

Maximal inequalities and some applications*

Franziska Kühn

*DB Regio AG, Richard-Wagner-Straße 1,
04109 Leipzig, Germany*
e-mail: franziska.fr.kuehn@deutschebahn.com

and

René L. Schilling

*Institut für Mathematische Stochastik, Fakultät Mathematik
Technische Universität Dresden, 01062 Dresden, Germany*
e-mail: rene.schilling@tu-dresden.de

Abstract: A maximal inequality is an inequality which involves the (absolute) supremum $\sup_{s \leq t} |X_s|$ or the running maximum $\sup_{s \leq t} X_s$ of a stochastic process $(X_t)_{t \geq 0}$. We discuss maximal inequalities for several classes of stochastic processes with values in an Euclidean space: Martingales, Lévy processes, Lévy-type – including Feller processes, (compound) pseudo Poisson processes, stable-like processes and solutions to SDEs driven by a Lévy process –, strong Markov processes and Gaussian processes. Using the Burkholder–Davis–Gundy inequalities we also discuss some relations between maximal estimates in probability and the Hardy–Littlewood maximal functions from analysis.

MSC2020 subject classifications: Primary 60E15; secondary 60G15, 60G44, 60G51, 60G53, 60J25, 42B25, 42A61.

Keywords and phrases: Maximal inequality, maximal function, moment estimate, tail estimate, concentration function, Burkholder–Davis–Gundy inequality, Doob’s maximal inequality, good- λ inequality, Hardy–Littlewood maximal function, martingale inequality, Gaussian process, Lévy process, Feller process, Lévy-type process, Markov process, stochastic differential equation.

Received April 2022.

Contents

1	Introduction	383
2	Preliminaries	385
2.1	Censored moments of infinitely divisible random variables	385
2.2	Concentration functions	392
2.3	Good- λ Inequalities	398

arXiv: [2204.04690](https://arxiv.org/abs/2204.04690)

*Financial support through the DFG-NCN Beethoven Classic 3 project SCHI419/11-1 & NCN 2018/31/G/ST1/02252 is gratefully acknowledged.

3	Some general principles	404
4	Martingales	413
4.1	Doob's maximal inequalities	413
4.2	Khintchine and Burkholder–Davis–Gundy inequalities	420
5	Lévy processes	433
5.1	Background knowledge on Lévy processes	433
5.2	Reflection-type maximal inequalities	436
5.3	Generalized moments of Lévy processes	444
6	Lévy-type processes	449
6.1	Getting started: Basics on Lévy-type processes	450
6.2	Maximal tail estimates for Lévy-type processes	453
7	Markov processes	463
8	Gaussian processes	466
8.1	Concentration of measure methods	467
8.2	Metric entropy methods	470
	Index of notation	473
	Index	474
	Acknowledgments	475
	References	475

1. Introduction

A *maximal inequality* is, quite generally, an inequality that involves the running maximum $M_t := \sup_{s \leq t} X_s$ or the absolute supremum $X_t^* := \sup_{s \leq t} |X_s|$ (and $M := M_\infty$, $X^* := X_\infty^*$) of a stochastic process $(X_t)_{t \geq 0}$. Frequently, this happens in the form of a *tail estimate*, i.e. estimates for $\mathbb{P}(X_t^* > \lambda)$, $\mathbb{P}(M_t \leq \lambda)$ etc. or as a *generalized moment*, i.e. bounds of $\mathbb{E}[g(X_t^*)]$ or $\mathbb{E}[\sup_{s \leq t} g(X_s)]$ or in similar expressions.

There are many situations where maximal inequalities appear naturally, for example when studying the growth and asymptotics of $t \mapsto X_t$ as $t \rightarrow 0$ or $t \rightarrow \infty$, or the Hausdorff dimension of the range. One reason for the usefulness of maximal inequalities are inclusions of the type

$$\{X_t^* < r\} \subseteq \{\sigma_r > t\} \subseteq \{X_t^* \leq r\} \subseteq \{\sigma_r \geq t\},$$

which should be seen as a kind of duality between the maximum process X_t^* and the first exit time $\sigma_r := \inf\{s > 0 \mid |X_s| > r\}$ from the closed ball $\overline{B}_r(0)$ with centre 0 and radius $r > 0$; we assume that $t \mapsto X_t$ is càdlàg, i.e. right continuous with finite left hand limits. Each of these inclusions can be strict due to the presence of jumps.

Surprisingly, maximal tail inequalities are often easier to obtain than ordinary tail inequalities, or even heat kernel bounds. In many situations maximal tail inequalities may serve as a good substitute of a heat kernel estimate. In Remark 6.16 we give a (by no means complete) list of potential applications and some relevant references. Having said this, we would like to point out that for

many processes maximal (tail) estimates are equivalent to ordinary tail estimates: For Brownian motion this is just the reflection principle, for Lévy processes we can use Etemadi's inequality, see Theorem 5.11.

In this survey paper we discuss various forms of maximal inequalities for several classes of stochastic processes: Martingales, Lévy processes, Lévy-type & Feller processes, Markov processes and Gaussian processes – and in Chapters 4 through 8 each class is dealt with in a chapter on its own. We try to be as self-contained as sensibly possible, and each chapter starts with a brief introduction to the class of processes under investigation. On the one hand, this makes the chapters independent of each other and allows the novice and non-specialist to follow the presentation easily, on the other hand it allows us to fix the notation and to go in our presentation from the more special to the more general. This may not always be the most efficient approach, but it enables us to exploit the peculiarities of each class and to get clearer proofs. Chapters 2 and 3 are some kind of exception, as we collect in Chapter 2 the basic tools of the trade, which are frequently used but rarely systematically exposed. Chapter 3 explores some basic principles for rather general processes without assuming too much structure. We hope that this is a useful resource for researchers and offers, even for the specialist, some new vistas on known material.

Most results and proofs are (more or less well) known. Wherever possible, we give precise references on the source(s) and point out further reading. Frequently, the arguments are streamlined, sometimes with minor improvements, but it would be utterly presumptive to claim any originality but in the composition.

A few words on *what we did not (even try to) do* are in order: We do not cover infinite-dimensional state spaces and quite often we restrict ourselves to dimension $d = 1$, since the multivariate version is pretty obvious and leads only to notational inconveniences. Naturally, areas where we contributed ourselves to the research, e.g. Lévy and Lévy-type processes, are covered in greater depth than fields like Gaussian processes, where we give only a glimpse into some of the methods – this is partly also due to the fact that Gaussian processes require very different methods than the other classes of stochastic processes discussed here. Although it can be seen as a maximal tail estimate, the theory of large deviations is also excluded: this is too wide a field, and it is pretty difficult to surpass an exposition like Dembo & Zeitouni [29]. Finally, we do not discuss discrete time settings and refer instead to Lin & Bai [111], Csörgö & Révész [24] or Petrov [136, 137].

Notation Overall, we try to stick to standard notation, and more specialized symbols are introduced locally and when needed. Here we would like to highlight some conventions which are used throughout this paper: If $(X_t)_{t \geq 0}$ is a stochastic process, we write

$$X_t^* := \sup_{s \leq t} |X_s| \quad \text{and} \quad X^* := \sup_{s \geq 0} |X_s|.$$

If X and Y are random variables, we denote by $\mathbb{E}X$ and $\mathbb{V}X$ the expectation and the variance of X ; by $X \sim Y$ we indicate that X and Y have the same

probability distribution.

Throughout, we assume the *usual conditions* on the filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F}_t)$: The underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is complete (i.e. every subset of a \mathbb{P} -null set is measurable) and the filtration $(\mathcal{F}_t)_{t \geq 0}$ is right-continuous (i.e. $\mathcal{F}_t = \mathcal{F}_{t+} = \bigcap_{u > t} \mathcal{F}_u$) and \mathcal{F}_0 contains all \mathbb{P} -null sets. This convention ensures, in particular, that the notions of optional time and stopping time coincide, and that first entrance times and first hitting times of Borel sets are stopping times.

The Fourier transform and its inverse are denoted by

$$\widehat{f}(\xi) := (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\xi \cdot y} f(y) dy \quad \text{and} \quad \check{g}(y) := \int_{\mathbb{R}^d} e^{i\eta \cdot y} g(\eta) d\eta,$$

respectively; this convention allows us to identify the inverse Fourier transform with the characteristic function of a random variable, and to stay compatible with most formulae used in harmonic analysis and the theory of pseudo differential operators.

We use ‘positive’, ‘negative’, ‘monotone’, ‘increasing’ etc. always in the non-strict sense; the natural numbers \mathbb{N} are $\{1, 2, 3, \dots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. As usual, $a \wedge b$ and $a \vee b$ denote the minimum and the maximum of $a, b \in \mathbb{R}$. With $f \asymp g$ or $f(t) \asymp g(t)$ we indicate that there are constants $0 < c \leq C < \infty$ such that $cf(t) \leq g(t) \leq Cf(t)$ (for a given range of t 's). The positive and negative part of a function is $f^+ := f \vee 0$ and $f^- := -(f \wedge 0)$. By $\lfloor x \rfloor$ we denote the largest integer $n \in \mathbb{Z}$ such that $n \leq x$. We write $B_r(x)$ for the open ball with centre x and radius $r > 0$. Finally, we agree that $\inf \emptyset := \infty$.

An **index of notation** is included on page 473 at the end of the paper.

2. Preliminaries

2.1. Censored moments of infinitely divisible random variables

We restrict ourselves to the one-dimensional setting, since the extension to higher dimensions is often straightforward.

Infinite divisibility and characteristic exponents Recall that a random variable X is *infinitely divisible*, if $X \sim X_1 + \dots + X_n$ for any $n \in \mathbb{N}$ where X_1, \dots, X_n are iid copies of some random variable $X^{(n)}$. It is easy to see that X is infinitely divisible, if the characteristic function $\phi_X(\xi) = \mathbb{E}e^{i\xi X}$ satisfies $\phi_X(\xi) = \phi_{(n)}(\xi)^n$ where $\phi_{(n)}(\xi)$ is itself a characteristic function. Since the characteristic function determines the law of a random variable, we have $\phi_{(n)} = \phi_{X^{(n)}}$ for the random variable $X^{(n)}$ from above. A classical result of Lévy and Khintchine shows that X is infinitely divisible if, and only if, $\phi_X(\xi) = \exp(-\psi_X(\xi))$ with the *characteristic exponent* given by the *Lévy–Khintchine formula*

$$\psi_X(\xi) = -ib\xi + \frac{1}{2}\sigma^2\xi^2 + \int_{y \neq 0} (1 - e^{iy\xi} + iy\xi\mathbf{1}_{(0,1)}(|y|)) \nu(dy). \quad (2.1)$$

The Lévy triplet (b, σ^2, ν) – comprising $b \in \mathbb{R}$, $\sigma \geq 0$ and a Borel measure ν on $\mathbb{R} \setminus \{0\}$ such that $\int_{0 < |y| < 1} y^2 \nu(dy) + \int_{|y| \geq 1} \nu(dy) < \infty$ – determines ψ_X , hence ϕ_X and X , uniquely. Note that $2 \operatorname{Re} \psi_X(\xi)$ is the characteristic exponent of the symmetrization of X , i.e. $X - X'$ where X' is an iid copy of X . We refer to Sato [143] as standard reference, also for the connection of infinitely divisible random variables and Lévy processes.

A further characterization of the exponent ψ_X is to say that it is a *continuous negative definite function* with $\psi_X(0) = 0$; negative definite means that the matrix $(\psi(\xi_i) + \overline{\psi(\xi_j)} - \psi(\xi_i - \xi_j))_{i,j}$ is for any $n \in \mathbb{N}$ and $\xi_1, \dots, \xi_n \in \mathbb{R}$ positive hermitian.

We remark that, in particular, $\operatorname{Re} \psi_X(\xi) \geq 0$ and $\xi \mapsto \sqrt{|\psi_X(\xi)|}$ is subadditive, i.e. $\sqrt{|\psi_X(\xi + \eta)|} \leq \sqrt{|\psi_X(\xi)|} + \sqrt{|\psi_X(\eta)|}$ for all $\xi, \eta \in \mathbb{R}$. Standard references on (continuous) negative definite functions are Berg & Forst [9] and Jacob [66, Vol. 1].

We will also need the following *maximum functions* which describe the maximum of $|\psi_X(\xi)|$ and $\operatorname{Re} \psi_X(\xi)$ on balls with radius $\ell > 0$:

$$\psi_X^*(\ell) := \sup_{|\xi| \leq \ell} \operatorname{Re} \psi_X(\xi) \quad \text{and} \quad |\psi_X|^*(\ell) := \sup_{|\xi| \leq \ell} |\psi_X(\xi)|. \quad (2.2)$$

These maximum functions are also important for estimates of the transition densities of Lévy and Lévy-type processes, see e.g. Knopova & Kulik [74, 75], Knopova [73], Grzywny [51] or Grzywny & Szczypkowski [52].

Censored second moments Thinking of variances of random variables, it is often useful to require finite second moments – but this is not always the case. The following construction will serve as a partial substitute when there are no finite second moments. If μ is a measure on \mathbb{R} , we define its *truncated second moment*, the *tail* and the *censored second moment*, respectively, as

$$\begin{aligned} K(\mu; \ell) &:= \frac{1}{\ell^2} \int_{0 \leq |y| \leq \ell} y^2 \mu(dy), & G(\mu; \ell) &:= \int_{|y| > \ell} \mu(dy), \\ \text{and } D(\mu; \ell) &:= K(\mu; \ell) + G(\mu; \ell), & \ell &> 0. \end{aligned} \quad (2.3)$$

It is not hard to see that $D(\mu; \ell) \leq D(\mu; \ell_0)$ for all $0 < \ell_0 < \ell$.

For an infinitely divisible random variable with Lévy triplet (b, σ^2, ν) we (formally) insert $\nu_0(dy) = \sigma^2 y^{-2} \delta_0(dy) + \nu(dy)$ instead of μ , which means that

$$\begin{aligned} D(\nu_0; \ell) &= \frac{\sigma^2}{\ell^2} + \frac{1}{\ell^2} \int_{0 < |y| \leq \ell} y^2 \nu(dy) + \int_{|y| > \ell} \nu(dy) \\ &= \frac{\sigma^2}{\ell^2} + \int_{y \neq 0} \min \left\{ 1, \frac{y^2}{\ell^2} \right\} \nu(dy). \end{aligned} \quad (2.4)$$

Scholium 2.1 (probabilistic meaning of truncated moments). The interpretation of the truncated and censored moments in terms of moments of the random variable X is obvious if μ is the corresponding probability distribution. If X

is an infinitely divisible random variable with exponent ψ_X and Lévy triplet (b, σ^2, ν) , the truncated moments are moments of the characteristic triplet, and the probabilistic interpretation in terms of moments of X is less straightforward.

In order to appreciate the construction fully, it is useful to introduce yet another truncated moment that captures the non-symmetry of X , as suggested by Pruitt [140]:

$$\begin{aligned}
 M(b, \nu; \ell) &= \frac{1}{\ell} \left| b + \int_{|y| \leq \ell} y \mathbf{1}_{[1, \infty)}(|y|) \nu(dy) - \int_{|y| > \ell} y \mathbf{1}_{(0, 1)}(|y|) \nu(dy) \right| \\
 &= \frac{1}{\ell} \left| b + \left\{ \int_1^\ell + \int_{-\ell}^{-1} \right\} y \nu(dy) \right|; \tag{2.5}
 \end{aligned}$$

here we use the ‘Riemannian’ convention that $\int_a^b = -\int_b^a$ if $a > b$.

For fixed $\ell > 0$ we split the random variable X into a sum of two infinitely divisible random variables $Y + J$ where $Y = Y(\ell)$ and $J = J(\ell)$ are determined through their exponents:

$$\begin{aligned}
 \psi_J(\xi) &= \int_{|y| > \ell} (1 - e^{iy\xi}) \nu(dy) \\
 \psi_Y(\xi) &= -i\xi \left(b - \int_{|y| > \ell} y \mathbf{1}_{(0, 1)}(|y|) \nu(dy) \right) + \frac{1}{2} \sigma^2 \xi^2 \\
 &\quad + \int_{0 < |y| \leq \ell} (1 - e^{iy\xi} + iy\xi \mathbf{1}_{(0, 1)}(|y|)) \nu(dy).
 \end{aligned}$$

We can work out the moments of X , Y and J by differentiating the respective characteristic functions, e.g.

$$\mathbb{E}[Y] = i\psi'_Y(0), \quad \mathbb{E}[Y^2] = \psi''_Y(0) - \psi'^2_Y(0) \quad \text{and} \quad \mathbb{V}[Y] = \psi''_Y(0).$$

A short calculation reveals that $M(b, \nu; \ell)$ and $K(\nu_0; \ell)$ are the moments of the ‘truncated’ random variable $Y = Y(\ell)$:

$$|\mathbb{E}[\ell^{-1}Y(\ell)]| = M(b, \nu; \ell) \quad \text{and} \quad \mathbb{V}[\ell^{-1}Y(\ell)] = K(\nu_0; \ell).$$

We will now connect $D(\nu_0; \cdot)$ with the maximum exponent ψ^* . To do so, we need some preparations. A direct calculation, see e.g. Schilling & Schnurr [153, Lem. 6.1] yields:

Lemma 2.2. *The function $x^2/(1+x^2)$, $x \in \mathbb{R}$, is a characteristic exponent and its Lévy–Khintchine formula has the form*

$$\frac{x^2}{1+x^2} = \int_{\xi \neq 0} (1 - \cos(x\xi)) g(\xi) d\xi \tag{2.6}$$

where $g(\xi) = \frac{1}{2} \int_0^\infty (2\pi\lambda)^{-1/2} e^{-\xi^2/(2\lambda)} e^{-\lambda/2} d\lambda$ is an integrable function which has absolute moments of any order.

The next lemma contains a (by now) standard estimate. The short proofs are adapted from Schilling & Schnurr [153, Lem. 6.1] and Grzywny [51, Lem. 4].

Lemma 2.3. *Let $\psi: \mathbb{R} \rightarrow \mathbb{C}$ be a continuous negative definite function. Then*

$$|\psi(\xi)| \leq 2 \sup_{|\eta| \leq 1} |\psi(\eta)| (1 + |\xi|^2), \quad \xi \in \mathbb{R}. \quad (2.7)$$

If Ψ^* denotes one of the maximum functions ψ^* or $|\psi|^*$ from (2.2), then

$$\frac{1}{2} \frac{r^2}{1+r^2} \Psi^*(\ell) \leq \Psi^*(r\ell) \leq 2(1+r^2)\Psi^*(\ell), \quad r, \ell > 0. \quad (2.8)$$

Proof. Fix ξ and write it in the form $\xi = \eta + n\xi_0$ where ξ, η, ξ_0 have the same sign, $|\eta| < 1$, $|\xi_0| = 1$ and $n \in \mathbb{N}$. Using the subadditivity of $\sqrt{|\psi(\xi)|}$ we get

$$\sqrt{|\psi(\xi)|} \leq \sqrt{|\psi(\eta)|} + n\sqrt{|\psi(\xi_0)|} \leq (1 + |\xi|) \sup_{|\eta| \leq 1} \sqrt{|\psi(\eta)|},$$

and the first part of the lemma follows.

We will show the second part only for $\psi^*(\ell)$. Since $\operatorname{Re} \psi(\xi)$ is also continuous and negative definite, the upper estimate follows from the first part applied to $\xi \mapsto \operatorname{Re} \psi(\ell\xi)$:

$$\psi^*(r\ell) = \sup_{|\xi| \leq r} \operatorname{Re} \psi(\ell\xi) \leq 2 \sup_{|\eta| \leq 1} \operatorname{Re} \psi(\ell\eta) (1 + r^2) = 2(1 + r^2) \psi^*(\ell), \quad r, \ell > 0.$$

From this we get at once the lower bound, taking first $1/r$ instead of r , and then changing $\ell \rightsquigarrow \ell r$ in the resulting estimate. \square

The following estimate from Schilling [144, Rem. 4.8] is an improvement of results by Le Cam [100, Prop. 5] and Esseen [39, Lem. 2.1]. We follow again the presentation of Grzywny [51, Lem. 4].

Theorem 2.4. *Let $\psi = \psi_X$ be the characteristic exponent of an infinitely divisible real random variable X and denote by $D(\nu_0; \ell)$ the censored second moment of the measure $\nu_0 := \sigma^2 \delta_0 + \nu$; here σ^2 and ν are from the Lévy triplet of ψ . Then the following estimates are satisfied with an absolute constant $c < 1$.¹*

$$cD(\nu_0; \ell) \leq \psi^*(\ell^{-1}) \leq 2D(\nu_0; \ell). \quad (2.9)$$

Proof. Since the supremum is subadditive, we have

$$\psi^*(\ell^{-1}) \leq \sup_{|\xi| \leq 1/\ell} \frac{1}{2} \sigma^2 \xi^2 + \sup_{|\xi| \leq 1/\ell} \int_{y \neq 0} (1 - \cos(y\xi)) \nu(dy) \leq 2\psi^*(\ell^{-1}).$$

Therefore, it is enough to prove that for some constant c

¹This result remains valid if X takes values in \mathbb{R}^d ; then we can take $c = 1/(8(1 + 2d))$, cf. Grzywny [51, p. 9].

$$\begin{aligned} c \int_{y \neq 0} \min \left\{ 1, \frac{y^2}{\ell^2} \right\} \nu(dy) &\leq \sup_{|\xi| \leq 1/\ell} \int_{y \neq 0} (1 - \cos(y\xi)) \nu(dy) \\ &\leq 2 \int_{y \neq 0} \min \left\{ 1, \frac{y^2}{\ell^2} \right\} \nu(dy). \end{aligned}$$

Thus, without loss of generality, we may assume $\sigma = 0$. Combining the elementary estimate $\min\{1, y^2\} \leq 2y^2/(1 + y^2)$ with Lemma 2.2 and Tonelli's theorem shows

$$\begin{aligned} \int_{y \neq 0} \min \left\{ 1, \frac{y^2}{\ell^2} \right\} \nu(dy) &\leq 2 \int_{y \neq 0} \frac{y^2 \ell^{-2}}{1 + y^2 \ell^{-2}} \nu(dy) \\ &\leq 2 \int \int_{y \neq 0} \left(1 - \cos \frac{\xi y}{\ell} \right) \nu(dy) g(\xi) d\xi \\ &= 2 \int \operatorname{Re} \psi(\xi/\ell) g(\xi) d\xi. \end{aligned}$$

Since $\operatorname{Re} \psi$ is continuous and negative definite, we can use Lemma 2.3 to get

$$\int_{y \neq 0} \min \left\{ 1, \frac{y^2}{\ell^2} \right\} \nu(dy) \leq 4 \sup_{|\eta| \leq 1/\ell} \operatorname{Re} \psi(\eta) \int (1 + \xi^2) g(\xi) d\xi = c\psi^*(\ell^{-1}).$$

Using the estimate $1 - \cos u = 2 \sin^2(\frac{1}{2}u) \leq 2 \min\{1, u^2\}$ we see for all $|\xi| \leq 1/\ell$

$$\operatorname{Re} \psi(\xi) \leq \sup_{|\xi| \leq 1/\ell} 2 \int_{y \neq 0} \min \{1, y^2 \xi^2\} \nu(dy) = 2 \int_{y \neq 0} \min \left\{ 1, \frac{y^2}{\ell^2} \right\} \nu(dy)$$

finishing the proof. □

There is a similar estimate for $|\psi|^*(\ell^{-1})$ that takes into account the asymmetry of the random variable X .

Theorem 2.5. *Let $\psi = \psi_X$ be the characteristic exponent of an infinitely divisible real random variable X with Lévy triplet (b, σ^2, ν) and denote by $D(\nu_0; \ell) = K(\nu_0; \ell) + G(\nu; \ell)$ and $M(b, \nu; \ell)$ the censored and truncated moments of the measure $\nu_0 := \sigma^2 \delta_0 + \nu$. Then the following estimates are satisfied with absolute constants:²*

$$|\psi|^*(\ell^{-1}) \asymp K(\nu_0; \ell) + G(\nu; \ell) + M(b, \nu; \ell). \tag{2.10}$$

Proof. Since we know already the result for the real part from Theorem 2.4, it is enough to estimate the imaginary part. Recall that

$$-\operatorname{Im} \psi(\xi) = b\xi + \int_{y \neq 0} (\sin(y\xi) - y\xi \mathbf{1}_{(0,1)}(|y|)) \nu(dy).$$

²In the d -dimensional analogue of this result, the constants depend only on the space dimension d .

We show the upper bound for $\ell < 1$, the case $\ell \geq 1$ differs only by a sign within the definition of $M(b, \nu; \ell)$. Fix $0 < \ell < 1$ and take any $|\xi| < 1/\ell$.

Using the elementary estimate $|\sin(x) - x| \leq \frac{1}{6}|x|^3$, we get

$$\begin{aligned} & |\operatorname{Im} \psi(\xi)| \\ &= \left| \int_{0 < |y| \leq \ell} (\sin(y\xi) - y\xi) \nu(dy) + b\xi - \int_{\ell < |y| < 1} y\xi \nu(dy) + \int_{|y| > \ell} \sin(y\xi) \nu(dy) \right| \\ &\leq \frac{1}{6}K(\nu_0; \ell) + M(b, \nu; \ell) + G(\nu; \ell). \end{aligned}$$

Since $|\xi| \leq 1/\ell$ is arbitrary, the upper bound follows if we use the upper bound in (2.9).

