langle u, \Sigma_m u \rangle + \int_{\mathbb{R}^k} (e^{i\langle u, x \rangle} - 1 - i\langle u, x \rangle) \nu_m(x) dx, \quad m \in \{0, n\}.$$

Note that  $\psi_m$  is real valued because  $\nu_m$  is even. Using Taylor's formula, we obtain

$$\begin{aligned} \varphi_n(u) - \varphi_0(u) &= \mathcal{L}(-\psi_n(u)) - \mathcal{L}(-\psi_0(u)) \\ &= -(\psi_n(u) - \psi_0(u)) \int_0^1 \mathcal{L}'(-\psi_0(u) - t(\psi_n(u) - \psi_0(u))) dt. \end{aligned}$$

Defining  $\Psi_{n,t}(u) := -\psi_0(u) - t(\psi_n(u) - \psi_0(u))$ , we thus have

$$\partial_j^L(\varphi_n - \varphi_0)(u) = \sum_{r=0}^L \partial_j^r(\psi_n(u) - \psi_0(u)) \int_0^1 \partial_j^{L-r} \mathcal{L}'(\Psi_{n,t}(u)) dt,$$

where the partial derivatives of the composition  $\mathcal{L}' \circ \Psi_{n,t}$  can be computed with Faà di Bruno's formula. Since  $\int (e^{i\langle u, x \rangle} - 1 - i\langle u, x \rangle) \Delta K(x/\delta_n) dx = -\delta_n^{2+k} |u|^2 \mathcal{F}K(\delta_n u)$ , we have

$$\psi_n(u) - \psi_0(u) = a\delta_n^{2+s} |u|^2 (1 - \mathcal{F}K(\delta_n u))$$

and in particular  $\Psi_{n,t}(u) = -\psi_0(u)(1 + o(1))$  uniformly over  $u$  and  $t$  for  $\delta_n \rightarrow 0$ . Taking into account the properties

$$\begin{aligned} \mathcal{F}K(u) &= 0 \quad \text{for } |u| > 2/\delta_n, \\ \partial_j^r \mathcal{F}K(u) &= 0 \quad \text{for } |u| \leq 1/\delta_n \text{ and } |u| > 2/\delta_n, r = 1, \dots, L, \\ |\partial_j^r \mathcal{F}K(u)| &\leq C \quad \text{for } 1/\delta_n < |u| \leq 2/\delta_n, r = 0, 1, \dots, L, \end{aligned}$$

for all  $j = 1, \dots, k$  and some  $C > 0$ , we see that  $\psi_n(u) - \psi_0(u)$  is zero for  $|u| < 1/\delta_n$  and that  $|\partial_j \psi_m(u)| \lesssim 1 + |u_j|$ ,  $|\partial_j^r \psi_m(u)| \lesssim 1$  for  $r = 2, \dots, L$  and  $m \in \{0, n\}$ . We conclude

$$\|\partial_j^L(\varphi_n - \tilde{\varphi}_n)\|_{L^2(\mathbb{R}^k)}^2 \lesssim a\delta_n^{2s+4} \int_{|u| > 1/\delta_n} |u|^4 \sum_{r=0}^L |\mathcal{L}^{(1+r)}(-\psi_n(u)(1 + o(1)))|^2 |u|^{2r} du.$$

Due to monotonicity of  $\mathcal{L}'(-x)$  for  $x > 0$  and  $\mathcal{L}^{(r+1)}(x)/\mathcal{L}^{(r)}(x) = O(1/|x|)$  for  $|x| \rightarrow \infty, r = 1, \dots, L$ , the previous estimate and Step 2 yield as  $n \rightarrow \infty$

$$n\chi^2(\mathbb{P}_n, \mathbb{P}_0) \lesssim an\delta_n^{2(s+2)} \int_{|u|>1/\delta_n} |\mathcal{L}'(-\psi_0(u))|^2 |u|^4 \, du \lesssim an\delta_n^{2s+4+4\eta-k}$$

if  $4\eta > k$ . Hence,  $\chi^2(\mathbb{P}_n^{\otimes n}, \mathbb{P}_0^{\otimes n})$  remains bounded for  $\delta_n = n^{-1(2s+4+4\eta-k)}$  and with some sufficiently small  $a > 0$ .

*Step 4:* Noting that

$$\|\Sigma_0 - \tilde{\Sigma}_n\|_2 = 2\sqrt{ka}\delta_n^{s+2},$$

the first lower bound in Theorem 4.6 follows from Theorem 2.2 in [39].