The lower bound can be shown in a similar way, using the lower triangle inequality and the observation that

$$|\sin x - x| = \left| \int_0^x (1 - \cos t) dt \right| \leq |x|(1 - \cos x) \leq (1 - \cos x), \quad |x| \leq 1.$$

Thus,

$$\begin{aligned} & |\operatorname{Im} \psi(\xi)| \\ &= \left| \int_{0 < |y| \leq \ell} (\sin(y\xi) - y\xi) \nu(dy) + b\xi - \int_{\ell < |y| < 1} y\xi \nu(dy) + \int_{|y| > \ell} \sin(y\xi) \nu(dy) \right| \\ &\geq M(b, \nu; \ell) - G(\nu; \ell) - \int_{0 < |y| \leq \ell} (1 - \cos(y\xi)) \nu(dy) \\ &\geq M(b, \nu; \ell) - G(\nu; \ell) - \psi^*(\ell^{-1}). \end{aligned}$$

So we get $M(b, \nu; \ell) \leq 2(|\psi|^*(\ell^{-1}) + G(\nu; \ell))$, and the lower estimate follows. \square

Comparability of truncated and censored second moments We want to compare the truncated moments $K(\mu; \ell)$ and the censored moments $D(\mu; \ell)$. Recall that

$$K(\mu; \ell) = \int_{|y| \leq \ell} \frac{y^2}{\ell^2} \mu(dy) \quad \text{and} \quad D(\mu; \ell) = \int_{\mathbb{R}} \min \left\{ 1, \frac{y^2}{\ell^2} \right\} \mu(dy).$$

Clearly, $K(\mu; \ell) \leq D(\mu; \ell)$. Let us introduce the condition

$$D(\mu; \ell) \leq \beta K(\mu; \ell) \quad \text{for some } \beta \geq 1 \text{ and all } \ell > 0. \quad (2.11)$$

We have $\frac{d}{dx} \min\{1, x\} = \mathbf{1}_{[0,1]}(x)$ Lebesgue a.e. on $[0, \infty)$. Therefore,

$$\frac{d}{d\ell} D(\mu; \ell) = - \int_{\mathbb{R}} \mathbf{1}_{[0,1]} \left(\frac{y^2}{\ell^2} \right) \frac{2y^2}{\ell^3} \mu(dy) = -\frac{2}{\ell} K(\mu; \ell) \quad \text{Leb. a.e. for } \ell > 0.$$

If we assume (2.11), we get for almost all $\ell > 0$

$$-\frac{2}{\ell} \leq \frac{D'(\mu; \ell)}{D(\mu; \ell)} \leq -\frac{2}{\beta\ell},$$

and, by integration,

$$\frac{1}{\ell^2} \leq \frac{D(\mu, \ell)}{D(\mu, 1)} \leq \frac{1}{\ell^{2/\beta}} \quad \text{if } \ell \geq 1, \text{ and } \quad \frac{1}{\ell^{2/\beta}} \leq \frac{D(\mu, \ell)}{D(\mu, 1)} \leq \frac{1}{\ell^2} \quad \text{if } \ell \leq 1.$$

If we combine this with Theorem 2.4 and use $\mu = \nu_0 := \sigma^2\delta_0 + \nu$, we finally arrive at the following result which was first observed by Knopova & Kulik [74, Lem. 3.1].

Theorem 2.6. *Let X be an infinitely divisible random variable with characteristic exponent ψ , maximum function ψ^* and Lévy triplet (b, σ^2, ν) .*

If $\int_{y \neq 0} \min\{\ell^2, y^2\} \nu(dy) \leq \beta \int_{|y| \leq \ell} y^2 \nu(dy)$ for some $\beta \geq 1$, then

$$\psi^*(r) \geq cr^{2/\beta} \quad \text{if } r \geq 1, \text{ and } \quad \psi^*(r) \leq Cr^{2/\beta} \quad \text{if } r \leq 1. \quad (2.12)$$

The sector condition Let us briefly discuss under which circumstances one can use ψ^* and $|\psi|^*$ interchangeably. Clearly,

$$\operatorname{Re} \psi(\xi) \leq |\psi(\xi)|, \quad \xi \in \mathbb{R} \quad \text{and} \quad \psi^*(\ell) \leq |\psi|^*(\ell), \quad \ell > 0$$

are always satisfied. For the converse estimates one needs the so-called *sector condition*.

Lemma 2.7. *Let $\psi: \mathbb{R} \rightarrow \mathbb{C}$ be a continuous negative definite function. If the sector condition with sector constant $\kappa \in (0, \infty)$ holds, i.e.*

$$|\operatorname{Im} \psi(\xi)| \leq \kappa \operatorname{Re} \psi(\xi), \quad \xi \in \mathbb{R}, \quad (2.13)$$

then $\operatorname{Re} \psi$ and $|\psi|$ as well as ψ^ and $|\psi|^*$ are comparable. Conversely, the comparability of $\operatorname{Re} \psi$ and $|\psi|$ entails the sector condition for some suitable $\kappa > 0$.*

Proof. The assertion follows immediately from the estimates

$$\operatorname{Re} \psi(\xi) \leq |\psi(\xi)| \leq (1 + \kappa) \operatorname{Re} \psi(\xi).$$

which remain true if we take the supremum over all $|\xi| \leq \ell$. The converse follows then from $|\operatorname{Im} \psi(\xi)| \leq |\psi(\xi)|$ and the comparability of $\operatorname{Re} \psi(\xi)$ and $|\psi(\xi)|$. \square

The sector condition has a few further interesting consequences. First of all, see Berg & Forst [8], it is equivalent to saying that the bilinear form $(u, v) \mapsto \int \psi(\xi) \widehat{u}(\xi) \overline{\widehat{v}(\xi)} d\xi$, $u, v \in L^2(\mathbb{R}; dx)$, is a Dirichlet form – the Dirichlet form that uniquely characterizes the Lévy process associated with ψ , see Jacob [66, Chapter 4.7. 4.8]. Stochastically, it is a condition that ensures that the ‘true random character’ is not dominated by the ‘drift’ which is essentially governed by the imaginary part $\operatorname{Im} \psi(\xi)$.

2.2. Concentration functions

Lévy's concentration function Concentration functions were introduced by Lévy [106, Chapter III.16] in order to study the convergence behaviour of sums and sequences of random variables.

Definition 2.8 (Lévy). Let X be a real-valued random variable. *Lévy's concentration function* is the function $Q_X : [0, \infty) \rightarrow [0, 1]$ defined by

$$Q_X(\ell) := \sup_{x \in \mathbb{R}} \mathbb{P}(X \in [x, x + \ell]), \quad \ell \geq 0. \quad (2.14)$$

Sometimes, the generalized inverse $D_X(p) := \inf \{\ell \mid Q_X(\ell) > p\}$ of Q_X is used instead of Q_X ; it is usually called *dispersion* or *scattering function*. The standard reference for Lévy's concentration functions is Hengartner & Theodorescu [55], further material can be found in Petrov [137, §§1.5, 1.8, 2.5] and LeCam [101, Chapter 15.2]. Hengartner & Theodorescu also discuss multivariate versions of the concentration function. We will summarize the most important facts from these sources.

Properties 2.9 (of Lévy's concentration function). *Let X be a real-valued random variable, $F_X(x) = \mathbb{P}(X \leq x)$ the (right continuous) distribution function and Q_X the concentration function.*

1. Q_X is right continuous, increasing and satisfies $\lim_{\ell \rightarrow \infty} Q_X(\ell) = 1$. In particular, Q_X is the distribution function of some random variable $Y \geq 0$, i.e. $Q_X = F_Y$.
2. $Q_X(0) = \sup_{x \in \mathbb{R}} \mathbb{P}(X = x)$, and Q_X is continuous if, and only if, the distribution function F_X is continuous.
3. Replacing in (2.14) the closed interval $[x, x + \ell]$ by $(x, x + \ell)$, $(x, x + \ell]$ or $[x, x + \ell)$ gives $Q_X^-(\ell) := Q_X(\ell^-)$ if $\ell > 0$ and $Q_X^-(0) = 0$.
4. Q_X is subadditive: $Q_X(\ell + \ell') \leq Q_X(\ell) + Q_X(\ell')$ for all $\ell, \ell' \geq 0$.
5. Every subadditive distribution function of a random variable $Y \geq 0$ is also the concentration function of some random variable X .

Let us mention some typical examples of concentration functions.

- Example 2.10.**
1. If $X \sim \mathcal{N}(m, \sigma^2)$ is normal with mean $m \in \mathbb{R}$, standard deviation $\sigma \geq 0$, and distribution function Φ , then $Q_X(\ell) = 2\Phi(\ell/2\sigma) - 1$, $\ell \geq 0$.
 2. If $X \sim \lambda\pi^{-1}(\lambda^2 + (x - \alpha)^2)^{-1}$ is Cauchy distributed ($\lambda > 0$, $\alpha \in \mathbb{R}$), then $Q_X(\ell) = 2\pi^{-1} \arctan(\ell/2\lambda)$, $\ell \geq 0$.
 3. If $X \sim \mathcal{U}[a, b]$ is a uniform random variable on the interval $[a, b]$, then $Q_X(\ell) = \min \{\ell/(b - a); 1\}$, $\ell \geq 0$.
 4. If $X \sim Y + c$ or $X \sim -Y$, then $Q_X = Q_Y$, i.e. the concentration function is invariant under shifts and reflections.

The concentration function Q_X is often used for sums of independent random variables. In this connection the following lemma is important. The proof is a rather straightforward calculation based on the definition of the concentration function.

Lemma 2.11. *Let X and Y be two independent real-valued random variables. Then*

$$Q_X(t\ell) \cdot Q_Y((1-t)\ell) \leq Q_{X+Y}(\ell) \leq \min \{Q_X(\ell); Q_Y(\ell)\}, \quad \ell \geq 0, t \in [0, 1].$$

If we consider the symmetrization of $X^{\text{sy}} := X - X'$ where $Y = X'$ is an iid copy of X , then we get for $t = \frac{1}{2}$ that $Q_X(\ell)^2 \leq Q_{X^{\text{sy}}}(2\ell) \leq Q_X(2\ell)$.

We are most interested in the case where X is *infinitely divisible*, see Section 2.1. We write $\phi_X(\xi) = \exp(-\psi(\xi))$ for the characteristic function, with the characteristic exponent given by the Lévy–Khintchine formula (2.1). The following lemma is, essentially, a refinement of the classical Lévy truncation estimate. The key to its proof is the observation that

$$u(x) = \max \{0, 1 - |x|\} \xleftrightarrow[\text{(inverse) Fourier transform}]{\text{characteristic function}} \hat{u}(\xi) = \left(\frac{\sin(\xi/2)}{\xi/2}\right)^2$$

is a Fourier transform pair consisting of positive functions such that

$$\frac{1}{2} \mathbb{1}_{[-1/2, 1/2]} \leq u(x) \leq \mathbb{1}_{[-1, 1]}(x) \quad \text{and} \quad \left(\frac{95}{96}\right)^2 \mathbb{1}_{[-1/2, 1/2]}(\xi) \leq \hat{u}(\xi) \leq \mathbb{1}_{[-1, 1]}(\xi),$$

for all $x \in \mathbb{R}$ and $\xi \in [-1, 1]$, respectively; see Fig. 1.

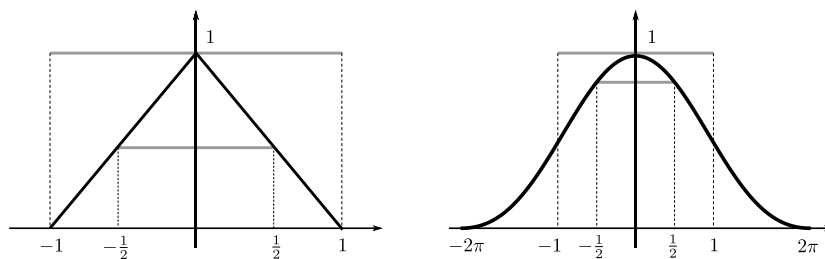


FIGURE 1. *The Fourier pair $u(x) = \max \{0, 1 - |x|\}$ (left) and $\hat{u}(\xi) = \left(\frac{\sin(\xi/2)}{\xi/2}\right)^2$ (right).*

Lemma 2.12 (Esseen [39]). *Let X be a real-valued random variable and denote by $\phi_X(\xi) = \mathbb{E}e^{i\xi X}$ its characteristic function. Then, for all $\ell, a > 0$*

$$\frac{95\ell}{256\pi(1+2a\ell)} \int_{-a}^a |\phi_X(\xi)|^2 d\xi \leq Q_X(\ell) \leq \left(\frac{96}{95}\right)^2 \max \{\ell, a^{-1}\} \int_{-a}^a |\phi_X(\xi)| d\xi.$$

We will now apply Lemma 2.12 to infinitely divisible random variables. Recall the definition of the truncated censored moment $D(\nu_0; \ell)$ where $\nu_0 = \sigma^2\delta_0 + \nu$, see (2.4) and the notation introduced in (2.3). The following result is due to Le Cam [100] and Esseen [39].

Theorem 2.13 (Le Cam; Esseen). *Let X be a (non-degenerate) infinitely divisible real random variable with characteristic exponent ψ and Lévy triplet*

(b, σ^2, ν) . Set $\nu_0 := \sigma^2 \delta_0 + \nu$. There are absolute constants $0 < c, C < \infty$ such that for all $\ell > 0$

$$c \min \left\{ 1, \frac{1}{\sqrt{K(\nu_0; \ell)}} \right\} e^{-4G(\nu; \ell)} \leq Q_X(\ell) \leq \frac{C}{\sqrt{D(\nu_0; \ell)}}. \tag{2.15}$$

Proof. We begin with an elementary estimate for the cosine. Since $\sin(u)$ is concave on $[0, \frac{\pi}{2}]$, we have $\sin(u) \geq \frac{\sin(\frac{\pi}{2})}{\frac{\pi}{2}} \cdot u \geq \frac{1}{2}u$, and so

$$1 - \cos t = \int_0^t \sin(u) \, du \geq \frac{1}{2} \int_0^t u \, du = \frac{1}{4} t^2, \quad t \in [0, \frac{\pi}{2}].$$

Since we want to use Lemma 2.12, we need to bound the characteristic function $|\phi_X(\xi)|$. Take $a = \ell^{-1}$ in the upper bound of Lemma 2.12 and use the Lévy–Khintchine formula (2.1) to get

$$\begin{aligned} \ell \int_{-1/\ell}^{1/\ell} |e^{-\psi_X(\xi)}| \, d\xi &= \int_{-1}^1 \exp[-\operatorname{Re} \psi_X(\xi/\ell)] \, d\xi \\ &= \int_{-1}^1 \exp \left[-\frac{\sigma^2 \xi^2}{2\ell^2} - \int_{|y| \leq \ell} \left(1 - \cos \frac{y\xi}{\ell}\right) \nu(dy) - \int_{|y| > \ell} \left(1 - \cos \frac{y\xi}{\ell}\right) \nu(dy) \right] \, d\xi \\ &\leq \int_{-1}^1 \exp \left[-\frac{1}{4} K(\nu_0; \ell) \xi^2 \right] \exp \left[- \int_{|y| > \ell} \left(1 - \cos \frac{y\xi}{\ell}\right) \nu(dy) \right] \, d\xi. \end{aligned}$$

Now we use Hölder’s inequality with

$$\frac{1}{p} = \frac{K(\nu_0; \ell)}{D(\nu_0; \ell)} \quad \text{and} \quad \frac{1}{q} = \frac{G(\nu; \ell)}{D(\nu_0; \ell)},$$

see (2.3) for the definition of K and G , and we obtain

$$\ell \int_{-1/\ell}^{1/\ell} |e^{-\psi_X(\xi)}| \, d\xi \leq (\text{I})^{1/p} \cdot (\text{II})^{1/q}.$$

The two factors can be estimated in the following way:

$$\text{I} = \int_{-1}^1 \exp \left[-\frac{1}{4} D(\nu_0; \ell) \xi^2 \right] \, d\xi \leq \frac{C_1}{\sqrt{D(\nu_0; \ell)}}.$$

If $\nu(|y| > \ell) = 0$, then $\text{II} = 1$, and we are done. Otherwise, we use the normalization $\int_A \cdots \, d\nu := \int_A \cdots \frac{d\nu}{\nu(A)}$, and Jensen’s inequality:

$$\begin{aligned} \text{II} &= \int_{-1}^1 \exp \left[-D(\nu_0; \ell) \int_{|y| > \ell} \left(1 - \cos \frac{y\xi}{\ell}\right) \nu(dy) \right] \, d\xi \\ &\leq \int_{-1}^1 \int_{|y| > \ell} \exp \left[-D(\nu_0; \ell) \left(1 - \cos \frac{y\xi}{\ell}\right) \right] \nu(dy) \, d\xi \\ &= 2 \int_{|y| > \ell} \int_0^{|y|/\ell} \exp[-D(\nu_0; \ell) (1 - \cos t)] \frac{dt}{|y|/\ell} \nu(dy). \end{aligned}$$

Since $1 - \cos t$ is periodic, symmetric (even) around $\frac{\pi}{2}$ in the interval $[0, \pi]$, and $|y|/\ell > 1$, we get

$$\Pi \leq 4 \int_{|y|>\ell} \frac{\lfloor \frac{|y|/\ell}{\pi} \rfloor + 1}{|y|/\ell} \int_0^{\pi/2} \exp \left[-D(\nu_0; \ell) \frac{1}{2} t^2 \right] dt \nu(dy) \leq \frac{C_2}{\sqrt{D(\nu_0; \ell)}}.$$

Since C_1, C_2 are absolute constants and $p^{-1} + q^{-1} = 1$, the upper bound in (2.15) follows.

For the lower bound in (2.15) we use the lower estimate in Lemma 2.12 with $a = \ell^{-1}$ and the elementary inequality $1 - \cos u \leq \frac{1}{2}u^2$.

$$\begin{aligned} Q_X(\ell) &\geq c\ell \int_{-1/\ell}^{1/\ell} \left| e^{-\psi_X(\xi)} \right|^2 d\xi \\ &= c \int_{-1}^1 \exp \left[-\frac{\sigma^2 \xi^2}{\ell^2} - 2 \int_{|y| \leq \ell} \left[1 - \cos \frac{y\xi}{\ell} \right] \nu(dy) - 2 \int_{|y| > \ell} \left[1 - \cos \frac{y\xi}{\ell} \right] \nu(dy) \right] d\xi \\ &\geq c \int_{-1}^1 \exp \left[-\xi^2 \left(\frac{\sigma^2}{\ell^2} + \frac{1}{\ell^2} \int_{|y| \leq \ell} y^2 \nu(dy) \right) - 4 \int_{|y| > \ell} \nu(dy) \right] d\xi \\ &\geq c \int_{-1}^1 e^{-\xi^2 K(\nu_0; \ell)} d\xi e^{-4G(\nu; \ell)} \\ &\geq c' \min \left\{ 1; \frac{1}{\sqrt{K(\nu_0; \ell)}} \right\} e^{-4G(\nu; \ell)}. \quad \square \end{aligned}$$

Remark 2.14. a) With a more careful argument based on Lemma 2.11 one can improve the lower bound in (2.15) and get $G(\nu, \ell)$ instead of $4G(\nu; \ell)$ in the exponent, see Hengartner & Theodorescu [55, pp. 53–54].

b) There is an interesting extension of the upper bound in (2.15) to arbitrary, not necessarily infinitely divisible, random variables X . The starting point is the elementary convexity estimate $(z + 1) \leq e^z, z \in \mathbb{R}$, which yields

$$e^z \geq 1 + z \iff -2x \leq -(1 - e^{-2x}) \iff \sqrt{e^{-2x}} \leq \sqrt{e^{-(1 - e^{-2x})}},$$

and this implies

$$|\phi_X(\xi)| \leq \exp \left[-\frac{1}{2} \left(1 - |\phi_X(\xi)|^2 \right) \right].$$

The right hand side is the characteristic function of a compound-Poisson type random variable, since the exponent is given by a Lévy–Khintchine formula:

$$\frac{1}{2} \left(1 - |\phi_X(\xi)|^2 \right) = \frac{1}{2} \int (1 - \cos(y\xi)) \rho * \tilde{\rho}(dy)$$

where $X \sim \rho$ and $-X \sim \tilde{\rho}$. The proof of Theorem 2.13 immediately gives

$$Q_X(\ell) \leq \frac{C}{\sqrt{D(\frac{1}{2}\rho * \tilde{\rho}; \ell)}}.$$

Remark 2.15. Combining the estimates from Theorem 2.13 with Theorem 2.4 gives at once the following estimates for an infinitely divisible random variable X with characteristic function ψ in terms of the censored second moment $D(\nu_0; \cdot)$, $\nu_0 = \sigma^2 \delta_0 + \nu$,

$$c \min \left\{ 1, \frac{1}{\sqrt{D(\nu_0; \ell)}} \right\} e^{-4D(\nu_0; \ell)} \leq Q_X(\ell) \leq \frac{C}{\sqrt{D(\nu_0; \ell)}}, \quad (2.16)$$

($0 < c < C < \infty$ are absolute constants) and the maximum function ψ^* :

$$c' \min \left\{ 1, \frac{1}{\sqrt{\psi^*(\ell^{-1})}} \right\} e^{-c' \psi^*(\ell^{-1})} \leq Q_X(\ell) \leq \frac{C'}{\sqrt{\psi^*(\ell^{-1})}} \quad (2.17)$$

($0 < c' < C'' < \infty$ and $c'' > 0$ are absolute constants).

Milman's concentration function A different type of concentration function was introduced by Amir & Milman [1] to study isoperimetric inequalities using the *concentration of measure phenomenon*. This approach goes back to Milman [128] who proved Dvoretzky's theorem on spherical sections of convex sets using an estimate for the volume of the n -dimensional spherical cap due to Lévy [105, pp. 260–285], see also Lévy [107, 209–234]: if S is a cap of the sphere \mathbb{S}^{n-1} in \mathbb{R}^n and $A \subseteq \mathbb{S}^{n-1}$ another set having the same Haar measure $\sigma(\cdot)$ as S , then the ϵ -enlargements $A_\epsilon := \{x \in \mathbb{S}^{n-1} \mid \text{dist}(x, A) \leq \epsilon\}$ satisfy $\sigma(A_\epsilon) \geq \sigma(S_\epsilon)$. Starting in 1988, in a series of papers Talagrand turned the concentration of measures approach into a universal tool in probability, analysis and the geometry of Banach spaces. Arguably, the contribution Talagrand [169] on the measure concentration on product spaces marks the broad breakthrough of the theory. Very good presentations are due to McDiarmid [123], Ledoux [103] and Boucheron *et al.* [15]. Below we restrict ourselves to the concentration of the law of a random variable.

Definition 2.16 (Milman). Let X be a random variable taking values in a metric space (M, d) . Denote by $K_\ell(0)$ the closed metric ball of radius ℓ and write $A_\ell = A + K_\ell(0)$ for the ℓ -enlargement of a Borel set $A \subseteq M$. *Milman's concentration function* is the function $\alpha_X: (0, \infty) \rightarrow [0, 1/2]$ defined by

$$\alpha_X(\ell) = \sup \left\{ 1 - \mathbb{P}(X \in A_\ell) \mid \mathbb{P}(X \in A) \geq \frac{1}{2} \right\}, \quad \ell > 0. \quad (2.18)$$

Properties 2.17 (Milman's concentration function). *Milman's concentration function α_X of an M -valued random variable X is decreasing, right continuous and satisfies $\alpha_X(0+) \leq \frac{1}{2}$ and $\lim_{\ell \rightarrow \infty} \alpha_X(\ell) = 0$.*

Proof. Let $\ell < \ell + \epsilon$. Since $A_\ell \subseteq A_{\ell+\epsilon}$ we have $\mathbb{P}(X \in A_{\ell+\epsilon}) \geq \mathbb{P}(X \in A_\ell)$, and so $\alpha_X(\ell + \epsilon) \leq \alpha_X(\ell)$. The properties $\alpha_X(0) \leq \frac{1}{2}$ and $\lim_{\ell \rightarrow \infty} \alpha_X(\ell) = 0$ are trivial.

Continuity from the right follows from the right continuity of measures and the fact that we can always interchange suprema:

$$\begin{aligned} \sup_{\epsilon > 0} \alpha_X(\ell + \epsilon) &= \sup \left\{ 1 - \inf_{\epsilon > 0} \mathbb{P}(X \in A_{\ell + \epsilon}) \mid \mathbb{P}(X \in A) \geq \frac{1}{2} \right\} \\ &= \sup \left\{ 1 - \mathbb{P}(X \in A_\ell) \mid \mathbb{P}(X \in A) \geq \frac{1}{2} \right\} = \alpha_X(\ell). \quad \square \end{aligned}$$

If $M = \mathbb{R}$ with the usual Euclidean distance, then we have the following relation between Lévy's and Milman's concentration functions.

Lemma 2.18. *Let X be a real random variable and denote by Q_X and α_X the concentration functions of Lévy and Milman, respectively. Then*

$$1 - 2\alpha_X(\ell) \leq Q_X(2\ell) \leq 2Q_X(\ell). \quad (2.19)$$

Proof. Let $m \in \mathbb{R}$ be a median of X , i.e. a value such that both $\mathbb{P}(X \geq m) \geq \frac{1}{2}$ and $\mathbb{P}(X > m) \leq \frac{1}{2}$ (or, equivalently, $\mathbb{P}(X \leq m) \geq \frac{1}{2}$). Taking $A = (-\infty, m]$ in the definition of α_X shows

$$1 - \mathbb{P}(X \in (-\infty, m + \ell]) \leq \alpha_X(\ell) \quad \text{and so} \quad \mathbb{P}(X - m > \ell) \leq \alpha_X(\ell).$$

In the same way we conclude with $A = [m, \infty)$ that $\mathbb{P}(X - m < -\ell) \leq \alpha_X(\ell)$, i.e.

$$\mathbb{P}(|X - m| > \ell) \leq \mathbb{P}(X - m > \ell) + \mathbb{P}(X - m < -\ell) \leq 2\alpha_X(\ell).$$

From this we get

$$1 - 2\alpha_X(\ell) \leq \mathbb{P}(|X - m| \leq \ell) \leq \sup_{x \in \mathbb{R}} \mathbb{P}(|X - x| \leq \ell) = Q_X(2\ell).$$

The estimate $Q_X(2\ell) \leq 2Q_X(\ell)$ follows from the subadditivity of Q_X , see Property 2.9 no. 4. \square

The following example shows that the estimate in Lemma 2.18 is not sharp and that one cannot expect estimates like $Q_X(2\ell) \leq C(1 - 2\alpha_X(\ell))$ or $Q_X(\ell) \leq C(1 - 2\alpha_X(\ell))$.

Example 2.19. Consider a Bernoulli random variable X with $\mathbb{P}(X = -1) = 1 - p$ and $\mathbb{P}(X = 1) = p$. Without loss of generality we assume that $p \geq \frac{1}{2}$. Then we have

$$\begin{aligned} Q_X(\ell) &= p\mathbf{1}_{[0,2)}(\ell) + \mathbf{1}_{[2,\infty)}(\ell), \quad \alpha_X(\ell) = (1 - p)\mathbf{1}_{[0,2)}(\ell), \\ \text{and} \quad 1 - 2\alpha_X(\ell) &= (2p - 1)\mathbf{1}_{[0,2)}(\ell) + \mathbf{1}_{[2,\infty)}(\ell). \end{aligned}$$

Thus, $1 - 2\alpha_X(\ell) \leq Q_X(2\ell)$ always holds, while $Q_X(2\ell) \leq c(1 - 2\alpha_X(2\ell))$ is only possible for $p > \frac{1}{2}$ and with a constant depending on p .

2.3. Good- λ Inequalities

Good λ -inequalities were introduced by Burkholder & Gundy in order to study both (maximal) inequalities for martingales and maximal functions in analysis. Roughly speaking, good- λ inequalities provide a means to get from estimates for distribution functions to estimates of L^p -norms $\|X\|_{L^p}^p = \mathbb{E}[|X|^p]$, or more general moments, of the random variable X . The key tool is the following *layer cake formula*, which expresses (generalized) moments of a random variable via the distribution function. As usual, we allow integrals of positive functions to attain values in $[0, \infty]$.

Lemma 2.20. *Let $X \geq 0$ be a positive random variable and $\phi: [0, \infty) \rightarrow [0, \infty]$ a positive, right continuous increasing function with $\phi(0) = 0$. Then*

$$\mathbb{E}[\phi(X)] = \int_0^\infty \mathbb{P}(X \geq \lambda) d\phi(\lambda). \quad (2.20)$$

If, in addition, ϕ is absolutely continuous, then

$$\mathbb{E}[\phi(X)] = \int_0^\infty \phi'(\lambda) \mathbb{P}(X \geq \lambda) d\lambda = \int_0^\infty \phi'(\lambda) \mathbb{P}(X > \lambda) d\lambda. \quad (2.21)$$

Proof. Since ϕ is right continuous, $\nu_\phi((a, b]) = d\phi((a, b]) := \phi(b) - \phi(a)$ defines a Borel measure. An application of Tonelli's theorem gives

$$\begin{aligned} \mathbb{E}[\phi(X)] &= \int_\Omega \int_0^\infty \mathbf{1}_{(0, X]}(\lambda) d\phi(\lambda) d\mathbb{P} = \int_0^\infty \int_\Omega \mathbf{1}_{[\lambda, \infty)}(X) d\mathbb{P} d\phi(\lambda) \\ &= \int_0^\infty \mathbb{P}(X \geq \lambda) d\phi(\lambda). \end{aligned}$$

If ϕ is absolutely continuous, the same argument with $\int_0^\infty \mathbf{1}_{(0, X]}(\lambda) d\phi(\lambda) = \phi(X) = \int_0^X \phi'(\lambda) d\lambda$ yields the second formula. Here we use that the set of $\lambda \geq 0$ for which $\mathbb{P}(X \geq \lambda) \neq \mathbb{P}(X > \lambda)$ is countable, hence a Lebesgue null set. \square

Let us begin with Doob's [33, Chapter VII.§3] classical argument, which illustrates the use of the layer cake formula.

Lemma 2.21 (Doob). *Let $X, Y \geq 0$ be positive random variables which satisfy*

$$\mathbb{P}(X > \lambda) \leq \frac{1}{\lambda} \int_{\{X \geq \lambda\}} Y d\mathbb{P} \quad \text{for all } \lambda > 0. \quad (2.22)$$

Then one has

$$\mathbb{E}[X^p] \leq \begin{cases} \left(\frac{p}{p-1}\right)^p \mathbb{E}[Y^p], & \text{if } p \in (1, \infty), \\ \frac{e}{e-1} (1 + \mathbb{E}[Y \log^+ Y]), & \text{if } p = 1. \end{cases} \quad (2.23)$$

Proof. Without loss of generality we may assume that X is bounded. To see this, we note that (2.22) remains valid if we replace $X \rightsquigarrow X \wedge n$ and observe that $\mathbb{E}[X^p] = \sup_{n \geq 1} \mathbb{E}[(X \wedge n)^p]$ by monotone convergence.

Let $\phi(x) = x^p$ or $\phi(x) = (x-1)^+$. Combining the layer cake formula with the estimate (2.22) yields

$$\begin{aligned} \mathbb{E}[\phi(X)] &= \int_0^\infty \phi'(\lambda) \mathbb{P}(X > \lambda) d\lambda \\ &\leq \int_0^\infty \phi'(\lambda) \left(\frac{1}{\lambda} \int_\Omega \mathbb{1}_{\{X > \lambda\}} Y d\mathbb{P} \right) d\lambda \\ &= \int_\Omega \int_0^X \frac{\phi'(\lambda)}{\lambda} d\lambda Y d\mathbb{P}. \end{aligned}$$

If $p \in (1, \infty)$ and $\phi(x) = x^p$, we see that

$$\int_0^X \frac{\phi'(\lambda)}{\lambda} d\lambda = \int_0^X p\lambda^{p-2} d\lambda = \frac{1}{p-1} X^{p-1}.$$

Therefore, we can use Hölder's inequality with $p^{-1} + q^{-1} = 1$,

$$\mathbb{E}[X^p] \leq \frac{1}{p-1} \mathbb{E}[Y X^{p-1}] \leq \frac{1}{p-1} (\mathbb{E}[Y^p])^{\frac{1}{p}} (\mathbb{E}[X^p])^{1-\frac{1}{p}}$$

which yields (2.23).

If $p = 1$ and $\phi(x) = (x-1)^+$, we have

$$\int_0^X \frac{\phi'(\lambda)}{\lambda} d\lambda = \mathbb{1}_{\{X \geq 1\}} \int_1^X \frac{1}{\lambda} d\lambda = \mathbb{1}_{\{X \geq 1\}} \log(X).$$

From the elementary inequality $x^{-1} \log(x) \leq e^{-1}$, $x > 0$,—just find the maximum of $x^{-1} \log(x)$ —we see that

$$\log(x) = \log\left(y \frac{x}{y}\right) = \log(y) + \log\left(\frac{x}{y}\right) \leq \log^+(y) + \frac{x}{ye},$$

hence $y \log(x) \leq y \log^+(y) + x/e$. Therefore,

$$\mathbb{E}[X] \leq 1 + \mathbb{E}[Y \mathbb{1}_{\{X \geq 1\}} \log(X)] \leq 1 + \mathbb{E}[Y \log^+(Y)] + \frac{1}{e} \mathbb{E}[X],$$

from which it is easy to get (2.23). \square

The following lemma will be a substitute for Lemma 2.21 if $0 < p < 1$.

Lemma 2.22. *Let $X, Y \geq 0$ be positive random variables which satisfy*

$$\mathbb{P}(X > \lambda) \leq \frac{1}{\lambda} \mathbb{E}Y \quad \text{for all } \lambda > 0. \quad (2.24)$$

Then one has

$$\mathbb{E}[X^p] \leq \frac{1}{1-p} (\mathbb{E}Y)^p \quad \text{for all } p \in (0, 1). \quad (2.25)$$

Proof. Using the layer-cake formula for $\phi(\lambda) = \lambda$ we get

$$\begin{aligned} \mathbb{E}[X^p] &= \int_0^\infty \mathbb{P}(X > \lambda^{1/p}) d\lambda \leq \int_0^x d\lambda + \mathbb{E}Y \int_x^\infty \lambda^{-1/p} d\lambda \\ &= x + \frac{p}{1-p} x^{1-1/p} \mathbb{E}Y. \end{aligned}$$

The right hand side becomes minimal if $x = (\mathbb{E}Y)^p$, and the claim follows. \square

The next lemma can be seen as a partial converse of (2.23) for $p = 1$.

Lemma 2.23. *Let $0 \leq X \leq Y$ be positive random variables such that for some constants $a, b > 1$*

$$\frac{1}{\lambda} \int_{\{Y > \lambda\}} X d\mathbb{P} \leq b \mathbb{P}(Y \geq \lambda) \quad \text{for all } \lambda > a. \quad (2.26)$$

Then $\mathbb{E}Y < \infty$ implies that $\mathbb{E}[X \log^+ X] < \infty$.

Proof. Assume that $\mathbb{E}Y < \infty$. From $X \leq Y$ we see that $\mathbb{E}X < \infty$. Let $a, b > 0$ be as in the statement of the lemma. The layer-cake formula shows

$$\begin{aligned} \mathbb{E}[X \log^+ X] &= \int_1^a \int_{\{X > \lambda\}} X d\mathbb{P} \frac{d\lambda}{\lambda} + \int_a^\infty \int_{\{X > \lambda\}} X d\mathbb{P} \frac{d\lambda}{\lambda} \\ &\leq \int_1^a \frac{d\lambda}{\lambda} \mathbb{E}X + \int_a^\infty \int_{\{Y > \lambda\}} X d\mathbb{P} \frac{d\lambda}{\lambda} \\ &\leq \log(a) \mathbb{E}X + \int_0^\infty b \mathbb{P}(X \geq \lambda) d\lambda \\ &= \log(a) \mathbb{E}X + b \mathbb{E}Y. \end{aligned}$$

This proves that $\mathbb{E}[X \log^+ X] < \infty$ if $\mathbb{E}[Y] < \infty$. \square

The following *good- λ -inequality* is due to Burkholder [16, Lem. 7.1], earlier versions can be found in Burkholder & Gundy [17, Proof Thm. 4.1], Burkholder, Davies & Gundy [19, Proof Lem. 3.1], and Burkholder & Gundy [18, Proof Thm. 1].

Lemma 2.24. *Let $0 < p < \infty$ and assume that $X, Y \geq 0$ are positive random variables. If there exist $\beta > 1$, $\delta \in (0, 1)$ and $0 < \epsilon < \frac{1}{2}\beta^{-p}$ such that*

$$\mathbb{P}(X > \beta\lambda, Y < \delta\lambda) \leq \epsilon \mathbb{P}(X \geq \lambda) \quad \text{for all } \lambda > 0 \quad (2.27)$$

holds, then there exists a constant $c = c(\beta, \delta, \epsilon, p)$ such that

$$\mathbb{E}[X^p] \leq c \mathbb{E}[Y^p]. \quad (2.28)$$

Proof. As in the proof of Lemma 2.21 we can assume that the random variable X is bounded—note that (2.27) remains valid, if we truncate X . We have

$$\begin{aligned}\mathbb{P}(X > \beta\lambda) &\leq \mathbb{P}(X > \beta\lambda, Y < \delta\lambda) + \mathbb{P}(Y \geq \delta\lambda) \\ &\leq \epsilon\mathbb{P}(X \geq \lambda) + \mathbb{P}(Y \geq \delta\lambda).\end{aligned}$$

An application of the layer cake formula for $\phi(x) = x^p$ shows that

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{\beta}X\right)^p\right] &\leq \epsilon\mathbb{E}[X^p] + \mathbb{E}\left[\left(\frac{1}{\delta}Y\right)^p\right] \\ &\leq \frac{1}{2\beta^p}\mathbb{E}[X^p] + \mathbb{E}\left[\left(\frac{1}{\delta}Y\right)^p\right].\end{aligned}$$

This yields (2.28) with $c = 2\beta^p/\delta^p$. \square

Remark 2.25. It is possible to replace $x \mapsto x^p$ in (2.28) by a positive, increasing function $\phi: [0, \infty) \rightarrow [0, \infty)$ with $\phi(0) = 0$ which satisfies a ‘doubling condition’, i.e.

$$\phi(2x) \leq c\phi(x) \quad \text{for some } c \geq 1 \text{ and all } x \geq 0.$$

The doubling condition implies that $\phi(x)$ has at most polynomial growth. In fact, if $\beta \geq 0$ and $\kappa = \log(\beta)/\log(2)$, then

$$\phi(\beta x) = \phi(2^{\kappa}x) \leq \phi(2^{\lfloor \kappa \rfloor + 1}x) \leq c^{\kappa}\phi(x) = c^{\beta \log(c)/\log(2)}\phi(x).$$

As a consequence, we see for $\beta > 1$, $\delta \in (0, 1)$, and suitable constants c_β and c_δ

$$\mathbb{E}[\phi(X/\beta)] \geq c_\beta\mathbb{E}[\phi(X)] \quad \text{and} \quad \mathbb{E}[\phi(Y/\delta)] \leq c_\delta\mathbb{E}[\phi(Y)],$$

which means that Lemma 2.24 remains valid for this class of functions. A thorough discussion for this and other classes can be found in Lenglart *et al.* [104].

The following variant of the good- λ -inequality reveals its (formal) connection with interpolation theorems. The proof (for $0 < p < 1$) is taken from Revuz & Yor [141, Lem. IV(4.6)], see also Schilling [151, Lem. 19.19], while the general case $0 < p < r$ follows the lines of Stein [166, Chapter I.§4].

Lemma 2.26. *Let $X, Y \geq 0$ be positive random variables such that for some $r > 0$*

$$\mathbb{P}(X > \lambda, Y < \lambda) \leq \frac{1}{\lambda^r} \int_{\{Y < \lambda\}} Y^r d\mathbb{P} \quad \text{for all } \lambda > 0, \quad (2.29)$$

or

$$\mathbb{P}(X > \lambda, Y < \lambda) \leq \frac{1}{\lambda^r} \int (Y \wedge \lambda)^r d\mathbb{P} \quad \text{for all } \lambda > 0, \quad (2.30)$$

holds. Then one has for all $0 < p < r$

$$\mathbb{E}[X^p] \leq \left(\frac{r}{r-p} + 1\right) \mathbb{E}[Y^p]. \quad (2.31)$$

Proof. Since

$$\frac{1}{\lambda^r} \int (Y \wedge \lambda)^r d\mathbb{P} = \frac{1}{\lambda^r} \int_{\{Y < \lambda\}} Y^r d\mathbb{P} + \mathbb{P}(Y \geq \lambda),$$

the argument below shows that either (2.29) or (2.30) yield (2.31), and that the respective constants in (2.31) differ by 1.

From (2.29) we get

$$\begin{aligned} \mathbb{P}(X > \lambda) &= \mathbb{P}(X > \lambda, Y \geq \lambda) + \mathbb{P}(X > \lambda, Y < \lambda) \\ &\leq \mathbb{P}(Y \geq \lambda) + \frac{1}{\lambda^r} \int_{\{Y < \lambda\}} Y^r d\mathbb{P}. \end{aligned}$$

At this point we see that (2.30) just adds one more term $\mathbb{P}(Y \geq \lambda)$ on the right hand side. The layer cake formula with $\phi(x) = x^p$ yields $\mathbb{E}[X^p]$ on the left, and $\mathbb{E}[Y^p]$ from the first summand on the right side. The second term contributes

$$\begin{aligned} \int_0^\infty p\lambda^{p-1}\lambda^{-r} \int_{\{Y < \lambda\}} Y^r d\mathbb{P} d\lambda &= \int_\Omega \int_Y^\infty p\lambda^{-1-(r-p)} d\lambda Y^r d\mathbb{P} \\ &= \frac{p}{r-p} \int_\Omega Y^p d\mathbb{P}; \end{aligned}$$

here we use that $p < r$. Together we arrive at (2.31) with the (slightly better) constant $r/(r-p)$. \square

Remark 2.27. Let us briefly make the connection to interpolation of operators in the scale $L^p(\mu)$, $1 \leq p < \infty$, where μ is a σ -finite Borel measure on \mathbb{R}^d . An operator T defined on $L^p(\mu)$ and taking values in the Borel-measurable functions $\mathcal{B}(\mathbb{R}^d)$ is said to be of *weak type* (p, q) , $1 \leq p, q < \infty$, if there is a constant $c = c_{p,q}$ such that

$$\mu(\{|Tf| > \lambda\}) \leq \left(\frac{c}{\lambda}\|f\|_{L^p}\right)^q \quad \text{for all } \lambda > 0, f \in L^p(\mu). \quad (2.32)$$

The operator is called of (*strong*) *type* (p, q) , if for some constant $C_{p,q}$

$$\|Tf\|_{L^q} \leq C_{p,q}\|f\|_{L^p} \quad \text{for all } f \in L^p(\mu). \quad (2.33)$$

The techniques developed in this section allow us to prove a simple version of the *Marcinkiewicz interpolation theorem*, see e.g. Stein [166, Chapter I.§4.2].³

Theorem (Marcinkiewicz). *Let T be an operator, which is defined on the space $L^1(\mu) + L^r(\mu) := \{f + g \mid f \in L^1(\mu), g \in L^r(\mu)\}$ for some $r \in (1, \infty)$. If T is sublinear, i.e. $|T(f + g)(x)| \leq |Tf(x)| + |Tg(x)|$, and both of weak-type $(1, 1)$ and of weak-type (r, r) , then it is of (strong) type (p, p) for any $1 < p < r$, i.e. $\|Tf\|_{L^p} \leq c_{r,p}\|f\|_{L^p}$.*

³There are also stronger versions of the Marcinkiewicz interpolation theorem for general weak-type (p_0, q_0) and weak-type (p_1, q_1) operators, cf. Bennett & Sharpley [7, Chapter 4.§4].

Proof. Because of the subadditivity of T it is enough to consider positive $f \geq 0$, since we have

$$\{|Tf| > 4\lambda\} \subseteq \{|T(f^+)| > 2\lambda\} \cup \{|T(f^-)| > 2\lambda\}.$$

Let $f \geq 0$ and use again subadditivity to see

$$\{|Tf| > 2\lambda\} \subseteq \{|T(f \wedge \lambda)| > \lambda\} \cup \{|T(f - f \wedge \lambda)| > \lambda\},$$

hence

$$\mu(\{|Tf| > 2\lambda\}) \leq \mu(\{|T(f \wedge \lambda)| > \lambda\}) + \mu(\{|T(f - f \wedge \lambda)| > \lambda\}). \quad (2.34)$$

For the measures appearing on the right hand side we use the weak- (r, r) resp. weak- $(1, 1)$ bounds to get

$$\mu(\{|T(f \wedge \lambda)| > \lambda\}) \leq \frac{c_{r,r}}{\lambda^r} \int (f \wedge \lambda)^r d\mu \quad (2.35)$$

as well as

$$\mu(\{|T(f - f \wedge \lambda)| > \lambda\}) \leq \frac{c_{1,1}}{\lambda} \int (f - f \wedge \lambda) d\mu \leq \frac{c_{1,1}}{\lambda} \int_{\{f \geq \lambda\}} f d\mu. \quad (2.36)$$

We can now apply the layer cake formula with $\phi(x) = x^p$ to (2.34) and use the estimates (2.34), (2.35). On the left-hand side of (2.34) we get $\frac{1}{2}\|Tf\|_{L^p}^p$. The strategy used in the proof of Lemma 2.21 works for $p > 1$, and yields the bound $K_{1,p}\|f\|_{L^p}^p$ for the term in (2.36). Indeed, we have

$$\begin{aligned} \int_0^\infty p\lambda^{p-1} \mu(\{|T(f - f \wedge \lambda)| > \lambda\}) d\lambda &\leq c_{1,1} \int_0^\infty p\lambda^{p-1} \left(\frac{1}{\lambda} \int \mathbb{1}_{\{f \geq \lambda\}} f d\mu \right) d\lambda \\ &= c_{1,1} \int f(x) \int_0^{f(x)} p\lambda^{p-2} d\lambda \mu(dx) \\ &= \frac{c_{1,1}p}{p-1} \|f\|_{L^p}^p. \end{aligned}$$

For the term (2.35) appearing on the right-hand side, we can literally use the calculation from the proof of Lemma 2.26 with $p < r$ to get the bound $\kappa_{r,p}\|f\|_{L^p}^p$ with $\kappa_{r,p} = r/(r-p)$.

As one would expect, the constants $\kappa_{r,p}$ and $K_{1,p}$ become infinite as $p \uparrow r$ and $p \downarrow 1$, respectively. \square

Further applications of the good- λ inequality in harmonic analysis are discussed in Torchinsky [173, Chapter XIII].

Good- λ inequalities play also an important role in the extrapolation problem for function spaces, see Geiss [44] and Hytönen *et al.* [62, Ch. 3]. That paper also contains a discussion how one can formulate certain weak-type L^1 -estimates in a BMO (functions of bounded mean oscillation, see Remark 3.11) setting in order to obtain strong L^p -estimates.

3. Some general principles

Let $(X_t)_{t \geq 0}$ be a stochastic process with values in \mathbb{R}^d defined on a filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F}_t)$; we assume that the process is *adapted* to the filtration, i.e. each random variable X_t is \mathcal{F}_t -measurable. Without further assumptions, the supremum processes $X_t^* := \sup_{s \leq t} |X_s|$ or $M_t := \sup_{s \leq t} X_s$ may not be measurable. Typically, if $t \mapsto X_t$ is left- or right continuous, this is not an issue. For a long time the notion of *separability* was used.

Separability A process is called *separable* with *separating set* $D \subseteq [0, \infty)$, if for some countable set D the set $\{(t, X_t(\omega)) \mid t \in D\}$ is for almost all $\omega \in \Omega$ dense (w.r.t. the Euclidean topology in $[0, \infty) \times \mathbb{R}^d$) in the graph $\{(t, X_t(\omega)) \mid t \geq 0\}$. In particular, this allows us to calculate suprema etc. along the countable set D , rendering everything measurable. A full discussion on separability can be found in Dellacherie & Meyer [26, pp. 153–161] and Loève [116, Vol. 2, pp. 170–178].

The notion of *separability* is hardly seen in modern texts. This is mainly due to the following two results:

Fact I (Doob). *Every stochastic process $(X_t)_{t \geq 0}$ with values in \mathbb{R}^d has a separable modification taking values in $[-\infty, +\infty]^d$.*

A proof of this can be found in Dellacherie & Meyer [26, Thm. 29, p. 157] or in Doob [33, Thm. II.2.4]. Note that the modification is, in general, not \mathbb{R}^d -valued; moreover, we have no control on what the separating set D will be.

Fact II. *Let $(X_t)_{t \geq 0}$ be a stochastic process with values in \mathbb{R}^d which is defined on a complete probability space.*

- (a) *If $(X_t)_{t \geq 0}$ is continuous in probability, i.e. $\lim_{s \rightarrow t} \mathbb{P}(|X_s - X_t| > \epsilon) = 0$ for all $\epsilon > 0$ and $t \geq 0$, then there is a separable modification taking values in $[-\infty, \infty]^d$.*
- (b) *If $\rho(s, t) := \sqrt{\mathbb{E}[|X_t - X_s|^2]}$ is finite for all $s, t \geq 0$ and the pseudo metric space $([0, \infty), \rho)$ is separable, then there is a separable modification taking values in \mathbb{R}^d .*

Moreover, if (a) or (b) holds, every countable dense set $D \subseteq [0, \infty)$ can be used as a separating set.

We refer to Wentzell [177, Satz 3, p. 79] and Lifshits [109, Prop. 1, p. 26] for the simple proof.

Comparison of $|X_t|$, X_t^* and $X_t, M_t := \sup_{s \leq t} X_s$. Frequently we state a theorem only for $|X_t|$ and X_t^* or for X_t and M_t . The following argument gives a useful criterion that compares $\mathbb{E}[X_t^*]$ and $\mathbb{E}[M_t]$.

Lemma 3.1. *Let $(X_t)_{t \geq 0}$ be a separable real-valued process such that $\mathbb{E}|X_t| < \infty$ for all $t \geq 0$. Then*

$$\mathbb{E}[X^*] < \infty \iff -\infty < \mathbb{E} \left[\inf_{t \geq 0} X_t \right] \leq \mathbb{E} \left[\sup_{t \geq 0} X_t \right] < \infty. \quad (3.1)$$

If the process is symmetric, i.e. $X_t \sim -X_t$, then

$$\begin{aligned} \mathbb{E} \left[\sup_{t \geq 0} |X_t| \right] &\leq 3\mathbb{E} \left[\sup_{t \geq 0} X_t \right] + 2 \inf_{t \geq 0} \mathbb{E} [|X_t|] \\ &\leq 3\mathbb{E} \left[\sup_{t \geq 0} X_t \right] + 2 \inf_{t \geq 0} \sqrt{\mathbb{V} [X_t]}. \end{aligned} \quad (3.2)$$

Proof. Note that $|X_t| = 2X_t^+ - X_t$ and

$$\sup_{t \geq 0} |X_t| \leq 2 \sup_{t \geq 0} X_t^+ + \sup_{t \geq 0} (-X_t) = 2 \sup_{t \geq 0} X_t^+ - \inf_{t \geq 0} X_t.$$

Moreover, $\sup_{t \geq 0} X_t^+ \leq \sup_{t \geq 0} X_t + \inf_{t \geq 0} |X_t|$. This clever estimate is taken from Lifshits [110, Prop. 10.2], and it follows from considering two cases: If $\sup_{t \geq 0} X_t \geq 0$, then $\sup_{t \geq 0} X_t = \sup_{t \geq 0} X_t^+$, while for $\sup_{t \geq 0} X_t < 0$ the left hand side is zero. Together we have

$$\sup_{t \geq 0} |X_t| \leq 2 \sup_{t \geq 0} X_t + 2 \inf_{t \geq 0} |X_t| - \inf_{t \geq 0} X_t \leq 5 \sup_{t \geq 0} |X_t|$$

from which the claim follows at once. Taking expectations, we get

$$\mathbb{E} \left[\sup_{t \geq 0} |X_t| \right] \leq 2\mathbb{E} \left[\sup_{t \geq 0} X_t \right] + 2 \inf_{t \geq 0} \mathbb{E} [|X_t|] - \mathbb{E} \left[\inf_{t \geq 0} X_t \right].$$

If the process is symmetric, then $-\mathbb{E} \left[\inf_{t \geq 0} X_t \right] = \mathbb{E} \left[\sup_{t \geq 0} (-X_t) \right] = \mathbb{E} \left[\sup_{t \geq 0} X_t \right]$, and (3.2) follows. \square

The Kolmogorov–Chentsov theorem When talking about maximal inequalities, the Kolmogorov–Chentsov continuity theorem sounds like an unlikely candidate. However, the formulation pioneered by Totoki [174] and widely used in the theory of stochastic flows, see e.g. Meyer [127, pp. 105–106] and Kunita [98, Appendix], clearly shows the ‘maximal estimate’ character of this result. Our statement follows the exposition in Revuz & Yor [141, Thm. I.(2.1)] and Schilling [151, Thm. 10.1], and we refer to these monographs for the (by now standard) proof.