*Step 5:* For the second case, i.e.,  $k > 2\eta$ , we modify our construction as follows: We use the jump measures  $\nu_0$  and  $\nu_n$  from before, but only on the first derivative. We use the same rank  $k$  diffusion matrix  $\Sigma_0$  with  $k$  ones on the diagonal, but choose the alternative as  $\Sigma_n = \text{diag}(1 + 2a\delta_n^{2+s}, 1, \dots, 1, 0, \dots, 0)$  where the last  $d - k$  entries are zero. Since the corresponding laws  $\mathbb{P}_0$  and  $\mathbb{P}_n$  are product measures which differ only on the first coordinate, the calculations from Step 2 and 3 yield

$$\begin{aligned} n\chi^2(\mathbb{P}_n, \mathbb{P}_0) &\lesssim an\delta_n^{2(s+2)} \int_{\mathbb{R} \setminus [-1/\delta_n, 1/\delta_n]} |\mathcal{L}'(-\psi_0(u))|^2 |u|^4 \, du \\ &\leq an\delta_n^{2(s+2)} \sup_{|u|>1/\delta_n} \{|\mathcal{L}'(-\psi_0(u))||u|^2\} \int_{\mathbb{R}} |\mathcal{L}'(-\psi_0(u))||u|^2 \, du \\ &\lesssim an\delta_n^{2(s+2)+2\eta} \int_{\mathbb{R}} (1 + |u|^2)^{-\eta-1} |u|^2 \, du, \end{aligned}$$

where the integral is finite by the assumption  $\eta > 1/2$ . Hence, we have shown the second lower bound Theorem 4.6(i).

The result in (ii) can be deduced analogously, choosing  $k = 1$ . In Step 3 we obtain under the corresponding assumption on  $\mathcal{L}$  that with some constant  $c > 0$

$$\begin{aligned} n\chi^2(\mathbb{P}_n, \mathbb{P}_0) &\lesssim an\delta_n^{2(s+2)} \int_{|u|>1/\delta_n} |\mathcal{L}'(-\psi_0(u))|^2 |u|^4 (1 + |u|^{2L}) \, du \\ &\lesssim an\delta_n^{2s+4} e^{-c\delta_n^{-2\eta}} \end{aligned}$$

which remains bounded if  $\delta_n \sim (\log n)^{-1/(2\eta)}$ . □

### 8.3. The mixing case: Proof of Theorem 5.1

We denote again  $\xi_U := U^2 \inf_{|u| \leq U} |\mathcal{L}'(-\psi(u))|$  and recall the event  $\mathcal{H}_{n,U}$  defined in Lemma 8.4. Applying (5.2) and Lemmas 8.3 and 8.4, we obtain on  $\mathcal{H}_{n,U}$

$$\begin{aligned} \|\mathcal{R}_n\|_\infty &\leq 4 \int_{\mathbb{R}^d} |u|^{-2} |\text{Re}(\mathcal{L}^{-1}(\varphi_n(u)) - \mathcal{L}^{-1}(\varphi(u)))| w_U(u) \, du + 4 \int_{\mathbb{R}^d} |u|^{-2} |\Psi(u) + \alpha| w_U(u) \, du \\ &\leq 4 \int_{\mathbb{R}^d} \frac{|\varphi_n(u) - \varphi(u)|}{|u|^2 |\mathcal{L}'(-\psi(u))|} w_U(u) \, du + 16C_L \int_{\mathbb{R}^d} \frac{|\varphi_n(u) - \varphi(u)|^2}{|u|^2 |\mathcal{L}'(-\psi(u))|^2} w_U(u) \, du \\ &\quad + 4 \int_{\mathbb{R}^d} |u|^{-2} |\Psi(u) + \alpha| w_U(u) \, du \\ &\leq C(\xi_U^{-1} \|\varphi_n - \varphi\|_U + U^2 \xi_U^{-2} \|\varphi_n - \varphi\|_U^2 + U^{-s-2}), \end{aligned}$$



where the constant  $C > 0$  depends only on  $w, C_L$  and  $C_v$ . Therefore,

$$\begin{aligned} & \mathbb{P}\left(\|\mathcal{R}_n\|_\infty \geq \frac{\kappa(\log U)^\rho}{\sqrt{n}\xi_U} + CU^{-s-2}\right) \\ & \leq \mathbb{P}\left(\sqrt{n}(\log U)^{-\rho}\|\varphi_n - \varphi\|_U \geq \frac{\kappa}{2C}\right) + \mathbb{P}\left((\sqrt{n}(\log U)^{-\rho}\|\varphi_n - \varphi\|_U)^2 \geq \frac{\sqrt{n}\kappa\xi_U}{2CU^2(\log U)^\rho\xi_U}\right) \\ & \quad + \mathbb{P}(\mathcal{H}_{n,U}^c). \end{aligned}$$

If  $\kappa \leq \sqrt{n}\xi_U(\log U)^{-\rho}U^{-2}$ , we have

$$\mathbb{P}(\mathcal{H}_{n,U}^c) \leq \mathbb{P}\left(\sqrt{n}(\log U)^{-\rho}\|\varphi_n - \varphi\|_U \geq \frac{2\kappa}{C_L}\right).$$

Theorem A.4 yields

$$\mathbb{P}\left(\|\mathcal{R}_n\|_\infty \geq \frac{\kappa(\log U)^\rho}{\sqrt{n}\xi_U} + CU^{-s-2}\right) \lesssim e^{-c\kappa^2} + n^{-p/2}$$

some  $c > 0$  and for any  $n \in \mathbb{N}$  and  $\kappa \in (\underline{\xi}\sqrt{d\log n}, \bar{\xi}\sqrt{n}/\log^2 n)$ . □