Theorem 3.2 (Kolmogorov–Chentsov–Totoki). *Let $(\xi(x))_{x \in \mathbb{R}^n}$ be a stochastic process taking values in \mathbb{R}^d and indexed by $x \in \mathbb{R}^n$. If*

$$\forall x, y \in \mathbb{R}^n : \mathbb{E} [|\xi(x) - \xi(y)|^\alpha] \leq c|x - y|^{n+\beta} \quad (3.3)$$

holds for some constants $c > 0$ and $\alpha, \beta > 0$, then $(\xi(x))_{x \in \mathbb{R}^n}$ has a modification with continuous sample paths. Moreover

$$\mathbb{E} \left[\sup_{\substack{0 < |x-y| < 1 \\ x, y \in [-T, T]^n}} \frac{|\xi(x) - \xi(y)|^\alpha}{|x - y|^\gamma} \right] < \infty \quad (3.4)$$

for all $T > 0$ and $0 \leq \gamma < \beta$. In particular, $x \mapsto \xi(x)$ is a.s. locally Hölder continuous of order γ/α .

Without real changes in the argument, we may in (3.4) replace $0 < |x - y| < 1$ by $0 < |x - y| < \kappa$ for some $\kappa < T$. Therefore, we may read (3.4) as

$$\mathbb{E} \left[\sup_{\substack{0 < |x-y| < \kappa \\ x, y \in [-T, T]^n}} |\xi(x) - \xi(y)|^\alpha \right] \leq C_T \kappa^\gamma, \quad 0 < \gamma < \beta.$$

Sometimes, a further variant is useful. Recall that $B_\kappa(z)$ is an open ball with centre z and radius κ . Since $\{x, y \in [-T, T]^n \mid x, y \in B_\kappa(z), z \in [-T, T]^n\} \subseteq \{x, y \in [-T, T]^n \mid |x - y| < 2\kappa\}$, we also have

$$\mathbb{E} \left[\sup_{z \in [-T, T]^n} \sup_{x, y \in B_\kappa(z) \cap [-T, T]^n} |\xi(x) - \xi(y)|^\alpha \right] \leq C_T \kappa^\gamma, \quad 0 < \gamma < \beta.$$

If we replace $n \rightsquigarrow 1+n$ and $x \rightsquigarrow (i, x) \in \mathbb{R} \times \mathbb{R}^n$, the above theorem also shows that the estimate $\mathbb{E} [|\xi(i, x) - \xi(i, y)|^\alpha] \leq c|x - y|^{1+n+\beta}$ for all $(i, x), (i, y) \in \mathbb{R} \times \mathbb{R}^n$ implies

$$\mathbb{E} \left[\sup_{\substack{i \in [-T, T], \\ x, y \in [-T, T]^n}} |\xi(i, x) - \xi(i, y)|^\alpha \right] \leq C_T \kappa^\gamma, \quad \gamma < \beta.$$

The supremum over the extra parameter i ‘costs us’ one extra power in the estimate (3.3).

There are several refinements of Theorem 3.2, which use methods that were discovered when studying Gaussian processes. A good survey is Pisier [138] and the paper Schilling [147] explains the connection to the Sobolev embedding theorem. The following result is due to Pisier [138, Thm. 2.1], we follow the strategy worked out in Lifshits [110, Thm. 10.1, Exercise 10.1]. For the proof we need the notion of a covering number (see also Definition 8.8 further down): Let (\mathbb{T}, d) be a compact (pseudo-)metric space. The covering number is defined as

$$N(\epsilon) := \min \{n \mid \mathbb{T} \subseteq \bigcup_{k=1}^n A_k, \text{diam}(A_k) \leq \epsilon\}, \quad \epsilon > 0. \quad (3.5)$$

Lemma 3.3 (Pisier). *Let $(X_t)_{t \geq 0}$ be a separable stochastic process taking values in \mathbb{R}^d and such that $m_t^p := \sup_{s \leq t} \mathbb{E} [|X_s|^p] < \infty$ for some $p \in (1, \infty)$ and all $t > 0$. Let $\rho(r, s) := \|X_r - X_s\|_{L^p}$ and denote by $N(\epsilon)$ the covering number induced by ρ on every compact set $[0, t]$. Then*

$$\mathbb{E} [X_t^*] \leq 4 \int_0^{m_t/2} N(\epsilon)^{1/p} d\epsilon. \quad (3.6)$$

Proof. We set up a chaining scheme.⁴ Fix $t > 0$, set $\epsilon_i := m_t 2^{-i}$, $i \in \mathbb{N}_0$, and cover $[0, t]$ by a minimal covering with sets of ρ -diameter less than ϵ_i . By the definition of the covering number, the covering contains $N(\epsilon_i)$ sets. Take in each covering set a point $t_k = t_k(i)$ and set $T_i = \{t_1, \dots, t_{N(\epsilon_i)}\}$. Now we construct a map

⁴An abstract discussion of ‘generic’ chaining schemes is in Talagrand [171].

$\pi_i: [0, t] \rightarrow T_i$ satisfying $\forall s \in [0, t] : \rho(s, \pi_i(s)) = \|X_s - X_{\pi_i(s)}\|_{L^p} \leq \epsilon_i$.

Clearly,

$$\sup_{s \in T_i} |X_s| \leq \sup_{s \in T_{i-1}} |X_s| + \sup_{s \in T_i} |X_s - X_{\pi_{i-1}(s)}|.$$

Taking expectations, we see that

$$\begin{aligned} \mathbb{E} \left[\sup_{s \in T_i} |X_s| \right] &\leq \mathbb{E} \left[\sup_{s \in T_{i-1}} |X_s| \right] + \left(\mathbb{E} \left[\sup_{s \in T_i} |X_s - X_{\pi_{i-1}(s)}|^p \right] \right)^{1/p} \\ &\leq \mathbb{E} \left[\sup_{s \in T_{i-1}} |X_s| \right] + \left(\sum_{s \in T_i} \mathbb{E} [|X_s - X_{\pi_{i-1}(s)}|^p] \right)^{1/p} \\ &\leq \mathbb{E} \left[\sup_{s \in T_{i-1}} |X_s| \right] + N(\epsilon_i)^{1/p} \epsilon_{i-1}. \end{aligned}$$

If we iterate this argument, observe that $N(\epsilon)$ is decreasing and that $\epsilon_i - \epsilon_{i+1} = \epsilon_{i+1} = \frac{1}{2}\epsilon_i$, then we get

$$\begin{aligned} \mathbb{E} \left[\sup_{s \in T_n} |X_s| \right] &\leq \sum_{i=1}^n N(\epsilon_i)^{1/p} \epsilon_{i-1} = 4 \sum_{i=1}^n \int_{\epsilon_{i+1}}^{\epsilon_i} N(\epsilon)^{1/p} d\epsilon \\ &\leq 4 \sum_{i=1}^n \int_{\epsilon_{i+1}}^{\epsilon_i} N(\epsilon)^{1/p} d\epsilon \leq 4 \int_0^{m_t/2} N(\epsilon)^{1/p} d\epsilon. \end{aligned}$$

If we apply the above argument to the process $(X_s)_{s \in D_n \cap [0, t]}$ defined on the finite set $D_n = \{k2^{-n} \mid k \in \mathbb{N}_0\} \cup \{t\}$, we get the estimate (3.6) first for $\mathbb{E} [\max_{s \in D_n \cap [0, t]} |X_s|]$; note that D_n is reached by some T_i after finitely many iterations. Since the right-hand side of (3.6) does not depend on D_n , we can use monotone convergence to bound $\mathbb{E} [\sup_{s \in D \cap [0, t]} |X_s|]$ for $D = \bigcup_n D_n$. Finally, by separability, $X_t^* = \sup_{s \in D \cap [0, t]} |X_s|$, finishing the proof. \square

The proof also works if we use a further metric $d(s, t)$ on $[0, \infty)$ such that $\rho(s, t) \leq cd(s, t)$. If $d(s, t) = d(0, t - s)$ is invariant under translations, the d -covering number is $N(\epsilon) \approx d(0, \epsilon)^{-1}$. If, in particular, $\|X_s - X_t\|_{L^p} \leq c|t - s|^\alpha$ for some $\alpha \leq 1$, we get

$$\mathbb{E} [X_t^*] \leq c \int_0^{m_t} \epsilon^{-\alpha/p} d\epsilon = c_{\alpha, p} m_t^{1-\alpha/p} = c_{\alpha, p} \left(\sup_{s \leq t} (\mathbb{E} |X_s|^p)^{1/p} \right)^{1-\alpha/p}.$$

This is essentially a refinement of the Kolmogorov–Chentsov theorem.

A systematic exposition of processes $(X_t)_{t \geq 0}$ satisfying an ‘increment condition’, i.e. an inequality of the form $\mathbb{P} (|X_t - X_s| > \lambda) \leq \exp [-\lambda^2/2d^2(s, t)]$ where $d(s, t)$ is some metric on the parameter space of the process $(X_t)_{t \geq 0}$, can be found in Talagrand [171, Ch. 2]. For example, for a Gaussian process we have $d(s, t) = \|X_s - X_t\|_{L^2}$, see Chapter 8.

A maximal inequality for processes with controlled increments The following theorem was found by Liu & Xi [115], without making the connection

to the Kolmogorov–Chentsov–Totoki theorem. Let $(X_t, \mathcal{F}_t)_{t \in [0, T]}$ be an adapted \mathbb{R}^d -valued stochastic process satisfying a *conditional increment control* with parameters (p, h) , $p > 1$, $h \in (0, 1)$ and $ph > 1$, if $\mathbb{E}[|X_t|^p] < \infty$ and if there is a constant $A_{p,h}$ such that

$$\forall 0 \leq s < t \leq T : \quad \mathbb{E}[|\mathbb{E}(X_t | \mathcal{F}_s) - X_s|^p] \leq A_{p,h}|t - s|^{ph}. \quad (3.7)$$

Theorem 3.4. *Let $(X_t, \mathcal{F}_t)_{t \in [0, T]}$ be an \mathbb{R}^d -valued càdlàg process with conditionally controlled increments with parameters (p, h) such that $p > 1$, $h \in (0, 1)$ and $ph > 1$. Define for $[u, v] \subseteq [0, T]$ the maximal process $X_{u,v}^* := \sup_{u \leq s \leq v} |X_s|$. There is a constant $C_{p,h,\epsilon}$ such that for any $0 < \epsilon < ph - 1$*

$$\mathbb{E}[(X_{u,v}^*)^p] \leq \frac{c_p p^p}{(p-1)^p} (C_{p,h,\epsilon} A_{p,h} |u - v|^{ph-1-\epsilon} + \mathbb{E}[|X_v|^p]).$$

The paper Liu & Xi [115] shows the result of Theorem 3.4 with $|u - v|^{ph}$ instead of $|u - v|^{ph-1-\epsilon}$, but the proof is much more complicated. The key is an improved version of Lemma 3.5, which is not directly deduced from Theorem 3.2, but the authors give a more refined version of the argument used in the proof of Theorem 3.2. If one is interested in large values of p , our shortcut is not too much of a loss. The following two auxiliary results are simplified versions of results in Liu & Xi [115].

Lemma 3.5. *Let $(X_t)_{t \in [0, T]}$ be a process as in Theorem 3.4. There is a constant $C_{p,h}$ such that for all $[u, v] \subseteq [0, T]$ and $t \in [u, v]$ and any $0 < \epsilon < ph - 1$*

$$\mathbb{E} \left[\sup_{u \leq s \leq v} |\mathbb{E}(X_t | \mathcal{F}_s) - X_s|^p \right] \leq C_{p,h,\epsilon} A_{p,h} |u - v|^{ph-1-\epsilon}.$$

Proof. Pick $0 \leq u < v \leq T$ and consider the process $Y_t := \mathbb{E}(X_t | \mathcal{F}_u)$ for $u \leq s < t \leq v$. By the tower property and the conditional Jensen inequality we have

$$\begin{aligned} \mathbb{E}[|Y_t - Y_s|^p] &= \mathbb{E}[|\mathbb{E}(X_t - X_s | \mathcal{F}_u)|^p] \\ &= \mathbb{E}[|\mathbb{E}(\mathbb{E}(X_t | \mathcal{F}_s) - X_s | \mathcal{F}_u)|^p] \\ &\leq \mathbb{E}[|\mathbb{E}(X_t | \mathcal{F}_s) - X_s|^p] \\ &\leq A_{p,h}|t - s|^{ph}. \end{aligned}$$

We can now apply Theorem 3.2, cf. see the discussion following the statement of that theorem, to conclude

$$\begin{aligned} \mathbb{E} \left[\sup_{u \in [0, T]} \sup_{t \in [u, v]} |\mathbb{E}(X_t | \mathcal{F}_u) - X_u|^p \right] &= \mathbb{E} \left[\sup_{u \in [0, T]} \sup_{t \in [u, v]} |Y_t - Y_u|^p \right] \\ &\leq \mathbb{E} \left[\sup_{u \in [0, T]} \sup_{u \leq s < t \leq v} |Y_t - Y_s|^p \right] \\ &\leq C_{p,h} |u - v|^{ph-1-\epsilon} \end{aligned}$$

from which the claim follows. \square

Lemma 3.6. *Let $(X_t, \mathcal{F}_t)_{t \in [0, T]}$ be an adapted càdlàg process with $\mathbb{E}|X_t| < \infty$ for all $t \in [0, T]$. For all $[u, v] \subseteq [0, T]$, $X_{u,v}^* := \sup_{u \leq s \leq v} |X_s|$ and $\lambda > 0$*

$$\mathbb{P}(X_{u,v}^* > \lambda) \leq \frac{1}{\lambda} \int_{X_{u,v}^* > \lambda} \left(\sup_{s \in [u, v]} |\mathbb{E}(X_v | \mathcal{F}_s) - X_s| + |X_v| \right) d\mathbb{P}. \quad (3.8)$$

Proof. Fix $\lambda > 0$ and define a stopping time $\tau = \inf \{s \in [u, v] \mid |X_s| > \lambda\} \wedge T$ with $\inf \emptyset := \infty$. Since the process has càdlàg paths,

$$\{\tau < T\} = \{u \leq \tau \leq v\} \supseteq \{X_{u,v}^* > \lambda\},$$

and $|X_\tau| > \lambda$ on the set $\{u \leq \tau \leq v\}$. Thus,

$$\begin{aligned} \mathbb{P}(X_{u,v}^* > \lambda) &\leq \frac{1}{\lambda} \int_{u \leq \tau \leq v} |X_\tau| d\mathbb{P} \\ &\leq \frac{1}{\lambda} \int_{u \leq \tau \leq v} \left[|\mathbb{E}(X_v | \mathcal{F}_\tau) - X_\tau| + |\mathbb{E}(X_v | \mathcal{F}_\tau)| \right] d\mathbb{P} \\ &\leq \frac{1}{\lambda} \int_{u \leq \tau \leq v} \left[\sup_{s \in [u, v]} |\mathbb{E}(X_v | \mathcal{F}_s) - X_s| + \mathbb{E}(|X_v| | \mathcal{F}_\tau) \right] d\mathbb{P} \\ &\leq \frac{1}{\lambda} \int_{u \leq \tau \leq v} \left[\sup_{s \in [u, v]} |\mathbb{E}(X_v | \mathcal{F}_s) - X_s| + |X_v| \right] d\mathbb{P}. \quad \square \end{aligned}$$

Proof of Theorem 3.4. Lemma 3.6 and Lemma 2.21 show for $p > 1$ with $ph > 1$

$$\mathbb{E}[(X_{u,v}^*)^p] \leq c_p \left(\frac{p}{p-1} \right)^p \mathbb{E} \left[\sup_{s \in [u, v]} |\mathbb{E}(X_v | \mathcal{F}_s) - X_s|^p + |X_v|^p \right].$$

From this the claim follows with the estimate in Lemma 3.5. \square

Remark 3.7. The appearance of the conditional increment control (3.7) in connection with the Kolmogorov-Chentsov-Totoki theorem is not too surprising. A similar condition appears in tightness and compactness theorems for the weak convergence of stochastic processes, see Ethier & Kurtz [41, p. 138, Rem. 3.8.7].

Another type of conditional domination can be used to get exponential moments for càdlàg processes with bounded jumps $\Delta X_t := X_t - X_{t-}$. The clever proof of the following lemma is due to Bass [4, Thm. I.(6.11)].

Lemma 3.8. *Let $(X_t, \mathcal{F}_t)_{t \in [0, \infty]}$ be an adapted càdlàg process with bounded jumps, i.e. $\sup_{t > 0} |\Delta X_t| \leq c$ for some constant $c > 0$. If there is some constant $C > 0$ such that*

$$\mathbb{E}(|X_\infty - X_{\tau-}| | \mathcal{F}_\tau) \leq C \quad \text{for all stopping times } \tau, \quad (3.9)$$

then there exists an exponent $p > 0$ such that $\mathbb{E}[\exp(pX^)] < \infty$.*

Proof. There is no loss of generality in assuming $c = C = 1$, as we can always consider the scaled process $(c + C)^{-1}X_t$. Let τ be any stopping time, $\lambda > 0$ and $F \in \mathcal{F}_\tau$. By the Markov inequality we get

$$\begin{aligned} \mathbb{P}(\{|X_\infty - X_{\tau-}| > \lambda\} \cap F) &\leq \frac{1}{\lambda} \int_F |X_\infty - X_{\tau-}| d\mathbb{P} \\ &= \frac{1}{\lambda} \int_F \underbrace{\mathbb{E}(|X_\infty - X_{\tau-}| \mid \mathcal{F}_\tau)}_{\leq 1} d\mathbb{P} \leq \frac{1}{\lambda} \mathbb{P}(F). \end{aligned}$$

We will now bound the tail of X^* . For this, we introduce stopping times

$$\sigma := \inf\{t > 0 \mid |X_t| \geq \lambda + 5\} \quad \text{and} \quad \tau := \inf\{t > 0 \mid |X_t| \geq \lambda + 15\},$$

and we use in the following calculation that $|X_{\tau-}| \geq \lambda + 14$ and $|X_{\sigma-}| \leq \lambda + 5$, since $t \mapsto X_t$ is càdlàg with jumps of size less than one. For any $\lambda > 0$

$$\begin{aligned} \mathbb{P}(X^* > \lambda + 20) &= \mathbb{P}(X^* > \lambda + 20, |X_\infty| \leq \lambda + 10) + \mathbb{P}(|X_\infty| > \lambda + 10) \\ &\leq \mathbb{P}(\tau < \infty, |X_\infty| \leq \lambda + 10) + \mathbb{P}(\sigma < \infty, |X_\infty| > \lambda + 10) \\ &\leq \mathbb{P}(\tau < \infty, |X_\infty - X_{\tau-}| > 4) + \mathbb{P}(\sigma < \infty, |X_\infty - X_{\sigma-}| > 5) \\ &\leq \frac{1}{4} \mathbb{P}(\tau < \infty) + \frac{1}{5} \mathbb{P}(\sigma < \infty). \end{aligned}$$

In the last inequality we use the estimate from the beginning of the proof. Since $\{\sigma < \infty\} \cup \{\tau < \infty\} \subseteq \{X^* > \lambda\}$, we get

$$\mathbb{P}(X^* > \lambda + 20) \leq \frac{1}{2} \mathbb{P}(X^* > \lambda),$$

and if we iterate this inequality, taking $\lambda = 20n$, $n \in \mathbb{N}_0$, we end up with $\mathbb{P}(X^* > 20n) \leq 2^{-n}$. Now we can apply the layer cake formula and conclude

$$\mathbb{E}\left[e^{pX^*}\right] = \sum_{n=0}^{\infty} \int_{20n}^{20(n+1)} p e^{p\lambda} \mathbb{P}(X^* > \lambda) d\lambda \leq p e^{20p} \sum_{n=0}^{\infty} e^{20pn} 2^{-n},$$

which is finite if p is small enough. □

Before we give an application of Lemma 3.8 to BMO (bounded mean oscillation) martingales, we want to record the analogue of Lemma 3.8 for increasing processes. This result is due to Dellacherie [25] and Lenglart *et al.* [104] who, in turn, credit Garsia and Neveu for the discrete-time case.

Lemma 3.9. *Let $(A_t, \mathcal{F}_t)_{t \geq 0}$ be an adapted, increasing càdlàg process such that $A_0 = 0$, $A_\infty := \lim_{t \rightarrow \infty} A_t$, $\mathbb{E}A_\infty < \infty$ and*

$$\mathbb{E}[A_\infty - A_{\tau-} \mid \mathcal{F}_\tau] \leq X \mathbf{1}_{\{\tau < \infty\}} \tag{3.10}$$

for some integrable random variable $X \geq 0$ and all stopping times τ . Then, for every positive convex function F , whose right derivative is denoted by $f = F'_+$, one has⁵

$$\mathbb{E}[F(A_\infty)] \leq \mathbb{E}[Xf(A_\infty)] + F(0).$$

If we use the lemma for $X \equiv c$, $\kappa \wedge A_t$ (note that $\kappa \wedge A_\infty - \kappa \wedge A_t \leq A_\infty - A_t$ since $A_t \leq A_\infty$), $F(x) = e^{px}$ and $f(x) = pe^{px}$ for some $p < 1/c$, we get

$$\mathbb{E}[\exp(p\kappa \wedge A_\infty)] \leq pc\mathbb{E}[\exp(p\kappa \wedge A_\infty)] + 1,$$

hence

$$\mathbb{E}[\exp(p\kappa \wedge A_\infty)] \leq (1 - pc)^{-1};$$

now we use Fatou's lemma to let $\kappa \rightarrow \infty$. This gives $\mathbb{E}[\exp(pA_\infty)] \leq (1 - pc)^{-1}$ making the connection to Lemma 3.8.

Proof. We multiply both sides of (3.10) with $\mathbf{1}_{\{\tau < \infty\}}$ and take expectations to get $\mathbb{E}[(A_\infty - A_{\tau-})] \leq \mathbb{E}[X\mathbf{1}_{\{\tau < \infty\}}]$ – here we use that $A_\infty = A_{\infty-}$. Fix $\lambda > 0$ and use the stopping time $\tau = \inf\{t > 0 \mid A_t \geq \lambda\}$. Since $A_{\tau-} \leq \lambda$ and $\{A_\infty \geq \lambda\} \supseteq \{\tau < \infty\}$

$$\mathbb{E}[(A_\infty - \lambda)\mathbf{1}_{\{A_\infty \geq \lambda\}}] \leq \mathbb{E}[A_\infty - A_{\tau-}] \leq \mathbb{E}[X\mathbf{1}_{\{\tau < \infty\}}] \leq \mathbb{E}[X\mathbf{1}_{\{A_\infty \geq \lambda\}}].$$

Now we can integrate both sides with respect to $df(\lambda)$ and get

$$\begin{aligned} \int_0^{A_\infty} (A_\infty - \lambda) df(\lambda) &= \int_0^{A_\infty} \int_\lambda^{A_\infty} dx df(\lambda) = \int_0^{A_\infty} \int_0^x df(\lambda) dx \\ &= \int_0^{A_\infty} (f(x) - f(0)) dx = F(A_\infty) - F(0) - f(0)A_\infty \end{aligned}$$

on the left, and $X(f(A_\infty) - f(0))$ on the right. Thus,

$$\mathbb{E}[F(A_\infty)] - F(0) \leq \mathbb{E}[(A_\infty - X)f(0) + Xf(A_\infty)] \leq \mathbb{E}[Xf(A_\infty)]. \quad \square$$

We will finally apply Lemma 3.8 to derive the John–Nirenberg inequality for BMO martingales. For details on martingales we refer to the following Chapter 4. A martingale $(X_t, \mathcal{F}_t)_{t \geq 0}$, which is L^2 -bounded, i.e. $\sup_{t \geq 0} \mathbb{E}[X_t^2] < \infty$, and has continuous paths, is of class BMO (*bounded mean oscillation*), if there exists a constant $c > 0$ such that

$$\mathbb{E}[(X_\infty - X_\tau)^2 \mid \mathcal{F}_\tau] \leq c^2 \quad \text{for all stopping times } \tau. \quad (3.11)$$

We define the norm $\|X\|_{\text{BMO}}$ as the infimum of all constants c in (3.11).

⁵There is a misprint in Lenglart *et al.* [104, Lem. 1.2.a): $\mathbb{E}[A_\infty f(X)]$ should read $\mathbb{E}[f(A_\infty)X]$

Theorem 3.10 (John, Nirenberg). *Let $(X_t)_{t \geq 0}$ be a (continuous) martingale of class BMO such that $0 < \|X\|_{\text{BMO}} < \infty$. There exists a constant $p > 0$ such that*

$$\mathbb{E} [\exp (pX^* / \|X\|_{\text{BMO}})] < \infty.$$

Proof. Since $(X_t)_{t \geq 0}$ is L^2 -bounded, it is uniformly integrable and so the limit $X_\infty = \lim_{t \rightarrow \infty} X_t$ and X^* exist in $L^2(\mathbb{P})$. Since $t \mapsto X_t$ is continuous, the martingale has no jumps. Moreover, by the conditional Jensen inequality, we see for all stopping times

$$\mathbb{E} [|X_\infty - X_\tau| \mid \mathcal{F}_\tau] \leq \sqrt{\mathbb{E} [|X_\infty - X_\tau|^2 \mid \mathcal{F}_\tau]} \leq \|X\|_{\text{BMO}},$$

and the claim follows from Lemma 3.8. \square

A systematic study of (weighted) BMO-spaces for stochastic processes is in Geiss [45, Sec. 2], see also Geiss [44] for martingale difference BMO-spaces.

Remark 3.11. In (harmonic) analysis a measurable function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ is said to be in BMO if

$$\sup_Q \frac{1}{|Q|} \int_Q |f(x) - f_Q| dx \leq c < \infty \quad (3.12)$$

where the supremum ranges over all axis-parallel cubes $Q \subseteq \mathbb{R}^d$, $|Q|$ is the Lebesgue measure of Q and $f_Q = |Q|^{-1} \int_Q f(y) dy$ is the mean over Q . The BMO-norm $\|f\|_{\text{BMO}}$ is the infimum of all constants c appearing in the inequality (3.12). We will again encounter a similar kind of ‘dyadic maximal function’ in Example 4.8.

The original analyst’s version of the John–Nirenberg inequality reads as follows: *There is some constant $\alpha > 0$ such that for each axis-parallel cube $Q \subseteq \mathbb{R}^d$ the following inequality holds:*

$$|\{x \in Q \mid |f(x) - f_Q| > \lambda\}| \leq |Q| e^{-\alpha \lambda / \|f\|_{\text{BMO}}}, \quad \lambda > 0.$$

A martingale proof can be found in Bass [5, Prop. IV.(7.6)], the standard proof is in Torchinsky [173, Ch. VIII, Thm. 1.3].

Let us briefly mention the origins of this class. In analysis, the L^p scale is often used but the end points of this scale, the spaces L^1 and L^∞ , are special since they are not reflexive. In many problems in analysis, therefore, the space L^1 is replaced by a slightly smaller space such as the $L \log L$ class in \mathbb{R}^d (see also § 4.1) or the real Hardy space $H^1(\mathbb{R}^d)$. For definitions and details we refer to the monograph by Torchinsky [173, Ch. XIV, XV]. It turned out – and even led to a Fields medal for Charles Fefferman – that the space $\text{BMO}(\mathbb{R}^d)$ is the topological dual of $H^1(\mathbb{R}^d)$. Of course, there are many more useful applications of BMO in harmonic analysis and the study of partial differential equations.

4. Martingales

For most mathematicians, the first encounter with maximal inequalities in probability theory happens in connection with martingales. Here we will consider martingales indexed with $I = \mathbb{N}_0$ or $I = [0, \infty)$. Recall that the underlying filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F}_t)$ satisfies the usual conditions, i.e. $(\Omega, \mathcal{A}, \mathbb{P})$ is complete, \mathcal{F}_0 contains all \mathbb{P} -null sets and $(\mathcal{F}_t)_{t \geq 0}$ is right-continuous (if $I = [0, \infty)$).

Definition 4.1. Let $X = (X_t, \mathcal{F}_t)_{t \in I}$ be a real-valued stochastic process; if $I = [0, \infty)$, we assume, in addition, that $t \mapsto X_t(\omega)$ is càdlàg. The process X is said to be a *martingale*, if

1. for every $t \in I$, the random variable X_t is \mathcal{F}_t measurable;
2. $\mathbb{E}|X_t| < \infty$ for every $t \in I$;
3. $\int_F X_s d\mathbb{P} = \int_F X_t d\mathbb{P}$ for all $s, t \in I$ with $s < t$ and all $F \in \mathcal{F}_s$;

If we have ‘ \leq ’ or ‘ \geq ’ in Property 3, then we speak of a sub- or super-martingale, respectively.

Remark 4.2. a) Property 3 in Definition 4.1 can be equivalently expressed as $\mathbb{E}(X_t | \mathcal{F}_s) = X_s$ using the conditional expectation $\mathbb{E}(\cdot | \mathcal{F}_s)$. The present formulation has the advantage, that it naturally extends the notion of martingales to general σ -additive measure spaces – which is useful for many applications of martingales in real analysis, see e.g. Stein [167, Ch. IV]. Note that most martingale results, in particular convergence theorems and inequalities, remain valid in the σ -additive setting, see Schilling [149, Ch. 22–25, 27] for a systematic presentation. An interesting side effect is that this also allows one to treat martingales without formally introducing conditional expectations right from the beginning.

b) If $t \mapsto X_t$ is not càdlàg but satisfies the Properties 1–3 from Definition 4.1, it is always possible to find a càdlàg modification; this is due to the ‘usual conditions’ for the filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F}_t)$. For sub-martingales one needs, in addition, that $t \mapsto \mathbb{E}X_t$ is right continuous. For a proof see Revuz & Yor [141, Ch. II.§2] or Schilling [152, Thm. 46.11].

Since we focus on maximal inequalities, we assume the reader to be familiar with the standard martingale convergence results, see e.g. Revuz & Yor [141] or Schilling [152]. Throughout this chapter, we use the following standard notation for maximal functions:

$$X_t^* := \sup_{s \leq t} |X_s| \quad \text{and} \quad X^* := \sup_{t \in I} X_t^* = \sup_{t \in I} |X_t|. \quad (4.1)$$

4.1. Doob’s maximal inequalities

The basis of most martingale maximal inequalities is the following lemma.

Lemma 4.3 (Doob–Kolmogorov maximal estimate). *Let $(X_t, \mathcal{F}_t)_{t \in I}$ be a submartingale. For every $t \in I$ and $\lambda > 0$*

$$\mathbb{P}(X_t^* > \lambda) \leq \frac{1}{\lambda} \int_{\{X_t^* \geq \lambda\}} X_t^+ d\mathbb{P}. \quad (4.2)$$

If $(X_t, \mathcal{F}_t)_{t \in I}$ is a martingale or a positive submartingale such that $X_t \in L^p(\mathbb{P})$ for some $p \geq 1$, then

$$\mathbb{P}(X_t^* > \lambda) \leq \frac{1}{\lambda^p} \int_{\{X_t^* \geq \lambda\}} |X_t|^p d\mathbb{P}. \quad (4.3)$$

Proof. Fix $t \in I$, $\lambda > 0$ and define the stopping time $\tau := \inf\{s \leq t \mid X_s \geq \lambda\}$; as usual, $\inf \emptyset := \infty$. Note that

$$\{\tau < t\} \subseteq \{X_t^* > \lambda\} \subseteq \{\tau \leq t\} \subseteq \{X_t^* \geq \lambda\}.$$

These inclusions, along with the Markov inequality and the submartingale property for the stopped submartingale $(X_{t \wedge \tau}^+)_{t \in I}$ yield

$$\begin{aligned} \mathbb{P}(X_t^* > \lambda) &\leq \mathbb{P}(\tau \leq t, X_{\tau \wedge t}^+ \geq \lambda) \\ &\leq \frac{1}{\lambda} \int_{\{\tau \leq t\}} X_{\tau \wedge t}^+ d\mathbb{P} \\ &\leq \frac{1}{\lambda} \int_{\{\tau \leq t\}} X_t^+ d\mathbb{P} \\ &\leq \frac{1}{\lambda} \int_{\{X_t^* \geq \lambda\}} X_t^+ d\mathbb{P}. \end{aligned}$$

The estimate (4.3) follows from (4.2) since for a (positive sub-)martingale satisfying $\mathbb{E}[|X_t|^p] < \infty$ the process $(|X_t|^p, \mathcal{F}_t)$ is a submartingale. \square

Combining the estimates from Lemma 4.3 with the good- λ techniques from Section 2.3 leads to Doob's maximal L^p inequalities.

Theorem 4.4 (Doob). *Let $(X_t, \mathcal{F}_t)_{t \in I}$ be a martingale or a positive submartingale. Then*

$$\mathbb{E}[(X^*)^p] \leq \begin{cases} \left(\frac{p}{p-1}\right)^p \sup_{t \in I} \mathbb{E}[|X_t|^p], & \text{if } p > 1, \\ \frac{e}{e-1} \sup_{t \in I} (1 + \mathbb{E}[|X_t| \log^+ |X_t|]), & \text{if } p = 1, \\ \frac{1}{1-p} \sup_{t \in I} (\mathbb{E}|X_t|)^p, & \text{if } 0 < p < 1. \end{cases} \quad (4.4)$$

If the terms on the right hand side of (4.4) are finite, $X^ \in L^p(\mathbb{P})$ and the limit $X_\infty = \lim_{t \rightarrow \infty} X_t$ exists a.s.; if $p \geq 1$ the limit exists also in $L^p(\mathbb{P})$ and $\sup_{t \in I} \mathbb{E}[|X_t|^p] = \mathbb{E}[|X_\infty|^p]$.*

Proof. The estimates in (4.4) follow if we combine Lemma 4.3 and Lemma 2.21 or Lemma 2.22, respectively, using $X = X_t^*$, $Y = X_t$, and taking the supremum over all $t \in I$.

If the right hand side of (4.4) is finite, the family $(X_t)_{t \in I}$ is L^1 -bounded, and the a.s. convergence $X_t \rightarrow X_\infty$ follows from the standard martingale convergence theorem. If $p \geq 1$, we also have uniform integrability, because of the inequality $|X_t|^p \leq (X_t^*)^p \in L^1(\mathbb{P})$, and we get L^p -convergence from Vitali's convergence theorem, see e.g. Schilling [149, Thms. 22.9, 22.7]; in particular, $\sup_{t \in I} \mathbb{E}[|X_t|^p] = \mathbb{E}[|X_\infty|^p]$. \square

Using Lemma 2.23 we can obtain the following partial converse of the $L \log L$ -case (i.e. for $p = 1$) of Theorem 4.4. Originally, this assertion is due to Stein [165] in connection with the Hardy–Littlewood maximal function, see also Torchinsky [173, Ch. IV.5]; the probabilistic point of view is Gundy's [53], and our presentation follows Chersi [21].

Theorem 4.5 (Stein; Gundy). *Let $(X_t, \mathcal{F}_t)_{t \in I}$ be a uniformly integrable martingale such that $X_t \geq 0$. Assume that there exist constants $a, b > 0$ and a sequence $t_0 < t_1 < \dots < t_n < \dots < t_\infty = \infty$ such that*

$$X_0 \leq a \quad \text{and} \quad X_{t_n} \leq bX_{t_{n-1}}, \quad \text{a.s. and for all } n \in \mathbb{N}.$$

If $\mathbb{E}[X^] < \infty$, then $\mathbb{E}[X_\infty \log^+ X_\infty] < \infty$ for $X_\infty := \lim_{t \rightarrow \infty} X_t$.*

Proof. Since $(X_t)_{t \in I}$ is uniformly integrable, the limit $X_\infty = \lim_{t \rightarrow \infty} X_t$ exists a.s. and in $L^1(\mathbb{P})$. Set $Y := \sup_{n \in \mathbb{N}} X_{t_n} \leq X^*$ and define stopping times

$$\tau := t_\nu, \quad \nu := \inf \{n \mid X_{t_n} > \lambda\}, \quad \inf \emptyset = \infty.$$

By assumption,

$$X_{t_\nu} \leq bX_{t_{\nu-1}} \leq b\lambda$$

for all $\lambda > a$, and this estimate still holds if $\nu = \infty$. Since $\{Y > \lambda\} = \{t_\nu < \infty\}$, we can use optional stopping to see for all $\lambda > a$

$$\int_{\{Y > \lambda\}} X_\infty d\mathbb{P} = \int_{\{t_\nu < \infty\}} X_{t_\nu} d\mathbb{P} \leq b\lambda \mathbb{P}(t_\nu < \infty) = b\lambda \mathbb{P}(Y > \lambda).$$

From this the assertion follows with Lemma 2.23. \square

There is an interesting variant of the Doob–Kolmogorov maximal estimate in Lemma 4.3. It is a continuous-time version of the following inequality due to Chow [22] and Birnbaum & Marshall [12, Thm. 2.1]: Let $(X_n)_{n \in \mathbb{N}}$ be a positive submartingale and $(c_n)_{n \in \mathbb{N}}$ a decreasing sequence. For every $\lambda \geq 0$

$$\begin{aligned} \lambda \mathbb{P}\left(\max_{1 \leq k \leq n} c_k X_k > \lambda\right) &\leq \sum_{k=1}^{n-1} (c_k - c_{k+1}) \mathbb{E}[X_k] + c_n \mathbb{E}[X_n] \\ &= c_1 \mathbb{E}[X_1] + \sum_{k=2}^n c_k (\mathbb{E}[X_k] - \mathbb{E}[X_{k-1}]). \end{aligned} \tag{4.5}$$

Lemma 4.6. *Let $(X_t, \mathcal{F}_t)_{t \geq 0}$, $X_0 = 0$, be a positive submartingale with finite second moments $m(t) := \mathbb{E}[X_t^2]$, and $c: [0, \infty) \rightarrow [0, \infty)$ a right continuous, decreasing function. Then*

$$\mathbb{P}\left(\sup_{0 \leq s \leq t} c(s)X_s > 1\right) \leq \int_0^t c^2(s) dm(s), \quad t > 0. \quad (4.6)$$

Proof. Fix $t > 0$ and define $t_i := t_{i,n} := (it)/n$ for $i = 0, 1, \dots, n$ and $n \in \mathbb{N}$. By right continuity and with the inequality (4.5) applied to the submartingale $(X_{t_i}^2, \mathcal{F}_{t_i})_{i=0, \dots, n}$ we get

$$\begin{aligned} \mathbb{P}\left(\sup_{0 \leq s \leq t} c(s)X_s > 1\right) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\max_{i=1, \dots, n} c(t_i)X_{t_i} > 1\right) \\ &\leq \lim_{n \rightarrow \infty} \sum_{i=1}^n c^2(t_i) (\mathbb{E}[X_{t_i}^2] - \mathbb{E}[X_{t_{i-1}}^2]) \\ &= \int_0^t c^2(s) dm(s). \quad \square \end{aligned}$$

We will close this first part with two applications of Doob's maximal inequalities.

Example 4.7 (Kolmogorov's SLLN). Let us briefly sum up the standard proof of Kolmogorov's strong law of large numbers using martingales with an index set running to the left ('backwards or reverse martingale'), see e.g. Schilling [149, Ex. 24.8] or Schilling [152, Thms. 45.5, 45.6]:

Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of identically distributed, integrable and independent (or exchangeable) random variables, then $S_n := (X_1 + \dots + X_n)/n$ tends to $\mathbb{E}X_1$ a.s. and in L^1 -sense.

The key observation in the argument is that

$$M_{-n} := \frac{X_1 + \dots + X_n}{n}, \quad \mathcal{F}_{-n} := \sigma(S_n, X_{n+1}, X_{n+2}, \dots)$$

is a martingale with backwards directed index set. Such martingales are automatically uniformly integrable, and so the limit $M_{-n} \rightarrow M_{-\infty}$ exists a.s. and in L^1 . In order to identify the limit as $M_{-\infty} = \mathbb{E}X_1$, one uses either Kolmogorov's zero-one law (in the iid case) or the Hewitt-Savage zero-one law (in the exchangeable case).

There is also the usual converse: *If a sequence of iid random variables is such that $\lim_{n \rightarrow \infty} \frac{1}{n}(X_1 + \dots + X_n)$ exists a.s., then $\mathbb{E}[|X_1|] < \infty$.*

The standard proof – along with a short argument for positive random variables – can be found in Schilling [149, continuation of Ex. 24.8, pp. 296–7]. We are now interested in the following converse:

If, in the above setting, $X_1 \geq 0$ and $\mathbb{E}[M^*] = \mathbb{E}\left[\sup_{n \in \mathbb{N}} \frac{1}{n}(X_1 + \cdots + X_n)\right] < \infty$, then $\mathbb{E}[X_1 \log^+ X_1] < \infty$.

This follows at once from Theorem 4.5 – suitably modified for martingales with backwards directed index set: The roles of X_0 and X_∞ in this theorem are now played by $M_{-\infty} = \mathbb{E}X_1$ and M_1 , respectively. Moreover,

$$M_{-n} = \frac{X_1 + \cdots + X_n}{n} \leq \frac{n+1}{n} \frac{X_1 + \cdots + X_n + X_{n+1}}{n+1} \leq 2M_{-n-1}$$

shows that the condition of Theorem 4.5 holds with $a = \mathbb{E}X_1 > 0$ and $b = 2$. Thus, $\mathbb{E}[X_1 \log^+ X_1] < \infty$.

Example 4.8 (Hardy–Littlewood maximal functions). Write $C = [0, 1]^d$ for the half-open unit cube in \mathbb{R}^d , λ for Lebesgue measure on $(C, \mathcal{B}(C))$, and define the dyadic σ -algebras

$$\mathcal{F}_k := \sigma(\mathcal{G}_k), \quad \mathcal{G}_k := \left\{ (j_1, \dots, j_d)2^{-k} + [0, 2^{-k})^d \mid \begin{array}{l} j_i = 0, \dots, 2^k - 1, \\ i = 1, \dots, d \end{array} \right\}.$$

Clearly, $\mathcal{G}_\infty := \bigcup_{k \in \mathbb{N}} \mathcal{G}_k$ generates the Borel sets $\mathcal{B}(C)$. Each $f \in L^1(\lambda) = L^1(C, \lambda)$ gives rise to a uniformly integrable martingale by

$$f_k(x) := \mathbb{E}[f \mid \mathcal{F}_k](x) = \sum_{Q \in \mathcal{G}_k} \frac{1}{\lambda(Q)} \int_Q f \, d\lambda \mathbf{1}_Q(x),$$

and we can define the *dyadic maximal function* f^\diamond as

$$f^\diamond(x) := \sup_{k \in \mathbb{N}} \mathbb{E}[|f| \mid \mathcal{F}_k](x) = \sup_{Q \in \mathcal{G}_\infty, Q \ni x} \frac{1}{\lambda(Q)} \int_Q |f| \, d\lambda.$$

The usual *Hardy–Littlewood maximal function* is given by

$$f^*(x) = \sup_{B_r(x), r \leq 1} \frac{1}{\lambda(B)} \int_B |f| \, d\lambda, \quad x \in C;$$

if necessary, we extend f periodically to \mathbb{R}^d and use Lebesgue measure λ on \mathbb{R}^d . In the definition of f^* it does not matter whether the balls are centred around x or just contain them, and we also may only consider balls with rational centres and rational radii containing x , see Schilling [149, Lem. 25.16].

Theorem. *The following estimates for the Hardy–Littlewood maximal function hold true: For any $f \in L^1(C, \lambda)$*

$$\|f^*\|_{L^p} \leq \begin{cases} c_d \frac{p}{p-1} \|f\|_{L^p}, & \text{if } p \in (1, \infty); \\ c_d(1 + \|f \log^+ |f|\|_{L^1}), & \text{if } p = 1. \end{cases} \quad (4.7)$$

Moreover, $\|f^*\|_{L^p} < \infty$ if, and only if, we have $\|f \log^+ |f|\|_{L^1} < \infty$ ($p = 1$) or $\|f\|_{L^p} < \infty$ ($1 < p < \infty$), respectively.

If we can show that the dyadic maximal function f^\diamond satisfies $f^\diamond \geq cf^*$, then the estimates (4.7) follow from Doob's inequalities (4.4). The trouble is, that a ball $B_r(x)$ of radius $r < \frac{1}{2}2^{-k}$, $k = -1, 0, 1, 2, 3, \dots$, need not entirely fall into any single square of our lattice \mathcal{G}_k , and it even need not be entirely contained in C .

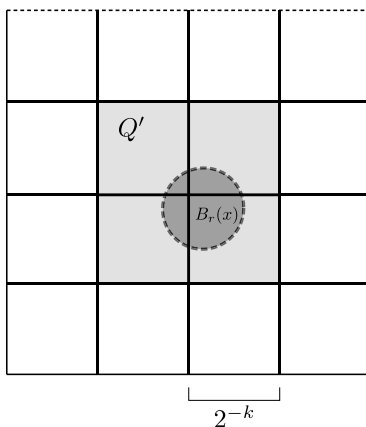


FIGURE 2. The picture shows a square from \mathcal{G}_{k-2} which is subdivided into four resp. sixteen squares from \mathcal{G}_{k-1} and \mathcal{G}_k , respectively. A ball $B_r(x)$ of radius $\frac{1}{2}2^{-k-1} \leq r < \frac{1}{2}2^{-k}$, $k = -1, 0, 1, 2, \dots$, need not entirely fall into any single square of side-length 2^{-k} , but it hits at most 2^d such squares (shaded), each of them is a corner of a square from \mathcal{G}_{k-1} with side-length 2^{1-k} , and all of them are in the square from \mathcal{G}_{k-2} with side-length 2^{2-k} . Note that the pictures shows the worst possible case as $B_r(x)$ may already fall into a square from \mathcal{G}_{k-1} (or even \mathcal{G}_k).

The latter problem can be fixed by extending f and \mathcal{G}_k , hence f_k and \mathcal{F}_k , periodically. This allows us to reduce everything to the situation where $B_r(x)$, $\frac{1}{2}2^{-k-1} \leq r < \frac{1}{2}2^{-k}$, is within C , but fails to be inside a single $Q \in \mathcal{G}_k$. In this case, $B_r(x)$ intersects at most 2^d of the sets in \mathcal{G}_k (see Fig. 2). Denote by $Q' \in \mathcal{G}_{k-2}$ a cube of side length $4 \cdot 2^{-k}$ such that $B_r(x) \subseteq Q'$. Then,

$$\frac{1}{\lambda(B_r(x))} \int_{B_r(x)} |f| d\lambda \leq \frac{\lambda(Q')}{\lambda(B_r(x))} \frac{1}{\lambda(Q')} \int_{Q'} |f| d\lambda \leq c_d f^\diamond(x).$$

The constant c_d comes from the comparison of $\lambda(Q')$ and $\lambda(B_r(x))$, and depends only on the space dimension as $\frac{1}{2}2^{-k-1} \leq r < \frac{1}{2}2^{-k}$; this constant is of the order 4^d .

This proves $f^*(x) \leq c_d f^\diamond(x)$. A similar reasoning also gives the converse inequality $f^\diamond(x) \leq c'_d f^*(x)$.

In order to see the second part of the assertion, we note first that the martingale convergence theorem gives $\lim_{k \rightarrow \infty} f_k(x) = f(x)$ for almost all x . Note that this essentially proves Lebesgue's differentiation theorem. Therefore, $\|f\|_{L^p} \leq \|f^\diamond\|_{L^p}$ follows easily for all $p \in [1, \infty)$. The assertion for $p = 1$ follows with Theorem 4.5, if we use $|f|$ instead of f .

A further Doob-type maximal result on martingales is discussed on connection with BMO spaces in Theorem 3.10.

Example 4.9 (The ‘sharp’ maximal function). We continue Example 4.8. Recall that $C = [0, 1)^d$ is the half-open unit cube in \mathbb{R}^d , λ is Lebesgue measure on $(C, \mathcal{B}(C))$, and $\mathcal{F}_k = \sigma(\mathcal{G}_k)$ is the σ -algebra generated by the dyadic cubes \mathcal{G}_k in C with side-length 2^{-k} . For $f \in L^1(C, \lambda)$ we define the *sharp maximal function* as

$$f^\#(x) := \sup_{Q \ni x} \frac{1}{\lambda(Q)} \int_Q |f - f_Q| d\lambda \quad \text{where} \quad f_Q := \frac{1}{\lambda(Q)} \int_Q |f| d\lambda;$$

the supremum ranges over all axis-parallel cubes $Q \subset C$ such that $x \in Q$. Compared with the Hardy–Littlewood maximal function f^* , we have now an additional centering with f_Q . This is useful if we want to deal with BMO-spaces, see Remark 3.11; trivially $f \in \text{BMO}$ if, and only if, $f^\# \in L^\infty$. It is straightforward to see that

$$f^\#(x) \leq 2f^*(x) \quad \text{and, therefore,} \quad \|f^\#\|_{L^p} \leq \frac{2c_d p}{p-1} \|f\|_{L^p} \quad (1 < p < \infty)$$

using the results from Example 4.8. Again, $p = 1$ and $f^*, f^\# \in L^1$, requires $f \log^+ |f| \in L^1$. But now we also have the converse estimate.

Theorem (Fefferman–Stein [42]). *Let $f \in L^r(C, \lambda)$ for some $1 < r < \infty$. For every $r < p < \infty$ there is a constant $\kappa_{d,p}$ such that the sharp maximal function satisfies*

$$\|f\|_{L^p} \leq \|f^\circ\|_{L^p} \leq \kappa_{d,p} \|f^\#\|_{L^p}.$$

The proof of this theorem relies on the good- λ -inequality (2.27), which can be established in the following form

$$\lambda(\{x \in C \mid f^\circ(x) > 2\lambda, f^\#(x) < \gamma\lambda\}) \leq \gamma 2^d \lambda(\{x \in C \mid f^\circ(x) \geq \lambda\}),$$

for $\gamma \in (0, 1)$, see e.g. Grafakos [49, Thm. 7.4.4]. Take $\gamma < 2^{-p-d-1}$ and use Lemma 2.24 to conclude that

$$\|f^\circ\|_{L^p} \leq \frac{2 \cdot 2^{1/p}}{\gamma} \|f^*(x)\|_{L^p}, \quad \gamma < 2^{-p-d-1},$$

and since $\|f\|_{L^p} \leq \|f^\circ\|_{L^p}$ – cf. the last paragraph in the previous Example 4.8 – the claim follows.