### Appendix: Multivariate uniform bounds for the empirical characteristic function

#### A.1. I.i.d. sequences

Let us recall the usual multi-index notation. For a multi-index  $\beta = (\beta_1, \dots, \beta_d) \in \mathbb{N}^d$ , a vector  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  and a function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  we write

$$\begin{aligned} |\beta| &:= \beta_1 + \dots + \beta_d, & x^\beta &:= x_1^{\beta_1} \dots x_d^{\beta_d}, & |x|^\beta &:= |x_1|^{\beta_1} \dots |x_d|^{\beta_d}, \\ \partial^\beta f &:= \partial_1^{\beta_1} \dots \partial_d^{\beta_d} f. \end{aligned}$$

We need a multivariate (straight forward) generalisation of Theorem 4.1 by Neumann and Reiß [30]. For a sequence of independent random vectors  $(Y_j)_{j \geq 1} \subseteq \mathbb{R}^d$  we define the empirical process corresponding to the empirical characteristic function by

$$C_n(u) := \sqrt{n}(\varphi_n(u) - \varphi(u)) = n^{-1/2} \sum_{j=1}^n (e^{i\langle u, Y_j \rangle} - \mathbb{E}[e^{i\langle u, Y_1 \rangle}]), \quad u \in \mathbb{R}^d, n \geq 1.$$

**Proposition A.1.** *Let  $\beta \in \mathbb{N}^d$  be a multi-index and let  $Y_1, \dots, Y_n \in \mathbb{R}^d$  be i.i.d.  $d$ -dimensional random vectors satisfying  $\mathbb{E}[|Y_1|^{2\beta}|Y_1|^\gamma] < \infty$  for some  $\gamma > 0$ . Using the weight function  $w(u) = (\log(e + |u|))^{-1/2-\delta}$ ,  $u \in \mathbb{R}^d$ , for some  $\delta > 0$ , there is a constant  $C > 0$  such that*

$$\sup_{n \geq 1} \mathbb{E}[\|w(u)\partial^\beta C_n(u)\|_\infty] \leq C\sqrt{d}(\sqrt{\log d} + 1).$$

**Proof.** The proof relies on a bracketing entropy argument and we first recall some definitions. For two functions  $l, u : \mathbb{R}^d \rightarrow \mathbb{R}$  a bracket is given by  $[l, u] := \{f : \mathbb{R}^d \rightarrow \mathbb{R} \mid l \leq f \leq u\}$ . For a set of functions  $G$  the  $L^2$ -bracketing number  $N_{[]}(\varepsilon, G)$  denotes the minimal number of brackets  $[l_k, u_k]$  satisfying  $\mathbb{E}[(u_k(Y_1) - l_k(Y_1))^2] \leq \varepsilon^2$  which are necessary to cover  $G$ . The bracketing integral is given by

$$J_{[]}(\delta, G) := \int_0^\delta \sqrt{N_{[]}(\varepsilon, G)} \, d\varepsilon.$$

A function  $F : \mathbb{R}^d \rightarrow \mathbb{R}$  is called envelop function of  $G$  if  $|f| \leq F$  for any  $f \in G$ .

Decomposing  $C_n$  into the real and the imaginary part, we consider the set  $G_\beta := \{g_u : u \in \mathbb{R}^d\} \cup \{h_u : u \in \mathbb{R}^d\}$  where

$$g_u : \mathbb{R}^d \rightarrow \mathbb{R}, \quad y \mapsto w(u) \frac{\partial^\beta}{\partial u^\beta} \cos(\langle u, y \rangle), \quad h_u : \mathbb{R}^d \rightarrow \mathbb{R}, \quad y \mapsto w(u) \frac{\partial^\beta}{\partial u^\beta} \sin(\langle u, y \rangle).$$

Noting that  $G_\beta$  has the envelop function  $F(y) = |y|^\beta$ , Lemma 19.35 in [40] yields

$$\mathbb{E}[\|w(u)\partial^\beta C_n(u)\|_\infty] \lesssim J_\square(\mathbb{E}[F(Y_1)^2], G_\beta).$$

Since the real and the imaginary part can be treated analogously, we concentrate in the following on  $\{g_u : u \in \mathbb{R}^d\}$ . Owing to  $|g_u(y)| \leq w(u)|y|^\beta$ , we have  $\{g_u : |u| > B\} \subseteq [g_0^-, g_0^+]$  for  $g_0^\pm(y) := \pm \varepsilon |y|^\beta$  and

$$B := B(\varepsilon) := \inf\left\{b > 0 : \sup_{|u| \geq b} w(u) \leq \varepsilon\right\}.$$

To cover  $\{g_u : |u| \leq B\}$ , we define for some grid  $(u_j)_{j \geq 1} \subseteq \mathbb{R}^d$  and  $j \geq 1$

$$g_j^\pm(y) := \left( w(u_j) \frac{\partial^\beta}{\partial u^\beta} \cos(\langle u_j, y \rangle) \pm \varepsilon |y|^\beta \right) \mathbb{1}_{\{|y| \leq M\}} \pm |y|^\beta \mathbb{1}_{\{|y| > M\}}$$

with

$$M := M(\varepsilon) := \inf\{m > 0 : \mathbb{E}[|Y_1|^{2\beta} \mathbb{1}_{\{|Y_1| > m\}}] \leq \varepsilon^2\}.$$

We have  $\mathbb{E}[|g_j^+(Y_1) - g_j^-(Y_1)|^2] \leq 4\varepsilon^2(\mathbb{E}[|Y_1|^{2\beta}] + 1)$  for  $j \geq 0$ . Denoting the Lipschitz constant of  $w$  by  $L$ , it holds

$$\left| w(u) \frac{\partial^\beta}{\partial u^\beta} \cos(\langle u, y \rangle) - w(u_j) \frac{\partial^\beta}{\partial u^\beta} \cos(\langle u_j, y \rangle) \right| \leq |y|^\beta (L + |y|) |u - u_j|.$$

Therefore,  $g_u \in [g_j^-, g_j^+]$  if  $(L + M)|u - u_j| \leq \varepsilon$ . Since any (Eukclidean) ball in  $\mathbb{R}^d$  with radius  $B$  can be covered with fewer than  $(B/\tilde{\varepsilon})^d$  cubes with edge length  $2\tilde{\varepsilon}$  and each of these cubes can be covered with a ball of radius  $\sqrt{d}\tilde{\varepsilon}$  (use  $|\cdot| \leq \sqrt{d}\|\cdot\|_{\ell^\infty}$ ), we choose  $\tilde{\varepsilon} = \varepsilon d^{-1/2}/(L + M)$  to see that

$$N_\square(\varepsilon, G_\beta) \leq 2 \left( \frac{\sqrt{d}B(L + M)}{\varepsilon} \right)^d + 2.$$

By the choice of  $w$  it holds  $B \leq \exp(\varepsilon^{-1/(1/2+\delta)})$  and Markov's inequality yields  $M \leq (\varepsilon^{-2}\mathbb{E}[|Y_1|^{2\beta} \|Y_1\|^\gamma])^{1/\gamma}$ . The bracketing entropy is thus bounded by

$$\log N_\square(\varepsilon, G_\beta) \lesssim d(\log d + \varepsilon^{-1/(1/2+\delta)} + \log(\varepsilon^{-2/\gamma-1})) \lesssim d(\log d + \varepsilon^{-2/(1+2\delta)})$$

and the entropy integral can be estimated by

$$J_\square(\mathbb{E}[F(Y_1)^2], G_\beta) \lesssim \sqrt{d} \left( \sqrt{\log d} + \int_0^{\mathbb{E}[|Y_1|^{2\beta}]} \varepsilon^{-1/(1+2\delta)} d\varepsilon \right) \lesssim \sqrt{d}(\sqrt{\log d} + 1). \quad \square$$

Applying Talagrand's inequality, we conclude the following concentration result, see also Proposition 3.3 in [7, Chapter 1].

**Theorem A.2.** *Let  $Y_1, \dots, Y_n \in \mathbb{R}^d$  be i.i.d.  $d$ -dimensional random vectors satisfying  $\mathbb{E}[|Y_1|^\gamma] < \infty$  for some  $\gamma > 0$ . For any  $\delta > 0$  there is some numerical constant  $c > 0$  independent of  $d, n, U$  such that*

$$\mathbb{P}\left(\sup_{|u| \leq U} |C_n(u)| \geq \kappa\right) \leq 2e^{-c\kappa^2},$$

for any  $\kappa \in [\sqrt{d}(\sqrt{\log d} + 1)(\log U)^{1/2+\delta}, \sqrt{n}]$ .

**Proof.** We will apply Talagrand’s inequality in Bousquet’s version (cf. [25, (5.50)]). Let  $T \subseteq [-U, U]^d$  be a countable index set. Noting that  $Z_{j,u} := n^{-1/2}(e^{i\langle u, Y_j \rangle} - \mathbb{E}[e^{i\langle u, Y_1 \rangle}])$  are centred and i.i.d. random variables satisfying  $|Z_{k,u}| \leq 2n^{-1/2}$ , for all  $u \in T, k = 1, \dots, n$ , as well as  $\sup_{u \in T} \text{Var}(\sum_{k=1}^n Z_{k,u}) \leq 1$ , we have for all  $\kappa > 0$

$$\mathbb{P}\left(\sup_{u \in T} \left| \sum_{k=1}^n Z_{k,u} \right| \geq 4\mathbb{E}\left[\sup_{u \in T} \left| \sum_{k=1}^n Z_{k,u} \right|\right] + \sqrt{2\kappa} + \frac{4}{3}n^{-1/2}\kappa\right) \leq 2e^{-\kappa}.$$

Proposition A.1 yields  $\mathbb{E}[\sup_{u \in T} |\sum_{k=1}^n Z_{k,u}|] \leq C(\log U)^{1/2+\delta} \sqrt{d}(\sqrt{\log d} + 1)$  for some  $\delta, C > 0$ . Choosing  $T = \mathbb{Q}^d \cap [-U, U]^d$ , continuity of  $u \mapsto Z_{j,u}$  yields

$$\mathbb{P}\left(\sup_{|u| \leq U} |C_n(u)| \geq C\sqrt{d \log d}(\log U)^{1/2+\delta} + \sqrt{2\kappa} + \frac{4}{3}n^{-1/2}\kappa\right) \leq 2e^{-\kappa}. \quad \square$$

### A.2. Mixing sequences

If the sequence is not i.i.d., but only  $\alpha$ -mixing, there is no Talagrand-type inequality to work with. At least Merlevède et al. [27] have proven to following Bernstein-type concentration result. The bound of the constant  $v^2$  has been derived by Belomestny [6].