The sharp maximal function was introduced by Fefferman & Stein [42] in order to study interpolation pairs when one end-point is the space BMO. A detailed modern presentation is given by Grafakos [49, Ch. 7.4], see also Torch-

insky [173, Ch. VIII.1,2]. Recently, Kamiński & Osękowski [71] have found an optimal bound for the constant $\kappa_{1,p}$ in dimension $d = 1$.

4.2. Khintchine and Burkholder–Davis–Gundy inequalities

Before we tend to the general case of the Burkholder–Davis–Gundy inequalities, we want to discuss an inequality which is frequently used in analysis. As it turns out it is related to the concentration of measure method, see Section 2.2, as well as to Doob’s maximal inequality, cf. Theorem 4.4. The following inequality is Azuma’s [2] generalization of Hoeffding’s inequality [57] for sums of bounded, independent random variables.

Lemma 4.10 (Azuma). *Let $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}_0}$ be a martingale such that the increments are bounded: $|X_n - X_{n-1}| \leq c_n$ for a sequence $(c_n)_{n \in \mathbb{N}} \subseteq [0, \infty)$. Set $d_n^2 := c_1^2 + \dots + c_n^2$. Then*

$$\mathbb{P}(X_n - X_0 \geq \lambda) \leq \exp \left[-\frac{\lambda^2}{2d_n^2} \right], \quad \lambda \geq 0, \quad n \in \mathbb{N}, \quad (4.8)$$

$$\mathbb{P}(|X_n - X_0| \geq \lambda) \leq 2 \exp \left[-\frac{\lambda^2}{2d_n^2} \right], \quad \lambda \geq 0, \quad n \in \mathbb{N}. \quad (4.9)$$

Further generalizations, e.g. if the c_n are random variables, are due to Hitczenko [56, Thm. 3.1, Lem. 4.3].

Proof. Since $(-X_n)_{n \in \mathbb{N}_0}$ is also a martingale with bounded increments, it is enough to prove (4.8) and add the resulting inequalities for $(\pm X_n)_{n \in \mathbb{N}_0}$ to get (4.9).

The function $e_\lambda(x) := e^{\lambda x}$, $x \in \mathbb{R}$, $\lambda \geq 0$, is convex. Thus, the graph of e_λ is on the segment $[-c, c]$ below the line connecting the points $(-c; e_\lambda(-c))$ and $(c; e_\lambda(c))$, i.e.

$$e^{\lambda x} \leq \frac{e^{\lambda c} - e^{-\lambda c}}{2c} x + \frac{1}{2} (e^{\lambda c} + e^{-\lambda c}), \quad x \in [-c, c].$$

Set $x = (X_n - X_{n-1})$ and $c = c_n$, and take on both sides conditional expectation $\mathbb{E}(\cdot \cdot \cdot | \mathcal{F}_{n-1})$. This gives

$$\begin{aligned} & \mathbb{E} \left(e^{\lambda(X_n - X_{n-1})} \mid \mathcal{F}_{n-1} \right) \\ & \leq \frac{e^{\lambda c_n} - e^{-\lambda c_n}}{2c_n} \mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) + \frac{e^{\lambda c_n} + e^{-\lambda c_n}}{2} \\ & = \frac{e^{\lambda c_n} + e^{-\lambda c_n}}{2} \leq e^{\lambda^2 c_n^2 / 2}. \end{aligned}$$

The tower property of conditional expectation and a pull out argument yield

$$\begin{aligned}
\mathbb{E} \left(e^{\lambda(X_n - X_0)} \right) &= \mathbb{E} \left(\prod_{i=1}^n e^{\lambda(X_i - X_{i-1})} \right) \\
&= \mathbb{E} \left(\prod_{i=1}^{n-1} e^{\lambda(X_i - X_{i-1})} \mathbb{E} \left[e^{\lambda(X_n - X_{n-1})} \mid \mathcal{F}_{n-1} \right] \right) \\
&\leq \mathbb{E} \left(\prod_{i=1}^{n-1} e^{t(X_i - X_{i-1})} \right) e^{\lambda^2 c_n^2 / 2} \\
&\leq \dots \leq e^{\lambda^2 d_n^2 / 2}.
\end{aligned}$$

Using the Markov inequality for $t, \xi \geq 0$ gives

$$\mathbb{P}(X_n - X_0 \geq \lambda) = \mathbb{P} \left(e^{\xi(X_n - X_0)} \geq e^{\lambda \xi} \right) \leq e^{-\lambda \xi} \mathbb{E} e^{\xi(X_n - X_0)} \leq e^{-\lambda \xi + \xi^2 d_n^2 / 2}.$$

Since the right hand side becomes minimal at $\xi = \lambda/d_n^2$, we get (4.8). \square

With the help of Azuma's inequality we can get *Khinchine's inequalities*. The upper bound in Theorem 4.11 was first proved in Khinchine [72], the beautiful 'duality argument' for the lower bound is due to Littlewood [114] and Paley & Zygmund [135]. In analysis, the iid random variables appearing in Theorem 4.11 are often represented by the Rademacher functions on the probability space $([0, 1], \mathcal{B}[0, 1], \lambda)$, hence the alternative name *Rademacher inequalities*.

Theorem 4.11 (Khinchine). *Let $(c_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}$ be a sequence of real numbers and $(\xi_n)_{n \in \mathbb{N}}$ iid Bernoulli random variables such that $\mathbb{P}(\xi_1 = \pm 1) = \frac{1}{2}$. For all $0 < p < \infty$*

$$\gamma_p \left(\sum_{i=1}^n c_i^2 \right)^{p/2} \leq \mathbb{E} \left(\left| \sum_{i=1}^n c_i \xi_i \right|^p \right) \leq \Gamma_p \left(\sum_{i=1}^n c_i^2 \right)^{p/2}, \quad n \in \mathbb{N}. \quad (4.10)$$

The constants $0 < \gamma_p \leq \Gamma_p < \infty$ depend only on p .

Proof. Upper estimate. $S_0 := 0$, $S_n := \xi_1 + \dots + \xi_n$ is a martingale, and so is the martingale transform

$$X_n := c \bullet S_n := \sum_{i=1}^n c_i (S_i - S_{i-1}) = \sum_{i=1}^n c_i \xi_i.$$

Azuma's inequality (Lemma 4.10) yields

$$\mathbb{P}(|X_n| \geq \lambda) \leq 2e^{-\lambda^2 / 2d_n^2}, \quad d_n^2 = c_1^2 + \dots + c_n^2,$$

and a direct calculation along the lines of Lemma 2.21 shows for all $0 < p < \infty$

$$\mathbb{E}(|X_n|^p) \leq 2p \int_0^\infty t^{p-1} e^{-t^2 / 2d_n^2} dt = p2^{p/2} \Gamma(p/2) (d_n^2)^{p/2},$$

i.e. $\Gamma_p := p2^{p/2} \Gamma(p/2)$.

Lower estimate. The following ‘duality’ trick is due to Littlewood [114]: Using independence, we see

$$\mathbb{E}(X_n^2) = \mathbb{V}X_n = \sum_{i=1}^n \mathbb{V}(c_i \xi_i) = \sum_{i=1}^n c_i^2,$$

and the lower estimate can be rewritten as

$$\gamma_p \|X_n\|_{L^2}^p \leq \|X_n\|_{L^p}^p.$$

This inequality is obvious if $p \geq 2$, since the $L^p(\mathbb{P})$ norm is increasing in p . If $p \in (0, 2)$, we write

$$2 = p\alpha + 4\beta \quad \text{with suitable } \alpha + \beta = 1, \alpha, \beta \in (0, 1),$$

and use Hölder’s inequality in conjunction with the upper estimate:

$$\begin{aligned} \mathbb{E}(|X_n|^2) &= \mathbb{E}(|X_n|^{p\alpha+4\beta}) \leq \{\mathbb{E}(|X_n|^p)\}^\alpha \{\mathbb{E}(|X_n|^4)\}^\beta \\ &\leq \{\mathbb{E}(|X_n|^p)\}^\alpha \Gamma_4^\beta \{\mathbb{E}(|X_n|^2)\}^{2\beta}. \end{aligned}$$

From this we get the lower estimate for $0 < p < 2$ with $\gamma_p = \Gamma_4^{1-2/p}$. \square

Remark 4.12. The proof of the upper bound in (4.10) uses Azuma’s inequality (4.9). The following argument shows that both inequalities are equivalent – up to some constant in the exponent.

The proof of Theorem 4.11 yields the constant $\Gamma_p := p2^{p/2}\Gamma(p/2)$ in the upper bound of the Khintchine inequality, which is not far off the known best constant $2^{p/2}\Gamma((p+1)/2)/\sqrt{\pi}$ (for $p > 2$), see Haagerup [54].

In order to simplify our calculations we use the following slightly rough estimate, which will, however, be good enough for the point we want to make. Since $ye^{-y} \leq e^{-1}$, we see for all $\lambda > 0$ and $\alpha, \beta \in (0, 1)$ with $\alpha + \beta = 1$

$$\begin{aligned} p\Gamma\left(\frac{p}{2}\right) &= 2\Gamma\left(\frac{p}{2} + 1\right) = 2 \int_0^\infty x^{p/2} e^{-x} dx \\ &= 2p \cdot p^{p/2} \int_0^\infty (ye^{-2\alpha y})^{p/2} e^{-p\beta y} dy \\ &\leq \frac{2}{\beta} p^{p/2} (2\alpha e)^{-p/2}. \end{aligned}$$

Thus, assuming the upper bound in (4.10), we see with Markov’s inequality

$$\mathbb{P}(|c \bullet S_n| > \lambda) \leq \frac{1}{\lambda^p} \mathbb{E}[|(c \bullet S)_n|^p] \leq \frac{2}{\beta} \left(\frac{d_n^2}{2\alpha e \lambda^2} p \right)^{p/2},$$

where $d_n^2 = c_1^2 + \dots + c_n^2$. As a function of p , the right hand side becomes minimal if $(d_n^2/\alpha e \lambda^2) p = 1/e$, i.e. $p = \alpha \lambda^2/d_n^2$. Thus,

$$\mathbb{P}(|c \bullet S_n| > \lambda) \leq \frac{2}{\beta} e^{-\alpha \lambda^2/(2d_n^2)}, \quad \lambda > 0, \alpha + \beta = 1, \alpha, \beta \in (0, 1),$$

which is essentially Azuma’s inequality.

Remark 4.13. The following argument explains, why (4.10) can be seen as a maximal inequality. If $\sum_{i=1}^{\infty} c_i^2 < \infty$, we infer from (4.10) that the martingale $\sum_{i=1}^n c_i \xi_i$ is uniformly integrable, hence we get⁶

$$\left(\mathbb{E} \left[\left| \sum_{i=1}^{\infty} c_i \xi_i \right|^p \right] \right)^{1/p} \asymp \left(\sum_{i=1}^{\infty} c_i^2 \right)^{1/2}, \quad 0 < p < \infty.$$

On the other hand, by Doob's maximal inequality,

$$\left(\mathbb{E} \left[\left| \sum_{i=1}^{\infty} c_i \xi_i \right|^p \right] \right)^{1/p} \asymp \left(\mathbb{E} \left[\sup_{n \in \mathbb{N}} \left| \sum_{i=1}^n c_i \xi_i \right|^p \right] \right)^{1/p}, \quad p > 1.$$

Finally, since the L^p -norm of a probability measure is monotone in $p \in (0, \infty)$, we see that (4.10) implies

$$\left(\sum_{i=1}^{\infty} c_i^2 \right)^{1/2} \asymp \left(\mathbb{E} \left[\left| \sum_{i=1}^{\infty} c_i \xi_i \right|^p \right] \right)^{1/p} \asymp \left(\mathbb{E} \left[\sup_{n \in \mathbb{N}} \left| \sum_{i=1}^n c_i \xi_i \right|^p \right] \right)^{1/p}.$$

Using the sequence $(c_1, c_2, \dots, c_n, 0, 0, \dots)$ brings us back to (4.10).

The Burkholder–Davis–Gundy (BDG) inequalities can be seen as a generalization of the Khintchine inequalities to martingales, which are more general than the random walk martingale $S_n = \xi_1 + \dots + \xi_n$, and to martingale transforms $X_n := c \bullet S_n := \sum_{i=0}^n c_i (S_i - S_{i-1})$ with a previsible – and not just deterministic – sequence $(c_n)_{n \in \mathbb{N}}$, $c_n = c_n(\omega)$. In order to appreciate this extension, it is helpful to interpret the term $\sum_{i=1}^n c_i^2$ appearing in Khintchine's inequalities as the *quadratic variation* (or *compensator* or *square function*) of the martingale $X = c \bullet S$. Indeed, the compensator $[S]_n$ of the simple random walk martingale $S_n = \xi_1 + \dots + \xi_n$ appearing in Theorem 4.11 is $[S]_n = n$, and so

$$[c \bullet S]_n = c^2 \bullet [S]_n = \sum_{i=1}^n c_i^2 ([S]_i - [S]_{i-1}) = \sum_{i=1}^n c_i^2.$$

This point of view allows us to find a continuous-time extension of Khintchine's inequalities. We will distinguish between two cases: (i) $(X_t)_{t \in I}$ is a discrete martingale or it is a continuous martingale whose paths are càdlàg (right continuous & finite left limits) with jumps, and (ii) $(X_t)_{t \geq 0}$ is a martingale in continuous time and with continuous paths. The latter situation is much easier to handle, since the previsible quadratic variation $\langle X \rangle = (\langle X \rangle_t)_{t \geq 0}$ and the adapted quadratic variation $[X] = ([X]_t)_{t \geq 0}$ coincide.

We will give two proofs of the BDG inequalities, first the standard proof based on Itô's formula and then an alternative approach using the fact that continuous martingales are time-changed Brownian motions – a further explanation as to why the continuous case is easier than the jump case or the discrete case.

⁶We use the symbol $f \asymp g$ to indicate two-sided estimates $cf \leq g \leq Cf$, with comparison constants $0 < c < C < \infty$ depending only on p .

Theorem 4.14 (Burkholder–Davis–Gundy). *Let $(X_t, \mathcal{F}_t)_{t \geq 0}$, $X_0 = 0$, be a martingale with continuous paths and (predictable) quadratic variation process $(\langle X \rangle_t)_{t \geq 0}$. For any $p > 0$ we have*

$$c_p \mathbb{E} \left(\langle X \rangle_\infty^{p/2} \right) \leq \mathbb{E} \left((X^*)^p \right) \leq C_p \mathbb{E} \left(\langle X \rangle_\infty^{p/2} \right). \quad (4.11)$$

The constants $0 < c_p \leq C_p < \infty$ depend only on $p \in (0, \infty)$.

Remark 4.15. a) Using a stopping argument $X_t \rightsquigarrow X_t^\tau := X_{t \wedge \tau}$ we can change (4.11) to a finite (deterministic or random) time horizon $[0, \tau]$. Moreover, (4.11) is always understood as taking values in $[0, \infty]$, i.e. if the right hand side is finite, then $X^* \in L^p$, hence $(X_t)_{t \geq 0} \subseteq L^p$. If $p \geq 1$ and $X^* \in L^p$, then $(X_t)_{t \geq 0}$ is uniformly integrable, and we get, as in Doob's inequality (Theorem 4.4), that $\mathbb{E}[|X_\infty|^p]$ and $\mathbb{E}[(X^*)^p]$ are comparable.

b) There is a huge literature on the optimality of the constants in the BDG and related inequalities. For this we refer, in particular, to the monograph by Osełkowski [134]. Recent developments can be found in Kamiński & Osełkowski [71]. An approach to BDG-inequalities via the Garsia–Rodemich–Rumsey lemma is due to Barlow & Yor [3] yielding more general BDG-type inequalities for continuous (semi-)martingales.

Standard proof of Theorem 4.14: Gettoor & Sharpe [46]. By stopping with the stopping time

$$\tau_n := \inf \{ s \geq 0 \mid |X_s|^2 + \langle X \rangle_s \geq n \},$$

we can, without loss of generality, assume that both X and $\langle X \rangle$ are bounded,⁷ in particular X is uniformly integrable, and X_∞ is well-defined.

Assume first that $p \geq 2$ and denote by $q = p/(p-1)$ the conjugate exponent. Combining the fact that $X^2 - \langle X \rangle$ is a martingale with Doob's maximal inequality (Theorem 4.4) proves (4.11) for $p = 2$:

$$\mathbb{E}[\langle X \rangle_\infty] = \mathbb{E}[|X_\infty|^2] \leq \mathbb{E}[(X^*)^2] \leq 4 \sup_{s \geq 0} \mathbb{E}[|X_s|^2] = 4\mathbb{E}[\langle X \rangle_\infty].$$

Since $x \mapsto |x|^p$, $p > 2$, is a C^2 -function, Itô's formula, yields

$$|X_t|^p = p \int_0^t |X_s|^{p-1} \operatorname{sgn}(X_s) dX_s + \frac{1}{2} p(p-1) \int_0^t |X_s|^{p-2} d\langle X \rangle_s.$$

Because of the boundedness of $|X_s|^{p-1}$, the first integral on the right hand side is a martingale, and we see with Doob's maximal inequality and the Hölder inequality for $p/(p-2)$ and $p/2$

⁷As usual, the problem is the overshoot when crossing the level n . Since $t \mapsto X_t$ and $t \mapsto \langle X \rangle_t$ are continuous, we trivially have $|X_{\tau_n} - X_{\tau_n-}| = 0$ and $\langle X \rangle_{\tau_n} - \langle X \rangle_{\tau_n-} = 0$, but this will change dramatically for a jump process.

$$\begin{aligned}
\mathbb{E} \left[\sup_{s \geq 0} |X_s|^p \right] &\leq q^p \frac{p(p-1)}{2} \mathbb{E} \left[\int_0^\infty |X_s|^{p-2} d\langle X \rangle_s \right] \\
&\leq q^p \frac{p(p-1)}{2} \mathbb{E} \left[\sup_{s \geq 0} |X_s|^{p-2} \langle X \rangle_\infty \right] \\
&\leq q^p \frac{p(p-1)}{2} \left(\mathbb{E} \left[\sup_{s \geq 0} |X_s|^p \right] \right)^{1-2/p} \left(\mathbb{E} \left[\langle X \rangle_\infty^{p/2} \right] \right)^{2/p}.
\end{aligned}$$

Dividing by $(\mathbb{E} [\sup_{s \geq 0} |X_s|^p])^{1-2/p}$ gives the upper estimate in (4.11) with $C_p = [q^p p(p-1)/2]^{p/2}$.

In order to prove the lower estimate in (4.11) we define an auxiliary martingale $Y_t := \int_0^t \langle X \rangle_s^{(p-2)/4} dX_s$. Observe that

$$\langle Y \rangle_t = \int_0^t \langle X \rangle_s^{(p-2)/2} d\langle X \rangle_s = \frac{2}{p} \langle X \rangle_t^{p/2}$$

and

$$\mathbb{E} \left[\langle X \rangle_t^{p/2} \right] = \frac{p}{2} \mathbb{E} [\langle Y \rangle_t] = \frac{p}{2} \mathbb{E} [|Y_t|^2].$$

The Itô formula for $f(x, y) = xy$ and the bivariate process $(X_t, \langle X \rangle_t^{(p-2)/4})$ gives

$$\begin{aligned}
X_t \langle X \rangle_t^{(p-2)/4} &= \int_0^t \langle X \rangle_s^{(p-2)/4} dX_s + \int_0^t X_s d\langle X \rangle_s^{(p-2)/4} \\
&= Y_t + \int_0^t X_s d\langle X \rangle_s^{(p-2)/4}.
\end{aligned}$$

If we rearrange this equality, we get $|Y_t| \leq 2 \sup_{s \leq t} |X_s| \langle X \rangle_t^{(p-2)/4}$. Finally, using Hölder's inequality with $p/2$ and $p/(p-2)$, shows

$$\begin{aligned}
\frac{2}{p} \mathbb{E} \left[\langle X \rangle_t^{p/2} \right] &= \mathbb{E} [|Y_t|^2] \leq 4 \mathbb{E} \left[\sup_{s \leq t} |X_s|^2 \langle X \rangle_t^{(p-2)/2} \right] \\
&\leq 4 \left(\mathbb{E} \left[\sup_{s \leq t} |X_s|^p \right] \right)^{2/p} \left(\mathbb{E} \left[\langle X \rangle_t^{p/2} \right] \right)^{1-2/p}.
\end{aligned}$$

Divide both sides by $(\mathbb{E} [\langle X \rangle_t^{p/2}])^{1-2/p}$ and let $t \rightarrow \infty$ to get the lower bound in (4.11) with $c_p = (2p)^{-p/2}$.

Now we assume that $0 < p < 2$. In this case, we use the good- λ argument from Lemma 2.26, setting $X = (X^*)^2$ and $Y = \langle X \rangle_\infty$. Let $\tau := \inf \{s \geq 0 \mid \langle X \rangle_s \geq \lambda\}$. We have

$$P := \mathbb{P} \left((X^*)^2 \geq \lambda, \langle X \rangle_\infty < \lambda \right) = \mathbb{P} \left((X^*)^2 \geq \lambda, \tau = \infty \right) \leq \mathbb{P} \left((X_\tau^*)^2 \geq \lambda \right),$$

and with optional stopping, Doob's maximal inequality (4.4) and the definition of τ , we find

$$P \leq \frac{1}{\lambda} \mathbb{E} \left(|X_\tau|^2 \right) = \frac{1}{\lambda} \mathbb{E} (\langle X \rangle_\tau) = \frac{1}{\lambda} \mathbb{E} (\langle X \rangle_\infty \wedge \lambda).$$

Now we can apply Lemma 2.26 and we get

$$\mathbb{E} [(X^*)^p] \leq \frac{4-p}{2-p} \mathbb{E} [\langle X \rangle_\infty^{p/2}].$$

This proves the upper bound in (4.11) for $0 < p < 2$.

For the lower bound we use again Lemma 2.26, this time with $X = \langle X \rangle_\infty$ and $Y = (X^*)^2$. Define $\sigma := \inf\{s > 0 \mid |X_s|^2 \geq \lambda\}$. Then

$$P' := \mathbb{P} \left(\langle X \rangle_\infty \geq \lambda, (X^*)^2 < \lambda \right) \leq \mathbb{P} (\langle X \rangle_\infty \geq \lambda, \sigma = \infty) \leq \mathbb{P} (\langle X \rangle_\sigma > \lambda).$$

Using the Markov inequality and the fact that $X^2 - \langle X \rangle$ is a martingale, we get

$$P' \leq \frac{1}{\lambda} \mathbb{E} [\langle X \rangle_\sigma] \leq \frac{1}{\lambda} \mathbb{E} [(X_\sigma^*)^2] \leq \frac{1}{\lambda} \mathbb{E} [(X_\sigma^*)^2 \wedge \lambda],$$

and Lemma 2.26 gives

$$\mathbb{E} [(X^*)^p] \geq \frac{2-p}{4-p} \mathbb{E} [\langle X \rangle_\infty^{p/2}],$$

which is the lower bound in (4.11) for $0 < p < 2$. \square

Time change proof of Theorem 4.14. ⁸By the Dambis & Dubins–Schwarz theorem, we can represent any continuous martingale $(X_t)_{t \in I}$ with $X_0 = 0$ as a time-changed Brownian motion, see e.g. Schilling [151, Thm. 19.17] and Revuz & Yor [141, Thm. V.1.6]; in fact, we have $X_t = B_{\langle X \rangle_t}$, for some Brownian motion $(B_t)_{t \in I}$ which is, in general, not independent of $(X_t)_{t \in I}$. Therefore, it is enough to show (4.11) for Brownian motion $(B_t)_{t \geq 0}$ evaluated at some stopping time τ . That is, we have to show that

$$c_p \mathbb{E} [\tau^{p/2}] \leq \mathbb{E} [(B_\tau^*)^p] \leq C_p \mathbb{E} [\tau^{p/2}]. \quad (4.11')$$

Let $\sigma := \inf\{t \geq 0 \mid B_t^* > \lambda\}$. We have for all $\lambda > 0$ and $\delta \in (0, 1)$

$$\begin{aligned} \mathbb{P} (B_\tau^* > 2\lambda, \sqrt{\tau} < \delta\lambda) &= \mathbb{P} (B_\tau^* > 2\lambda, \sqrt{\tau} < \delta\lambda, \sigma < \tau) \\ &\leq \mathbb{P} \left(\sup_{\sigma \leq t \leq \tau} |B_t - B_\sigma| > \lambda, \sqrt{\tau} < \delta\lambda, \sigma < \tau \right) \\ &\leq \mathbb{P} \left(\sup_{\sigma \leq t \leq \delta^2 \lambda^2} |B_t - B_\sigma| > \lambda, \sigma < \tau \right) \\ &= \mathbb{E} \left[\mathbb{P}^{B_\sigma} \left(\sup_{t \leq \delta^2 \lambda^2} |B_t - B_0| > \lambda \right) \mathbf{1}_{\{\sigma < \tau\}} \right] \\ &\leq \sup_{x \in \mathbb{R}} \mathbb{P}^x \left(\sup_{t \leq \delta^2 \lambda^2} |B_t - x| > \lambda \right) \mathbb{P}(\sigma < \tau). \end{aligned}$$

⁸It seems that Burkholder & Gundy [17, pp. 109–110 & §7] were the first to point out this strategy.