**Proposition A.3 ([27]).** *Let  $(X_k, k \geq 1)$  be a strongly mixing sequence of centred real-valued random variables on the probability space  $(\Omega, \mathcal{F}, P)$  with mixing coefficients satisfying*

$$\alpha(k) \leq \bar{\alpha}_0 \exp(-\bar{\alpha}_1 k), \quad k \geq 1, \bar{\alpha}_0, \bar{\alpha}_1 > 0. \tag{A.1}$$

If  $\sup_{k \geq 1} |X_k| \leq M$  a.s., then there is a positive constant depending on  $\bar{\alpha}_0$ , and  $\bar{\alpha}_1$  such that

$$\mathbb{P}\left(\sum_{k=1}^n X_k \geq \zeta\right) \leq \exp\left[-\frac{C\zeta^2}{nv^2 + M^2 + M\zeta \log^2(n)}\right]$$

for all  $\zeta > 0$  and  $n \geq 4$ , where

$$v^2 := \sup_k \left( \mathbb{E}[X_k]^2 + 2 \sum_{j \geq k} \text{Cov}(X_k, X_j) \right).$$

Moreover, there is a constant  $C' > 0$  such that

$$v^2 \leq \sup_k \mathbb{E}[X_k]^2 + C' \sup_k \mathbb{E}[X_k^2 \log^{2(1+\varepsilon)}(|X_k|^2)], \tag{A.2}$$

provided the expectations on the right-hand side are finite.

Let  $Z_j, j = 1, \dots, n$ , be a sequence of random vectors in  $\mathbb{R}^d$  with corresponding empirical characteristic function

$$\varphi_n(u) = \frac{1}{n} \sum_{j=1}^n \exp(i\langle u, Z_j \rangle), \quad u \in \mathbb{R}^d.$$

**Theorem A.4.** *Suppose that the following assumptions hold:*

(AZ1) *The sequence  $Z_j, j = 1, \dots, n$ , is strictly stationary and  $\alpha$ -mixing with mixing coefficients  $(\alpha_Z(k))_{k \in \mathbb{N}}$  satisfying*

$$\alpha_Z(k) \leq \bar{\alpha}_0 \exp(-\bar{\alpha}_1 k), \quad k \in \mathbb{N},$$

*for some  $\bar{\alpha}_0 > 0$  and  $\bar{\alpha}_1 > 0$ .*

(AZ2) *It holds  $\mathbb{E}[|Z_j|^p] < \infty$  for some  $p > 2$ .*

For arbitrary  $\delta > 0$  let the weighting function  $w : \mathbb{R}^d \rightarrow \mathbb{R}_+$  be given by

$$w(u) = \log^{-(1+\delta)/2}(e + |u|), \quad u \in \mathbb{R}. \quad (\text{A.3})$$

Then there are  $\underline{\xi}, \bar{\xi} > 0$  depending only on the characteristics of  $Z$  and  $\delta$ , such that for any  $n \in \mathbb{N}$  and for all  $\xi \in (\underline{\xi}\sqrt{d\log n}, \bar{\xi}\sqrt{n}/\log^2 n)$  the inequality

$$\mathbb{P}\left(\sqrt{n} \sup_{u \in \mathbb{R}^d} w(u) |\varphi_n(u) - \mathbb{E}[\varphi_n(u)]| > \xi\right) \leq C(e^{-c\xi^2} + n^{-p/2}) \quad (\text{A.4})$$

holds for constants  $C, c > 0$  independent of  $\xi, n$  and  $d$ .

**Proof.** We introduce the empirical process

$$\mathcal{W}_n(u) = \frac{1}{n} \sum_{j=1}^n w(u) (\exp(i\langle u, Z_j \rangle) - \mathbb{E}[\exp(i\langle u, Z_j \rangle)]), \quad u \in \mathbb{R}^d.$$

Consider the sequence  $A_k = e^k, k \in \mathbb{N}$ . As discussed in the proof of Proposition A.1, we can cover each ball  $\{U \in \mathbb{R}^d : |u| < A_k\}$  with  $M_k = (d^{1/2}A_k/\gamma)^d$  small balls with radius  $\gamma > 0$  and some centres  $u_{k,1}, \dots, u_{k,M_k}$ .

Since  $|\mathcal{W}_n(u)| \leq 2w(u) \downarrow 0$  as  $|u| \rightarrow \infty$ , there is for any  $\lambda > 0$  some finite integer  $K = K(\lambda)$  such that  $\sup_{|u| > A_k} |\mathcal{W}_n(u)| < \lambda$ . For  $|u| \leq A_k$  we use the bound

$$\begin{aligned} \max_{k=1, \dots, K} \sup_{A_{k-1} < |u| \leq A_k} |\mathcal{W}_n(u)| &\leq \max_{k=1, \dots, K} \max_{|u_{k,m}| > A_{k-1}} |\mathcal{W}_n(u_{k,m})| \\ &\quad + \max_{k=1, \dots, K} \max_{1 \leq m \leq M_k} \sup_{u: |u - u_{k,m}| \leq \gamma} |\mathcal{W}_n(u) - \mathcal{W}_n(u_{k,m})| \end{aligned}$$

to obtain

$$\begin{aligned} &\mathbb{P}\left(\sup_{u \in \mathbb{R}^d} |\mathcal{W}_n(u)| > \lambda\right) \\ &\leq \sum_{k=1}^K \sum_{|u_{k,m}| > A_{k-1}} \mathbb{P}(|\mathcal{W}_n(u_{k,m})| > \lambda/2) + \mathbb{P}\left(\sup_{|u-v| < \gamma} |\mathcal{W}_n(v) - \mathcal{W}_n(u)| > \lambda/2\right). \end{aligned} \quad (\text{A.5})$$

It holds for any  $u, v \in \mathbb{R}^d$

$$\begin{aligned} |\mathcal{W}_n(v) - \mathcal{W}_n(u)| &\leq 2|w(v) - w(u)| + \frac{1}{n} \sum_{j=1}^n |\exp(i\langle v, Z_j \rangle) - \exp(i\langle u, Z_j \rangle)| \\ &\quad + \frac{1}{n} \sum_{j=1}^n |\mathbb{E}[\exp(i\langle v, Z_j \rangle) - \exp(i\langle u, Z_j \rangle)]| \\ &\leq |u - v| \left( 2L_w + \frac{1}{n} \sum_{j=1}^n |Z_j| + \frac{1}{n} \sum_{j=1}^n \mathbb{E}[|Z_j|] \right), \end{aligned} \quad (\text{A.6})$$

where  $L_w$  is the Lipschitz constant of  $w$ . Markov's inequality and the moment inequality by Yokoyama [44] yield

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n} \sum_{j=1}^n (|Z_j| - \mathbb{E}[|Z_j|]) > c\right) &\leq c^{-p} n^{-p} \mathbb{E}\left[\left|\sum_{j=1}^n (|Z_j| - \mathbb{E}[|Z_j|])\right|^p\right] \\ &\leq C_p(\alpha) c^{-p} n^{-p/2} \end{aligned}$$

for any  $c > 0$  and where  $C_p(\alpha)$  is some constant depending on  $p$  and  $\alpha = (\bar{\alpha}_0, \bar{\alpha}_1)$  from Assumption (AZ1). In combination with (A.6) we obtain

$$\begin{aligned} \mathbb{P}\left(\sup_{|u-v|<\gamma} |\mathcal{W}_n(v) - \mathcal{W}_n(u)| > \lambda/2\right) &\leq \mathbb{P}\left(\frac{1}{n} \sum_{j=1}^n (|Z_j| - \mathbb{E}[|Z_j|]) > \frac{\lambda}{2\gamma} - 2(L_w + \mathbb{E}[|Z_1|])\right) \\ &\leq C_p(\alpha)n^{-p/2} \left(\frac{\lambda}{2\gamma} - 2(L_w + \mathbb{E}[|Z_1|])\right)^{-p}. \end{aligned}$$

Setting  $\gamma = \lambda/(6(L_w + \mathbb{E}[|Z_1|]))$ , we conclude

$$P\left(\sup_{|u-v|<\gamma} |\mathcal{W}_n(v) - \mathcal{W}_n(u)| > \lambda/2\right) \leq B_1 n^{-p/2}$$

with some constant  $B_1$  depending neither on  $\lambda$  nor  $n$ .

We turn now to the first term on the right-hand side of (A.5). If  $|u_{k,m}| > A_{k-1}$ , then it follows from Proposition A.3

$$\mathbb{P}(|\operatorname{Re}(\mathcal{W}_n(u_{k,m}))| > \lambda/4) \leq 2 \exp\left(-\frac{B_2 \lambda^2 n}{w^2(A_{k-1}) \log^{2(1+\delta)}(w(A_{k-1})) + \lambda \log^2(n) w(A_{k-1})}\right),$$

with some constant  $B_2 > 0$  depending only on the characteristics of the process  $Z$  and  $\delta > 0$ . The same bound holds true for  $\operatorname{Im}(\mathcal{W}_n(u_{k,m}))$ . Choosing  $\lambda = \xi n^{-1/2}$  for any  $0 < \xi \lesssim \sqrt{n}/\log^2(n)$  and taking into account the choice of  $\gamma$  from above, we get