In the penultimate step we use the strong Markov property of Brownian motion. Since Brownian motion is translation invariant and scales like $B_{\lambda^2 t} \sim \lambda B_t$, we get

$$\mathbb{P}(B_\tau^* > 2\lambda, \sqrt{\tau} < \delta\lambda) \leq \mathbb{P}(B_{\delta^2}^* > 1) \mathbb{P}(\sigma < \tau) \leq \epsilon \mathbb{P}(B_\tau^* \geq \lambda).$$

Here we use that for a Brownian motion $\mathbb{P}(B_{\delta^2}^* > 1) \rightarrow 0$ as $\delta \rightarrow 0$. The upper estimate in (4.11') now follows from Lemma 2.24.

The lower estimate can be shown in a similar fashion, switching the roles of $\sqrt{\tau}$ and B_τ^* . We have

$$\begin{aligned} \mathbb{P}(\sqrt{\tau} > 2\lambda, B_\tau^* < \delta\lambda) &\leq \mathbb{P}\left(\sqrt{\tau} > 2\lambda, \sup_{\lambda^2 \leq t \leq 2\lambda^2} |B_t - B_{\lambda^2}| < 2\delta\lambda\right) \\ &\leq \mathbb{E}\left[\mathbf{1}_{\{\sqrt{\tau} > \lambda\}} \mathbb{P}^{B_{\lambda^2}}\left(\sup_{t \leq \lambda^2} |B_t - B_0| < 2\delta\lambda\right)\right] \\ &\leq \mathbb{P}(\tau > \lambda^2) \sup_{x \in \mathbb{R}} \mathbb{P}^x\left(\sup_{t \leq \lambda^2} |B_t - x| < 2\delta\lambda\right) \\ &\leq \mathbb{P}(B_1^* < 2\delta) \mathbb{P}(\tau > \lambda^2) \\ &\leq \epsilon \mathbb{P}(\tau > \lambda^2). \end{aligned}$$

using again the Markov property, translation invariance and Brownian scaling. A further application of Lemma 2.24 gives the lower bound in (4.11). \square

Remark 4.16. The time-change technique can also be used to get sharp ‘deviation’ estimates for continuous martingales. This complements our earlier Remark 4.12 on the role of Azuma’s inequalities in relation to Khintchine’s inequality.

Let $(X_t)_{t \geq 0}$ be a continuous martingale and write it as $X_t - X_0 = B_{\langle X \rangle_t}$. Assume that $\langle X \rangle_T \leq d_T^2$ a.s. for some $T > 0$ and a constant d_T^2 . Using the scaling property of Brownian motion and the reflection principle we get

$$\begin{aligned} \mathbb{P}((X - X_0)_T^* > \lambda) &= \mathbb{P}(B_{\langle X \rangle_T}^* > \lambda) \leq \mathbb{P}(B_{d_T^2}^* > \lambda) \\ &\leq \mathbb{P}(B_1^* > \lambda/d_T) \\ &= 2\mathbb{P}(B_1 > \lambda/d_T). \end{aligned}$$

With the usual Gaussian tail estimate, we finally get

$$\mathbb{P}((X - X_0)_T^* > \lambda) \leq 2(2\pi)^{-1/2} \frac{d_T}{\lambda} e^{-\lambda^2/(2d_T^2)}.$$

This is sometimes called McKean’s inequality, see Krylov [86, Ch. IV.§2].

Remark 4.17. If $X_t = \int_0^t f(s) dB_s$ is a stochastic integral with respect to Brownian motion satisfying $\mathbb{E}[\langle X \rangle_t^{p/2}] = \mathbb{E}\left[\left(\int_0^t |f(s)|^2 ds\right)^{p/2}\right] < \infty$ for some $p > 1$, then Novikov [131] showed that

$$c_p \mathbb{E}[\langle X \rangle_t^{p/2}] \leq \mathbb{E}[|X_t|^p] \leq C_p \mathbb{E}[\langle X \rangle_t^{p/2}] \quad (4.12)$$

for some constants $c_p, C_p > 0$. Hence, by the BDG inequalities,

$$\mathbb{E} \left[\langle X \rangle_t^{p/2} \right] \asymp \sup_{s \leq t} \mathbb{E} [|X_s|^p] \asymp \mathbb{E} \left[\sup_{s \leq t} |X_s|^p \right].$$

While the upper bound in (4.12) holds for any $p > 0$, cf. Novikov [131], the lower bound breaks down for $p = 1$. This can be seen by considering the martingale $X_t := B_{t \wedge \tau}$ with $\tau := \inf\{t \geq 0 \mid B_t = 1\}$. On the one hand, we have

$$\mathbb{E} [|X_t|] = \mathbb{E} [2B_{t \wedge \tau}^+ - B_{t \wedge \tau}] = 2\mathbb{E} [B_{t \wedge \tau}^+] \leq 2.$$

On the other hand, $\mathbb{E} [\langle X \rangle_t^{1/2}] = \mathbb{E} [(t \wedge \tau)^{1/2}] \rightarrow \mathbb{E} [\tau^{1/2}] = \infty$ as $t \rightarrow \infty$,⁹ cf. Novikov [133], so that the lower bound cannot hold for $p = 1$.

The BDG inequalities for jump processes In the final part of this section, we will discuss the BDG inequalities for processes with jumps. Let us first explain why the presence of jumps creates difficulties.

Our first proof of the BDG inequalities for processes (see p. 424 *et seq.*) with continuous sample paths heavily relied on a stopping argument: we could assume without loss of generality that the process X and its quadratic variation $\langle X \rangle$ are bounded by considering the stopped process X^τ for the stopping time $\tau = \{s \geq 0 \mid |X_s|^2 + \langle X \rangle_s \geq n\}$. This reasoning is no longer applicable if $t \mapsto X_t$ has jumps. While $|X_s|^2 + \langle X \rangle_s < n$ is still true for $s < \tau$, we do not have any bound on $|X_\tau|^2$ because the jumps create an overshoot which cannot be controlled.¹⁰ The same kind of problem arises for discrete-time processes $(X_n)_{n \in \mathbb{N}}$, and, in fact, the BDG inequalities for discrete-time processes are essentially as hard to derive as the BDG inequalities for jump processes in continuous time.

From now on we will focus on continuous-time processes with jumps; an exposition of the discrete-time setting can be found in Schilling [150].

Example 4.18. Among the simplest examples of martingales with jumps are the compound Poisson processes. For a probability distribution μ on $\mathbb{R} \setminus \{0\}$, let H_1, H_2, \dots be iid random variables with $H_i \sim \mu$, and let $(N_t)_{t \geq 0}$ be an independent Poisson process with intensity $\mathbb{E}N_1 = \mathbb{V}N_1 = 1$. Then the sample paths of the *compound Poisson process*

$$J_t := \sum_{i=1}^{N_t} H_i, \quad t \geq 0,$$

are piecewise constant with jumps occurring whenever the value of N_t increases by 1. The compound Poisson process $(J_t)_{t \geq 0}$ has stationary and independent

⁹A simple proof that $\mathbb{E}[\tau^a] = \infty$ for all $a \geq \frac{1}{2}$ (for random walks) can be found in Schilling [150, Bsp. 11.10, p. 115–116]. Alternatively, one can use a direct calculation and the fact that $\tau \sim (2\pi s^3)^{-1/2} e^{-1/2s} ds$, see e.g. Schilling [151, Thm. 6.10]

¹⁰We note, in passing, that also the time-change proof of the BDG inequalities, see 426 *et seq.*, breaks down. Although there are quite general embedding results for right continuous (semi-)martingales into Brownian motion, see Monroe [129, 130] and Dubins [35] or Schilling [151, Thm. 14.7] for the discrete-time case, the stopping times appearing in the embedding frequently have unbounded moments of order $p/2$.

increments, i.e. it is a Lévy process, see Section 5.1, Scholium 2.1, and also Schilling [148, Chapter 3]. Using the independence and stationarity of the increments, it is not difficult to show that $X_t := J_t - bt$ with $b = \mathbb{E}(J_1) = \int y \mu(dy)$ is a martingale with quadratic variation $\langle X \rangle_t = t \int y^2 \mu(dy)$; of course, we assume that the first and second moments exist. A naive application of Theorem 4.14 would give

$$\mathbb{E} \left(\sup_{s \leq t} |X_s|^p \right) \asymp t^{p/2} \left(\int y^2 \mu(dy) \right)^{p/2}, \quad t \geq 0, p > 0, \quad (4.13)$$

but this estimate fails to hold. In fact, it can happen that the left hand side is infinite while the right hand side is finite. To see this, let us consider $p = 4$. The moment $\mathbb{E}(|X_t|^4)$ exists if, and only if, the characteristic function $\phi(\xi) = \mathbb{E}e^{i\xi X_t}$ is four times differentiable at $\xi = 0$. Since $\phi(\xi) = \exp(-t\psi(\xi))$ with $\psi(\xi) = \int_{y \neq 0} (1 - e^{iy\xi} - iy\xi) \mu(dy)$, we find that

$$\begin{aligned} \mathbb{E}(|X_t|^4) < \infty &\iff \psi \text{ is four times differentiable at } \xi = 0 \\ &\iff \int_{|y| \geq 1} |y|^4 \mu(dy) < \infty, \end{aligned} \quad (4.14)$$

see e.g. Kühn [88, Lem. 4.3] for a detailed proof or, indeed Theorem 5.19 further down. This already indicates that (4.13) cannot hold for $p = 4$. In fact, a somewhat lengthy, but easy, calculation yields

$$\mathbb{E}(X_t^4) = \frac{d^4}{d\xi^4} \phi(\xi) \Big|_{\xi=0} = 3t^2 \left(\int y^2 \mu(dy) \right)^2 + t \int y^4 \mu(dy).$$

While the first term on the right hand side looks familiar from (4.13), the second one is new. As (4.14) shows, this second term is actually needed to ensure that $\mathbb{E}(X_t^4)$ is finite.

The following version of the BDG inequalities for (possibly discontinuous) martingales is due to Dzhaparidze & Valkeila [37, Lem. 2.1].

Theorem 4.19. *Let $(X_t)_{t \geq 0}$ be a real-valued martingale with càdlàg sample paths and $X_0 = 0$. For every $p \geq 2$, there are constants $0 < c_p \leq C_p < \infty$ such that*

$$c_p \mathbb{E} \left[\langle X \rangle_t^{p/2} + |(\Delta X)_t^*|^p \right] \leq \mathbb{E} [(X_t^*)^p] \leq C_p \mathbb{E} \left[\langle X \rangle_t^{p/2} + |(\Delta X)_t^*|^p \right]; \quad (4.15)$$

here $(\Delta X)_t = X_t - X_{t-}$ is the jump height at time t .

Using a standard stopping argument, Theorem 4.19 can be extended to stopping times τ .

There are two further variants of the BDG inequalities for discontinuous martingales. We begin with Novikov's [132] result, which applies to processes which can be written as a stochastic integral with respect to a Poisson random measure; this includes, for instance, Lévy processes and solutions to stochastic differential equations driven by Lévy processes, see Sections 5 and 6. In contrast to Theorem 4.19, the result by Novikov gives moment estimates for any $p > 0$.

Theorem 4.20 (Novikov). *Let $(X_t)_{t \geq 0}$ be a one-dimensional stochastic process of the form*

$$X_t = \int_0^t \int_{y \neq 0} F(s, y) \tilde{N}(dy, ds), \quad t \geq 0,$$

for a predictable stochastic process F and a compensated Poisson random measure $\tilde{N} = N - \hat{N}$ with compensator $\hat{N}(dy, ds) = \nu(dy) ds$ for a measure ν satisfying $\int_{y \neq 0} \min\{1, |y|^2\} \nu(dy) < \infty$.

1. If $\mathbb{E} \left[\int_0^T \int_{y \neq 0} |F(s, y)| \nu(dy) ds \right] < \infty$ for some $T > 0$, then

$$\mathbb{E} \left(\sup_{t \leq T} |X_t|^p \right) \leq C_{p, \alpha} \mathbb{E} \left[\left(\int_0^T \int_{y \neq 0} |F(s, y)|^\alpha \nu(dy) ds \right)^{p/\alpha} \right] \quad (4.16)$$

for any $\alpha \in [1, 2]$ and $p \in [0, \alpha] \subseteq [0, 2]$.

2. If $\mathbb{E} \left(\int_0^T \int_{y \neq 0} |F(s, y)|^2 \nu(dy) ds \right) < \infty$ for some $T > 0$, then

$$\begin{aligned} \mathbb{E} \left(\sup_{t \leq T} |X_t|^p \right) &\leq c_p \mathbb{E} \left[\left(\int_0^T \int_{y \neq 0} |F(s, y)|^2 \nu(dy) ds \right)^{p/2} \right] \\ &\quad + c_p \mathbb{E} \left[\int_0^T \int_{y \neq 0} |F(s, y)|^p \nu(dy) ds \right] \end{aligned} \quad (4.17)$$

for all $p \geq 2$.

The estimate (4.17) for $p \geq 2$ can be derived from Theorem 4.19 or by applying Itô's formula to $x \mapsto x^p$, see e.g. Kunita [98, Thm. 2.11]. For $p \in (0, 2)$, one can show that it suffices to consider $p = \alpha$ and then establish the estimate for this particular case using a change of variable formula which is, so to speak, the substitute for Itô's formula, cf. Novikov [132].

The other possibility to get BDG inequalities for a jump process $X = (X_t)_{t \geq 0}$ is to replace the predictable compensator $\langle X \rangle$ by the square bracket process

$$[X]_t := \lim_{n \rightarrow \infty} \left(X_0^2 + \sum_{i=1}^{k(n)} (X_{t \wedge \tau_i^n} - X_{t \wedge \tau_{i-1}^n})^2 \right) \quad (4.18)$$

(the limit is taken in probability) where $0 = \tau_0^n \leq \tau_1^n \leq \dots \leq \tau_{k(n)}^n \leq t$ are stopping times such that $\max_{1 \leq i \leq k(n)} |\tau_i^n - \tau_{i-1}^n| \rightarrow 0$ as $n \rightarrow \infty$. A full discussion is given in Protter [139, Ch. II.6]; here we just mention that $[X]_t$ is an adapted increasing process such that $[X]_t - \langle X \rangle_t$ is a martingale. Moreover, one has the following relation to stochastic integration with jumps:

$$\Delta[X]_t = (\Delta X)_t^2 \quad \text{and} \quad [X]_t - [X]_0 = X_t^2 - X_0^2 - 2 \int_0^t X_{s-} dX_s. \quad (4.19)$$

Example 4.22 below shows why we must restrict ourselves to $p \geq 1$ in the following 'square-bracket' version of the BDG inequalities.

Theorem 4.21. *Let $(X_t)_{t \geq 0}$ be a real-valued martingale with càdlàg sample paths and $X_0 = 0$. For every $p \geq 1$ there are universal constants $0 < c_p \leq C_p < \infty$ such that*

$$c_p \mathbb{E} \left[[X]_t^{p/2} \right] \leq \mathbb{E} [(X_t^*)^p] \leq C_p \mathbb{E} \left[[X]_t^{p/2} \right], \quad t \geq 0. \quad (4.20)$$

It goes without saying that, as in all cases discussed before, (4.20) is stable under stopping and under the limit as $t \rightarrow \infty$. Theorem 4.21 goes back to the seminal paper [19] by Burkholder, Davis & Gundy, see also Burkholder [16] for a simplified version in discrete time. The first continuous-time proof, including the hitherto unpublished ‘Davies’ case $p = 1$, seems to be due to Meyer [125, App. II]. Full proofs of Theorem 4.21 are fairly difficult to find in textbooks, notable exceptions are Cohen & Elliott [23, Thm. 11.5.5], Liptser & Shiryaev [113, pp. 70–83] or Schilling [150, Ch. 10] in the discrete-time setting; usually Delacherie & Meyer [27, p. 304, §VII.92], [28, pp. 328–329, XIII.14], Meyer [125, App. II, III] and [126, pp. 350–353] and Lenglart *et al.* [104, §2] serve as standard references. Marinelli & Röckner [121] present a proof of Theorem 4.21 along the lines of the stochastic calculus proof of the BDG inequalities for jump processes which also works for Hilbert-space valued random variables. As one would expect, the key ingredient is the careful (conditional) domination of the jumps. If the underlying martingale is given by a stochastic integral, the BDG inequalities for $p = 2$ are just Itô’s isometry; Dirksen [32] discusses generalizations of Itô’s isometry to L^p -spaces with $1 < p < \infty$.

Burkholder [16, §10–12] points out that it is the impact of big jumps which causes the breakdown of the BDG inequalities if $p \in (0, 1)$. This paper also contains an interesting discussion on how to tame the jumps. The following example is due to Marcinkiewicz & Zygmund [119], see also Burkholder & Gundy [17]; we follow the presentation in Schilling [150, Bsp. 10.5].

Example 4.22 (Marcinkiewicz & Zygmund). In order to illustrate the breakdown of the BDG inequalities, it is enough to consider a discrete-time martingale $(X_n)_{n \in \mathbb{N}_0}$; the analogue of the jumps are the increments $\xi_n := X_n - X_{n-1}$.

Let η be a random variable with probability law $\mathbb{P}(\eta = 1) = 1 - 1/m$ and $\mathbb{P}(\eta = 1 - m) = 1/m$ for some fixed $m \geq 2$. We construct sequences of independent random variables $(\xi_n)_{n \in \mathbb{N}}$ such that the inequalities (4.20) fail for the ‘random walk’ martingale $X_0 := 0$, $X_n := \xi_1 + \dots + \xi_n$ if $0 < p < 1$.

Breakdown of the lower bound Let $\xi_1 \sim \eta$, $\xi_2 \sim -\eta$ and $\xi_n = 0$ for $n \geq 3$ be independent random variables. For $n \geq 2$

$$\begin{aligned} \mathbb{E} (|X_n|^p) &= \mathbb{E} (|\xi_1 + \xi_2|^p) = \mathbb{E} (|\xi_1 + \xi_2|^p \mathbf{1}_{\{\xi_1 \neq -\xi_2\}}) \\ &= 2\mathbb{E} (|\xi_1 + \xi_2|^p \mathbf{1}_{\{\xi_1 = 1\} \cap \{\xi_2 = m-1\}}) \\ &\leq 2m^p \mathbb{P}(\xi_2 = m-1) = 2m^{p-1}. \end{aligned}$$

However, we always have $\xi_1^2 + \xi_2^2 \geq 2$, and so

$$\mathbb{E} \left([X]_\infty^{p/2} \right) = \mathbb{E} \left((\xi_1^2 + \xi_2^2)^{p/2} \right) \geq 1.$$

If $m \gg 1$ is large enough, the lower bound in (4.20) is violated.

Breakdown of the upper bound For $k, n \in \mathbb{N}$ let $\xi_1 \sim \eta, \dots, \xi_n \sim \eta$ and $\xi_{n+k} = 0$ be independent random variables. If $m \geq 2n$, then

$$|X_{n+k}| = \left| \sum_{i=1}^n \xi_i \right| \geq \begin{cases} n, & \forall i \in \{1, \dots, n\} : \xi_i = 1, \\ \frac{\geq (m-1) - (n-1)}{|\xi_j| - \sum_{i \neq j} |\xi_i|} \geq n, & \exists j \in \{1, \dots, n\} : \xi_j = 1 - m. \end{cases}$$

Moreover, $[X]_{n+k} = \sum_{i=1}^n \xi_i^2$ satisfies

$$\mathbb{P}([X]_{n+k} = n) = \left(1 - \frac{1}{m}\right)^n \quad \text{and} \quad \mathbb{P}(n < [X]_{n+k} \leq nm^2) = 1 - \left(1 - \frac{1}{m}\right)^n.$$

On the one hand, we have $\mathbb{E}(|X_{n+k}|^p) \geq n^p$, and on the other hand

$$\mathbb{E}\left([X]_{n+k}^{p/2}\right) \leq n^{p/2} \left(1 - \frac{1}{m}\right)^n + n^{p/2} m^p \left(1 - \left(1 - \frac{1}{m}\right)^n\right) \xrightarrow{m \rightarrow \infty} n^{p/2}.$$

Thus, the upper bound in (4.20) cannot hold if $m, n \gg 1$.

Remark 4.23. Let us briefly indicate an extension of the BDG inequalities which is contained in Lenglart *et al.* [104] and Meyer [28]. Rather than considering the moment $F(X_t^*)$ for the polynomially growing function $F(x) = |x|^p$, $p \geq 1$, one may consider convex functions $F: [0, \infty) \rightarrow [0, \infty)$ which are of *moderate growth*, i.e. functions that satisfy

$$\exists c > 0 : \sup_{x > 0} \frac{F(cx)}{F(x)} < \infty.$$

All concave functions are moderate. If F is convex with right derivative $f = F'_+$, then the above condition is equivalent to

$$\alpha = \sup_{x > 0} \frac{xf(x)}{F(x)} < \infty \quad \text{which implies that} \quad \forall c > 1 : \sup_{x > 0} \frac{F(cx)}{F(x)} \leq c^\alpha.$$

The BDG inequalities for convex moderate functions read

$$\mathbb{E}[F(X_t^*)] \asymp \mathbb{E}\left[F\left(\sqrt{[X]_t}\right)\right], \quad t \geq 0.$$

The proof relies, essentially, on the good- λ -inequality and a clever use of the layer-cake formula for moderate functions.

Using the optimization trick from Remark 4.12 it is easy to deduce exponential tail estimates from the BDG inequalities. The restriction $\alpha \lambda^2 \geq d^2(t)$ comes from the fact that the BDG inequalities hold, in general, only for $p \geq 1$. The following result is due to Kallenberg & Sztencel [70, Thm. 5.1, 5.3].¹¹

¹¹Note that Kallenberg & Sztencel [70] treat high-dimensional martingales and aim to get uniform constants for all dimensions. Thus, the proofs are much more complicated, and the resulting estimates are ‘only’ sub-Gaussian or require further symmetry assumptions to get Gaussian tails.

Corollary 4.24. *Let $(X_t)_{t \geq 0}$ be a real-valued càdlàg martingale such that the square bracket process is bounded, i.e. $[X]_t \leq d^2(t)$ for some increasing function $d(t) > 0$. Then*

$$\mathbb{P}(X_t^* > \lambda) \leq \frac{2}{\beta} e^{-\alpha \lambda^2 / d^2(t)}, \quad \lambda > \frac{d^2(t)}{\alpha}, \quad \alpha + \beta = 1, \quad \alpha, \beta \in (0, 1). \quad (4.21)$$

5. Lévy processes

Brownian motion is arguably one of the most important stochastic processes in continuous time. The class of Lévy processes shares many features with Brownian motion, and it is reasonable to treat Brownian motion in the wider context of Lévy processes. In many cases, ‘Brownian proofs’ extend to the Lévy setting almost literally, and in such cases we sometimes point out ‘Brownian’ references, leaving the obvious adaptations to the reader. Below we often work in dimension $d = 1$. Let us point out, however, that most results below remain valid for d -dimensional processes at the expense of larger, usually dimension-dependent, constants.

5.1. Background knowledge on Lévy processes

Definition 5.1. A Lévy process is a stochastic process $(X_t)_{t \geq 0}$ taking values in \mathbb{R}^d and such that its increments are stationary (L1), independent (L2) and the trajectories are continuous in probability (L3):

$$\forall s \leq t : \quad X_t - X_s \sim X_{t-s} \quad (\text{L1})$$

$$\forall n \in \mathbb{N}, t_0 \leq \dots \leq t_n : \quad (X_{t_k} - X_{t_{k-1}})_{k=1}^n \text{ independent} \quad (\text{L2})$$

$$\forall \epsilon > 0 : \quad \lim_{h \downarrow 0} \mathbb{P}(|X_h| > \epsilon) = 0 \quad (\text{L3})$$

Our standard reference for Lévy processes is Sato’s monograph [143], a short introduction is given in Schilling [148].

Note that (L1) implies that $X_0 \sim \delta_0$, i.e. $\mathbb{P}(X_0 = 0) = 1$; this is often separately stated as a further defining condition. Let us add a few further remarks.

Remark 5.2 (role of Brownian motion). Among all Lévy processes, Brownian motion is special for (i) its law is Gaussian, (ii) its paths $t \mapsto B_t$ are (a.s.) continuous, and (iii) its infinitesimal generator is a second-order differential operator with constant coefficients. In fact, the properties (i)–(iii) are equivalent. The implication (i) \Rightarrow (ii) is the Kolmogorov–Chentsov theorem (e.g. Revuz & Yor [141, Thm. I.(2,1)], Schilling [151, Thm. 10.1] or Theorem 3.2), (ii) \Rightarrow (i) is due to Lévy (e.g. Itô [64, Sec. 4.§2], Schilling [148, Thm. 8.4], it is essentially a uniform version of the central limit theorem), for (ii) \Rightarrow (iii) see e.g. Schilling [151, Thm. 7.38], and (iii) \Rightarrow (i) is a standard fact on the fundamental solution of 2nd order differential operators – just use the Fourier transform.

Remark 5.3 (filtration and measurability issues). In the definition of a Lévy process we have implicitly used the natural filtration $\mathcal{F}_t^X := \sigma(X_s, s \leq t)$ generated by the process itself. It is not hard to see that (L2) is equivalent to

$$\forall s \leq t : X_t - X_s \text{ is independent of } \mathcal{F}_s^X. \quad (\text{L2}')$$

If we want to regularize the sample paths, it is important to add null sets to the filtration. To do so, we complete the underlying probability space (if needed – but we do not change notation) and denote the family of \mathbb{P} -null sets in the complete space by \mathcal{N} . The filtration is then enlarged by setting $\mathcal{F}_t := \sigma(\mathcal{F}_t^X, \mathcal{N})$. Since we are adding only null sets and sets of full measure, it is clear that (L2'), hence (L2), remains valid; thus $(X_t, \mathcal{F}_t)_{t \geq 0}$ is indeed a *Lévy process with filtration*. One can show that this enlargement is right continuous, i.e. $\mathcal{F}_t = \mathcal{F}_{t+}$, see e.g. Protter [139, Thm. I.31], Revuz & Yor [141, Thm. III.(2.10)] or Schilling [151, Ch. 6.7]; thus, $(X_t)_{t \geq 0}$ can be defined on a filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}_t, \mathbb{P})$ which fulfils the *usual conditions*.