$$\begin{aligned} \sum_{|u_{k,m}|>A_{k-1}} \mathbb{P}(|\mathcal{W}_n(u_{k,m})| > \lambda/2) &\leq 4 \left(\frac{d^{1/2} A_k}{\gamma}\right)^d \exp\left(-\frac{B_3 \lambda^2 n}{w^{2(1+\delta)}(A_{k-1}) + \lambda \log^2(n) w(A_{k-1})}\right) \\ &\lesssim d^{d/2} A_k^d n^{d/2} \xi^{-d} \exp\left(-\frac{B \xi^2}{w^{2(1+\delta)}(A_{k-1})}\right) \end{aligned}$$

with positive constants  $B_3, B$ . Fix  $\vartheta > 0$  such that  $B\vartheta = 2d$  and compute

$$\begin{aligned} \sum_{|u_{k,m}|>A_{k-1}} \mathbb{P}(\sqrt{n}|\mathcal{W}_n(u_{k,m})| > \xi/2) &\lesssim d^{d/2} \xi^{-d} e^{dk-\vartheta B(k-1)} n^{d/2} e^{-B(k-1)(\xi^2-\vartheta)} \\ &\leq d^{d/2} e^{2d} \xi^{-d} e^{k(d-\vartheta B)} e^{-B(k-1)(\xi^2-\vartheta)+d \log(n)/2}. \end{aligned}$$

If  $\xi^2 > \vartheta$  we obtain for any  $K > 0$

$$\sum_{k=2}^K \sum_{|u_{k,m}|>A_{k-1}} \mathbb{P}(\sqrt{n}|\mathcal{W}_n(u_{k,m})| > \xi/2) \lesssim (d^{1/2} e^4)^d \xi^{-d} e^{-(B\xi^2-d \log(n)/2)}.$$

On the interval  $\xi \in (\underline{\xi} \sqrt{d \log n}, \bar{\xi} \sqrt{n}/\log^2 n)$  for appropriate  $\underline{\xi}, \bar{\xi} > 0$ , we thus get (A.4). □

### Acknowledgements

The authors thank Markus Reiß and two anonymous referees for helpful comments. D.B. acknowledges the financial support from the Russian Academic Excellence Project “5-100” and from the Deutsche Forschungsgemeinschaft (DFG) through the SFB 823 “Statistical modelling of nonlinear dynamic processes”. M.T. gratefully acknowledges the financial support by the DFG research fellowship TR 1349/1-1. A part of this paper has been written while M.T. was affiliated to the Université Paris-Dauphine.

## References

- [1] A. Agarwal, S. Negahban and M. J. Wainwright. Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. *Ann. Statist.* **40** (2) (2012) 1171–1197. [MR2985947](#)
- [2] Y. Ait-Sahalia and P. A. Mykland. The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* **71** (2) (2003) 483–549. [MR1958137](#)
- [3] O. E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance Stoch.* **2** (1) (1997) 41–68. [MR1804664](#)
- [4] O. E. Barndorff-Nielsen and A. N. Shiryaev. *Change of Time and Change of Measure*. World Scientific, Singapore, 2010. [MR3363697](#)
- [5] I. V. Basawa and P. J. Brockwell. Non-parametric estimation for non-decreasing Lévy processes. *J. R. Stat. Soc., B* **44** (2) (1982) 262–269. [MR0676217](#)
- [6] D. Belomestny. Statistical inference for time-changed Lévy processes via composite characteristic function estimation. *Ann. Statist.* **39** (4) (2011) 2205–2242. [MR2893866](#)
- [7] D. Belomestny, F. Comte, V. Genon-Catalot, H. Masuda and M. Reiß. *Lévy Matters. IV: Estimation for Discretely Observed Lévy Processes. Lecture Notes in Mathematics* **2128**. Springer, Cham, 2015. [MR3364253](#)
- [8] D. Belomestny and V. Panov. Estimation of the activity of jumps in time-changed Lévy models. *Electron. J. Stat.* **7** (2013) 2970–3003. [MR3151759](#)
- [9] D. Belomestny and M. Reiß. Spectral calibration of exponential Lévy models. *Finance Stoch.* **10** (4) (2006) 449–474. [MR2276314](#)
- [10] M. Bibinger, N. Hautsch, P. Malec and M. Reiss. Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *Ann. Statist.* **42** (4) (2014) 80–114. [MR3226158](#)
- [11] P. J. Bickel and E. Levina. Covariance regularization by thresholding. *Ann. Statist.* **36** (6) (2008) 2577–2604. [MR2485008](#)
- [12] A. Bücher and M. Vetter. Nonparametric inference on Lévy measures and copulas. *Ann. Statist.* **41** (3) (2013) 1485–1515. [MR3113819](#)
- [13] A. Bull. Estimating time-changes in noisy Lévy models. *Ann. Statist.* **42** (5) (2014) 2026–2057. [MR3262476](#)
- [14] P. Carr, H. Geman, D. B. Madan and M. Yor. Stochastic volatility for Lévy processes. *Math. Finance* **13** (3) (2003) 345–382. [MR1995283](#)
- [15] J. Chorowski and M. Trabs. Spectral estimation for diffusions with random sampling times. *Stochastic Process. Appl.* **126** (10) (2016) 2976–3008. [MR3542623](#)
- [16] I. Dattner, M. Reiß, M. Trabs et al. Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli* **22** (1) (2016) 143–192. [MR3449779](#)
- [17] J. Fan, Y. Liao and M. Mincheva. High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.* **39** (6) (2011) 3320–3356. [MR3012410](#)
- [18] J. Fan, Y. Liao and M. Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** (4) (2013) 603–680. With 33 discussions by 57 authors and a reply by Fan, Liao and Mincheva. [MR3091653](#)
- [19] J. E. Figueroa-López. Nonparametric estimation of time-changed Lévy models under high-frequency data. *Adv. in Appl. Probab.* **41** (4) (2009) 1161–1188. [MR2663241](#)
- [20] J. Jacod and M. Podolskij. A test for the rank of the volatility process: The random perturbation approach. *Ann. Statist.* **41** (5) (2013) 2391–2427. [MR3127870](#)
- [21] J. Jacod and M. Reiß. A remark on the rates of convergence for integrated volatility estimation in the presence of jumps. *Ann. Statist.* **42** (3) (2014) 1131–1144. [MR3224283](#)
- [22] J. Kappus. Adaptive nonparametric estimation for Lévy processes observed at low frequency. *Stochastic Process. Appl.* **124** (1) (2014) 730–758. [MR3131312](#)
- [23] J. Kappus. Nonparametric estimation for irregularly sampled Lévy processes. *Stat. Inference Stoch. Process.* **21** (1) (2015) 141–167. [MR3769835](#)
- [24] V. Koltchinskii, K. Lounici and A. B. Tsybakov. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.* **39** (5) (2011) 2302–2329. [MR2906869](#)
- [25] P. Massart. *Concentration Inequalities and Model Selection. Lecture Notes in Mathematics* **1896**. Springer, Berlin, 2007. [MR2319879](#)
- [26] H. Masuda. Ergodicity and exponential  $\beta$ -mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.* **117** (1) (2007) 35–56. [MR2287102](#)
- [27] F. Merlevède, M. Peligrad and E. Rio. Bernstein inequality and moderate deviations under strong mixing conditions. In *High Dimensional Probability V: The Luminy Volume* 273–292. *Inst. Math. Stat. Collect.* **5**. Inst. Math. Statist., Beachwood, OH, 2009. [MR2797953](#)
- [28] I. Monroe. Processes that can be embedded in Brownian motion. *Ann. Probab.* **6** (1978) 42–56. [MR0455113](#)
- [29] S. Negahban and M. J. Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.* **39** (2) (2011) 1069–1097. [MR2816348](#)
- [30] M. H. Neumann and M. Reiß. Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15** (1) (2009) 223–248. [MR2546805](#)
- [31] R. Nickl, M. Reiß, J. Söhl and M. Trabs. High-frequency Donsker theorems for Lévy measures. *Probab. Theory Related Fields* **164** (1–2) (2015) 61–108. doi:10.1007/s00440-014-0607-3. [MR3449386](#)
- [32] B. Recht. A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** (2011) 3413–3430. [MR2877360](#)
- [33] P. Rigollet and A. B. Tsybakov. Comment: “Minimax estimation of large covariance matrices under  $\ell_1$ -norm”. *Statist. Sinica* **22** (4) (2012) 1358–1367. [MR3027087](#)
- [34] A. Rohde and A. B. Tsybakov. Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** (2) (2011) 887–930. [MR2816342](#)
- [35] H. Rubin and H. Tucker. Estimating the parameters of a differential process. *Ann. Math. Stat.* **30** (1959) 641–658. [MR0110174](#)
- [36] K. Sato. *Lévy Processes and Infinitely Divisible Distributions. Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge, 2013. Translated from the 1990 Japanese original. Revised edition of the 1999 English translation. [MR3185174](#)

- [37] M. Tao, Y. Wang, Q. Yao and J. Zou. Large volatility matrix inference via combining low-frequency and high-frequency approaches. *J. Amer. Statist. Assoc.* **106** (495) (2011) 1025–1040. [MR2894761](#)
- [38] M. Tao, Y. Wang and H. H. Zhou. Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors. *Ann. Statist.* **41** (4) (2013) 1816–1864. [MR3127850](#)
- [39] A. B. Tsybakov. *Introduction to Nonparametric Estimation. Springer Series in Statistics.* Springer, New York, 2009. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. [MR2724359](#)
- [40] A. W. van der Vaart. *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics 3.* Cambridge Univ. Press, Cambridge, 1998. [MR1652247](#)
- [41] A. Veraart and M. Winkel. Time change. In *Encyclopedia of Quantitative Finance.* R. Cont (Ed.). Wiley, New York, 2010. DOI:10.1002/9780470061602.eqf19026.
- [42] Y. Wang and J. Zou. Vast volatility matrix estimation for high-frequency financial data. *Ann. Statist.* **38** (2) (2010) 943–978. [MR2604708](#)
- [43] G. A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra Appl.* **170** (1992) 33–45. [MR1160950](#)
- [44] R. Yokoyama. Moment bounds for stationary mixing sequences. *Z. Wahrsch. Verw. Gebiete* **52** (1) (1980) 45–57. [MR0568258](#)