In this setting, we can construct a modification of $(X_t, \mathcal{F}_t)_{t \geq 0}$ such that for all ω one has $X_0(\omega) = 0$ and (L3) is (up to a modification) equivalent to

$$\forall \omega \in \Omega : t \mapsto X_t \text{ is right continuous with finite left limits (càdlàg)}. \quad (\text{L3}')$$

We refer to Revuz & Yor [141, Thm. III.(2.7)] for the proof.

Remark 5.4 (strong Markov property). The independent and stationary increments property implies that the process $Y_t := X_{t+s} - X_s$ is, for each fixed $s \geq 0$, again a Lévy process, which is independent of \mathcal{F}_s , and that $(Y_t)_{t \geq 0}$ and $(X_t)_{t \geq 0}$ have the same finite dimensional distributions.

The independence of $(Y_t)_{t \geq 0}$ from the past \mathcal{F}_s allows us to realize a Lévy process as a Markov process. To do so, we define probability measures \mathbb{P}^x , $x \in \mathbb{R}^d$, on $(\Omega, \mathcal{F}_\infty)$ as projective limit of the finite dimensional distributions

$$\begin{aligned} \mathbb{P}^x(X_{t_k} \in B_k, k = 1, \dots, n) &:= \mathbb{P}(X_{t_k} + x \in B_k, k = 1, \dots, n), \\ 0 \leq t_1 < \dots < t_n, n \in \mathbb{N}, B_1, \dots, B_n &\in \mathcal{B}(\mathbb{R}^d). \end{aligned}$$

Then \mathbb{P}^x and $\mathbb{P}^\mu(\dots) := \int \mathbb{P}^x(\dots) \mu(dx)$ are the laws of the process $(X_t)_{t \geq 0}$ with initial distribution $X_0 \sim \delta_x$ and $X_0 \sim \mu$, respectively; we frequently use \mathbb{P} instead of \mathbb{P}^0 .

In the last two paragraphs we may replace the fixed time $s \geq 0$ by an \mathcal{F}_t -stopping time σ :

$$\begin{aligned} (X_{t+\sigma} - X_\sigma)_{t \geq 0} \text{ is, conditional on } \{\sigma < \infty\}, \text{ a Lévy process,} \\ \text{which is independent of } \mathcal{F}_\sigma \text{ and distributed like } (X_t)_{t \geq 0}. \end{aligned} \quad (5.1)$$

The property (5.1) is often called the *strong Markov property* of a Lévy process and, using standard arguments, one can see that it implies the usual formulation of the strong Markov property

$$\mathbb{E}^x [f(X_{t+\sigma}) \mid \mathcal{F}_\sigma] = \mathbb{E}^{X_\sigma} [f(X_t)] \quad \mathbb{P}^x\text{-a.s. on } \{\sigma < \infty\} \text{ for } f \in B_b(\mathbb{R}^d)$$

see e.g. Schilling [151, Ch. 6.2].

An elegant way to characterize Lévy processes can be based on the conditional characteristic function, see Schilling [148, Thm. 3.1].

Theorem 5.5. *A stochastic process $(X_t, \mathcal{F}_t)_{t \geq 0}$ with values in \mathbb{R}^d and càdlàg paths is a Lévy process with filtration $(\mathcal{F}_t)_{t \geq 0}$ if, and only if,*

$$\mathbb{E} \left[e^{i\xi \cdot (X_t - X_s)} \mid \mathcal{F}_s \right] = e^{-(t-s)\psi(\xi)}, \quad \xi \in \mathbb{R}^d, \quad s \leq t, \quad (5.2)$$

with a characteristic exponent $\psi: \mathbb{R}^d \rightarrow \mathbb{C}$ which is uniquely determined by the Lévy–Khintchine formula

$$\begin{aligned} \psi(\xi) = & -ib \cdot \xi + \frac{1}{2} \xi \cdot Q \xi + \int_{|y| \geq 1} (1 - e^{i\xi \cdot y}) \nu(dy) \\ & + \int_{0 < |y| < 1} (1 - e^{i\xi \cdot y} + i\xi \cdot y) \nu(dy). \end{aligned} \quad (5.3)$$

The Lévy triplet (b, Q, ν) appearing in (5.3) is unique and it comprises a vector $b \in \mathbb{R}^d$, a positive semidefinite matrix $Q \in \mathbb{R}^{d \times d}$ and a Radon measure ν on $\mathbb{R}^d \setminus \{0\}$ such that $\int_{y \neq 0} |y|^2 \wedge 1 \nu(dy) < \infty$.

If we take expectations on both sides of the formula (5.2), we see immediately that the random variables X_t are infinitely divisible, see also the discussion at the beginning of Section 2.1. In fact, and this can also be deduced from (5.2) and the structure of infinitely divisible laws, each infinitely divisible random variable Z can be ‘embedded’ into a Lévy process $(X_t)_{t \geq 0}$ such that $X_1 \sim Z$. For further details we refer to Sato [143, Ch. 2, §7–8] and Schilling [148, Ch. 2, Ch. 7].

Since a Lévy process is a translation invariant Markov process, the transition semigroup is given by $P_t f(x) = \mathbb{E}^x f(X_t) = \mathbb{E} f(X_t + x)$. The next theorem explains the structure of the infinitesimal generator A of the semigroup $P_t = e^{tA}$ and the structure of the sample paths. Proofs can be found in Sato [143, §19–20 and Thm. 31.5] and Schilling [148, Ch. 6–7].

We will use $\widehat{f}(\xi) := (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\xi \cdot y} f(y) dy$ to denote the Fourier transform of the function f .

Theorem 5.6. *Let $(X_t, \mathcal{F}_t)_{t \geq 0}$ be a Lévy process with values in \mathbb{R}^d and characteristic exponent ψ given by (5.3). The infinitesimal generator of the process is a pseudo differential operator*

$$Af(x) = -\psi(D)f(x) = - \int_{\mathbb{R}^d} \psi(\xi) \widehat{f}(\xi) d\xi, \quad f \in C_c^\infty(\mathbb{R}^d), \quad (5.4)$$

which can also be written as an integro-differential operator

$$\begin{aligned} Af(x) = & b \cdot \nabla f(x) + \frac{1}{2} \operatorname{tr} (Q \cdot \nabla^2 f(x)) + \int_{|y| \geq 1} (f(x+y) - f(x)) \nu(dy) \\ & + \int_{0 < |y| < 1} (f(x+y) - f(x) - y \cdot \nabla f(x)) \nu(dy), \quad f \in C_c^2(\mathbb{R}^d). \end{aligned} \quad (5.5)$$

The sample paths of a Lévy process enjoy the following Lévy-Itô decomposition

$$X_t = bt + \sqrt{Q}W_t + \sum_{\substack{0 < s \leq t \\ |\Delta X_s| \geq 1}} \Delta X_s + \lim_{\epsilon \downarrow 0} \left[\sum_{\substack{0 < s \leq t \\ \epsilon \leq |\Delta X_s| < 1}} \Delta X_s - t \int_{\epsilon \leq |y| < 1} y \nu(dy) \right] \quad (5.6)$$

where the four terms are Lévy processes, which are independent and whose Lévy exponents are given by the four terms in the representation (5.3) of ψ . The limit appearing in (5.6) is a limit in probability.

5.2. Reflection-type maximal inequalities

The reflection principle is one of the first results one comes across when studying Brownian motion or random walks. It is, of course, a maximal (in-)equality and its proof is closely related to maximal inequalities for Lévy processes.

Let $(X_t)_{t \geq 0}$ be a real-valued stochastic process. As before, $X_t^* := \sup_{s \leq t} |X_s|$ denotes the absolute maximum, and we write $M_t := \sup_{s \leq t} X_t$ for the running maximum. By $\tau_a := \inf \{s > 0 \mid X_s \geq a\}$ we denote the first passage time at the level a (also: the first hitting time of the closed set $[a, \infty)$).

Scholium 5.7 (classical reflection principle). Let us briefly review the proof of the *reflection principle* for a one-dimensional Brownian motion $(B_t)_{t \geq 0}$. For any $a > B_0 = 0$ and $t > 0$ we have $\{\tau_a \leq t\} = \{M_t \geq a\}$ and, using the strong Markov property,

$$\begin{aligned} \mathbb{P}^0(M_t \geq a) &= \mathbb{P}^0(M_t \geq a, B_t < a) + \mathbb{P}^0(M_t \geq a, B_t \geq a) \\ &= \mathbb{P}^0(M_t \geq a, B_t < a) + \mathbb{P}^0(B_t \geq a) \\ &= \mathbb{P}^0(\tau_a \leq t, B_t < a) + \mathbb{P}^0(B_t \geq a) \\ &= \int_{\tau_a \leq t} \mathbb{P}^{B_{\tau_a}(\omega)}(B_{t-\tau_a(\omega)} < a) \mathbb{P}^0(d\omega) + \mathbb{P}^0(B_t \geq a). \end{aligned}$$

Since Brownian motion is invariant under translations, and $B_{\tau_a} = a$, we get

$$\begin{aligned} \mathbb{P}^{B_{\tau_a}(\omega)}(B_{t-\tau_a(\omega)} < a) &= \mathbb{P}^0(B_{t-\tau_a(\omega)} < a - B_{\tau_a}(\omega)) \\ &= \mathbb{P}^0(B_{t-\tau_a(\omega)} < 0) = \frac{1}{2} \end{aligned} \quad (5.7)$$

and, using again $\{\tau_a \leq t\} = \{M_t \geq a\}$, we see that

$$\mathbb{P}^0(M_t \geq a) = 2\mathbb{P}^0(B_t \geq a) = \mathbb{P}^0(|B_t| \geq a), \quad a > 0. \quad (5.8)$$

In the second equality we use the symmetry of Brownian motion: $B_t \sim -B_t$.

Recall that a stochastic process $(X_t)_{t \geq 0}$ is called *self-similar* if for any $a > 0$ there is some $b > 0$ such that the processes $(X_{at})_{t \geq 0}$ and $(bX_t)_{t \geq 0}$ have the same finite dimensional distributions. In particular, $\mathbb{P}(X_t \geq 0) = \mathbb{P}(X_1 \geq 0) = p_0$ does not depend on $t > 0$. A Lévy process $(X_t)_{t \geq 0}$ is said to be *strictly stable*, if

each random variable X_t is stable, i.e. if $X_t \sim b(t)X_1$ for a suitable $b(t) > 0$. For Lévy processes the notions of *self-similarity* and *strict stability* coincide and, in this case, $b(t) = t^{1/\alpha}$ for some ‘stability index’ $\alpha \in (0, 2]$, see Schilling [148, Ch. 3] and Sato [143, Ch. 3.§13]. The following observation is from Meerschaert *et al.* [124, Thm. 6.1].

Lemma 5.8. *Let $(X_t)_{t \geq 0}$ be a strictly stable (hence, self-similar) Lévy process which is spectrally negative, i.e. it does not have positive jumps. Set $p_0 := \mathbb{P}(X_1 \geq 0)$. Then*

$$\mathbb{P}^0(M_t \geq a) = \frac{1}{1-p_0} \mathbb{P}^0(X_t \geq a), \quad a > 0. \quad (5.9)$$

Proof. We can follow the proof of the reflection principle for Brownian motion: First $\{M_t \geq a\} = \{\tau_a \leq t\}$ and $X_{\tau_a} = a$ follows from the fact that $(X_t)_{t \geq 0}$ does not have positive jumps, i.e. it can cross any level a , coming from below, only by ‘creeping’. Thus we only have to pay attention to (5.7). Due to self-similarity, we have now $\mathbb{P}(X_{t-\tau_a(\omega)} < 0) = \mathbb{P}(X_1 < 0) = 1 - p_0$ instead of $\frac{1}{2}$.¹² \square

The proof of the reflection principle (Scholium 5.7) can also be tweaked to give Lévy’s inequalities. These inequalities were originally shown by Lévy [106, Ch. VI.§44], see also Loève [116, Vol. 1, Sec. 18.C], when studying the convergence of sums of independent random variables. Here we present a variant for symmetric Lévy processes, i.e. a Lévy process with symmetric probability distributions: $X_t \sim -X_t$ for all $t > 0$.

Lemma 5.9. *Let $(X_t)_{t \geq 0}$ be a symmetric real-valued Lévy process. Then*

$$\mathbb{P}^0(M_t > a) \leq 2\mathbb{P}^0(X_t \geq a) \quad \text{and} \quad \mathbb{P}^0(X_t^* > a) \leq 2\mathbb{P}^0(|X_t| \geq a). \quad (5.10)$$

Proof. Since we still have $\{M_t > a\} \subseteq \{\tau_a \leq t\}$, we can follow the proof in Scholium 5.7 up to (5.7). Since $X_{\tau_a} \geq a$, this identity now becomes an inequality:

$$\mathbb{P}^{X_{\tau_a}(\omega)}(X_{t-\tau_a(\omega)} \leq a) = \mathbb{P}^0(X_{t-\tau_a(\omega)} \leq \underbrace{a - X_{\tau_a}(\omega)}_{\leq 0}) \leq \frac{1}{2}$$

and we get the first part of (5.10). Using symmetry, we have

$$\mathbb{P}^0(X_t^* > a) \leq \mathbb{P}^0\left(\sup_{s \leq t} X_s > a\right) + \mathbb{P}^0\left(\sup_{s \leq t} (-X_s) > a\right),$$

and $2\mathbb{P}^0(X_t \geq a) = \mathbb{P}^0(|X_t| \geq a)$; this reduces things to the first estimate. \square

Remark 5.10. Lemma 5.9 holds, of course, for Brownian motion. In this case, the first inequality in (5.10) is actually an equality. To get an exact expression for $\mathbb{P}^0(B_t^* > a)$ one needs more care. Using Lévy’s ‘triple law’ (see Schilling [151, Thm. 6.19 & p. 207] – this is also a ‘reflection’ result) one can show that

$$\mathbb{P}^0(B_t^* > a) = 1 - \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{\pi^2(2k+1)^2}{8a^2}\right).$$

¹²Let us point out that, without self-similarity, (5.9) remains true, if we use $\mathbb{P}^0(X_t \geq 0)$ instead of $1 - p_0 = \mathbb{P}^0(X_1 \geq 0)$.

It is possible, as Lévy did, to avoid the symmetry assumption in Lemma 5.9 by a centring argument, e.g. at the mean value or the median of X_t . There is, however, a much more powerful (but little known) argument due to Etemadi [40] and, independently, Kwapien & Woyczyński [99, Prop. 1.1.1, Prop. 8.2.1]; we refer to see Billingsley [11, Thm. 22.5] or Schilling [152, Thm. 33.1] for a textbook presentation of the discrete case. The following proof in the continuous-time setting combines Etemadi's original idea with the proof of Scholium 5.7.

Theorem 5.11 (Etemadi). *Let $(X_t)_{t \geq 0}$ be a real-valued Lévy process. Then, for all $a > 0$, $t > 0$*

$$\mathbb{P}^0(M_t > 3a) \leq \mathbb{P}^0(|X_t| \geq a) + \sup_{s \leq t} \mathbb{P}^0(X_s > a), \quad (5.11)$$

as well as

$$\mathbb{P}^0(X_t^* > 3a) \leq 3 \sup_{s \leq t} \mathbb{P}^0(|X_s| \geq a). \quad (5.12)$$

Proof. Set $\tau = \tau_{3a} = \inf\{s > 0 \mid X_s > 3a\}$ and use $\{\tau_{3a} \leq t\} \supseteq \{M_t > 3a\}$ along with the strong Markov property to get

$$\begin{aligned} \mathbb{P}^0(M_t > 3a) &\leq \mathbb{P}^0(\tau \leq t, X_t < a) + \mathbb{P}^0(X_t \geq a) \\ &= \int_{\tau \leq t} \mathbb{P}^{X_\tau(\omega)}(X_{t-\tau(\omega)} < a) \mathbb{P}^0(d\omega) + \mathbb{P}^0(X_t \geq a) \end{aligned}$$

Since $X_\tau \geq 3a$, we can use translation invariance to get

$$\begin{aligned} \mathbb{P}^{X_\tau(\omega)}(X_{t-\tau(\omega)} < a) &= \mathbb{P}^0(X_{t-\tau(\omega)} < a - X_\tau(\omega)) \\ &\leq \mathbb{P}^0(X_{t-\tau(\omega)} < -2a) \\ &\leq \sup_{s \leq t} \mathbb{P}^0(X_{t-s} < -2a). \end{aligned}$$

Since $X_{t-s} \sim X_t - X_s$ and $\{X_t - X_s < -2a\} \subseteq \{X_t < -a\} \cup \{-X_s < -a\}$ we see

$$\mathbb{P}^{X_\tau(\omega)}(X_{t-\tau(\omega)} < a) \leq \mathbb{P}^0(X_t < -a) + \sup_{s \leq t} \mathbb{P}^0(X_s > a).$$

Inserting this estimate into the integral formula which we got from the strong Markov property yields (5.11).

The inequality (5.12) follows along the same lines, we just have to use the stopping time $\sigma_{3a} = \inf\{s > 0 \mid |X_s| > 3a\}$ instead of τ_{3a} . The slightly worse factor 3 comes from the triangle inequality when dealing with $|X_{t-s}| \sim |X_t - X_s|$ in the last step of the argument. \square

Remark 5.12. A close inspection of the proof of Theorem 5.11 shows that it actually yields estimates of the type

$$\mathbb{P}^0(X_t^* > 3a) \leq \frac{\mathbb{P}^0(|X_t| \geq a)}{1 - \sup_{s \leq t} \mathbb{P}^0(|X_{t-s}| > 2a)} \leq \frac{\mathbb{P}^0(|X_t| \geq a)}{\mathbb{P}^0(|X_t| \leq a) - \sup_{s \leq t} \mathbb{P}^0(|X_s| > a)}.$$

Such inequalities – often stated with $a + 2b$ instead of $3a$, causing some obvious changes – are also called *Ottaviani–Skorokhod inequalities* (Billingsley [11, p. 296]) or *generalized Kolmogorov inequalities* (Skorokhod [163, I.§1.1]). See also the notes in Billingsley where the potency of the Etemadi and the Ottaviani–Skorokhod inequalities is compared: as one might expect from our proofs, they do have similar power.

In a series of papers starting in the 1980s Etemadi demonstrated how the maximal estimate (5.12) can be used to replace Kolmogorov’s maximal estimate for centred partial sums, resulting in shorter and often more general proofs of classical results. We will use now Etemadi’s estimate and some of our considerations on censored moments and concentration functions from Sections 2.1, 2.2 to give a short proof of Pruitt’s maximal estimates for Lévy processes, Pruitt [140]. Pruitt used the truncated moments K, G and M from (2.3) and (2.5), the connection with the characteristic exponent ψ was pointed out in Schilling [144], see also Theorem 2.5.

Theorem 5.13. *Let $(X_t)_{t \geq 0}$ be a real-valued Lévy process with characteristic exponent ψ given by (5.3) and denote by $|\psi|^*(\ell) = \sup_{|\xi| \leq \ell} |\psi(\xi)|$ the associated maximum function as in (2.2). Then, there is an absolute constant $c > 0$ such that for all $a > 0, t > 0$*

$$\mathbb{P}^0(X_t^* > a) \leq ct|\psi|^*(a^{-1}) \quad \text{and} \quad \mathbb{P}^0(X_t^* \leq a) \leq \frac{c}{t|\psi|^*(a^{-1})}. \quad (5.13)$$

Proof. Fix $a > 0$ and $t > 0$. For the first estimate we begin with Etemadi’s inequality:

$$\mathbb{P}(X_t^* > a) \leq 3 \sup_{s \leq t} \mathbb{P}(|X_s| > \frac{a}{3}) = 3 \sup_{s \leq t} \mathbb{P}(|X_s| \wedge a > \frac{a}{3}).$$

By the Markov inequality

$$\mathbb{P}(|X_s| \wedge a > \frac{a}{3}) \leq \frac{9}{a^2} \mathbb{E}(|X_s|^2 \wedge a^2) \leq 18 \mathbb{E} \left[\frac{a^{-2} X_s^2}{1 + a^{-2} X_s^2} \right].$$

Using Lemma 2.2 and Tonelli’s theorem we see that

$$\begin{aligned} \mathbb{P}(|X_s| \wedge a > \frac{a}{3}) &\leq 18 \int \mathbb{E} \left[1 - \cos \left(\frac{\xi}{a} X_s \right) \right] g(\xi) d\xi \\ &= 18 \int \operatorname{Re} \left(1 - e^{-s\psi(a^{-1}\xi)} \right) g(\xi) d\xi \\ &\leq 18s \int |\psi(a^{-1}\xi)| g(\xi) d\xi. \end{aligned}$$

The first inequality in (5.13) follows now easily from Lemma 2.3, see also the argument used in the proof of Theorem 2.4.

We will now turn to the second inequality. If we content ourselves with $\psi^*(a^{-1})$ on the right hand side, then the argument is rather straightforward:

$$\begin{aligned} \mathbb{P}(X_t^* \leq a) &\leq \mathbb{P}(|X_{t/2}| \leq a, |X_t| \leq a) \\ &\leq \mathbb{P}(|X_{t/2}| \leq a, |X_t - X_{t/2}| \leq 2a) \\ &\leq \mathbb{P}(|X_{t/2}| \leq 2a)^2; \end{aligned}$$

in the last step we use the independent and stationary increments property of a Lévy process. Now we can apply Esseen's estimate for Lévy's concentration function, Theorem 2.13 and Remark 2.15, to get

$$\mathbb{P}(X_t^* \leq a) \leq \mathbb{P}(|X_{t/2}| \leq 2a)^2 \leq Q_{X_{t/2}}(4a)^2 \leq \frac{C}{tD(\nu_0, 4a)} \leq \frac{C'}{t\psi^*((4a)^{-1})};$$

if we use Lemma 2.3, we can compare $\psi^*((4a)^{-1})$ and $\psi^*(a^{-1})$ to get the upper bound with $\psi^*(a^{-1})$. For many practical purposes this is enough and, quite often one has also the sector condition (Lemma 2.7) which allows one to exchange ψ^* for $|\psi|^*$.

Let us give the additional argument (due to Pruitt [140]) to get $|\psi|^*(a^{-1})$. This relies on estimates in terms of the truncated moments $K(\nu_0; a)$, $G(\nu; a)$ and, crucially, $M(b, \nu; a)$. First we note that we have already seen that

$$\mathbb{P}(X_t^* \leq a) \leq \frac{C}{tD(\nu_0, 4a)} \leq \frac{C'}{t} \left[\frac{1}{K(\nu_0, 4a)} \wedge \frac{1}{G(\nu_0, 4a)} \right].$$

To proceed, we decompose the Lévy process into two independent processes $X_t = Y_t + J_t$, where $J_t = J_t(4a) = \sum_{s \leq t} \Delta X_s \mathbb{1}_{(4a, \infty)}(|\Delta X_s|)$ is the compound Poisson process consisting of all jumps larger than $4a$, and $Y_t = X_t - J_t$. The processes $(Y_t)_{t \geq 0}$ and $(J_t)_{t \geq 0}$ are independent Lévy processes, the characteristic exponents of Y_1 and J_1 are those given in Scholium 2.1 with $\ell = 4a$. On the set $\{X_t^* \leq a\}$, there cannot happen any jumps of size $> 2a$, in particular $J_t^* = 0$, and we see

$$\mathbb{P}(X_t^* \leq a) = \mathbb{P}(X_t^* \leq a, J_t^* = 0) = \mathbb{P}(Y_t^* \leq a, J_t^* = 0) \leq \mathbb{P}(Y_t^* \leq a).$$

Assuming $tM(b, \nu; 4a) \geq 1$ or, equivalently, $|\mathbb{E}Y_t| \geq 4a$, see Scholium 2.1, we have

$$\mathbb{P}(X_t^* \leq a) \leq \mathbb{P}(Y_t^* \leq a) \leq \mathbb{P}(|Y_t| \leq 2a) \leq \mathbb{P}(|Y_t - \mathbb{E}Y_t| \geq \frac{1}{2}|\mathbb{E}Y_t|).$$

The last estimate uses $|Y_t - \mathbb{E}Y_t| \geq |\mathbb{E}Y_t| - |Y_t| \geq |\mathbb{E}Y_t| - 2a \geq \frac{1}{2}|\mathbb{E}Y_t|$ on the set $\{|Y_t| \leq 2a\}$ and $tM(b, \nu; 4a) \geq 1$. The Markov inequality and the moment formulae stated in Scholium 2.1 yield

$$\mathbb{P}(X_t^* \leq a) \leq \frac{4V(Y_t)}{|\mathbb{E}Y_t|^2} = \frac{4K(\nu_0; 4a)}{tM^2(b, \nu; 4a)}.$$

Finally, if $tM(b, \nu; 4a) \leq 1$, we trivially have

$$\mathbb{P}(X_t^* \leq a) \leq 1 \leq \frac{1}{tM(b, \nu; 4a)}.$$

Summing up we get, up to a constant Ct^{-1} , upper bounds of the form

$$\frac{1}{K(\nu_0; 4a)} \wedge \frac{1}{G(\nu; 4a)} \wedge \frac{K(\nu_0; 4a)}{M(b, \nu; 4a)^2} \quad \text{or} \quad \frac{1}{K(\nu_0; 4a)} \wedge \frac{1}{G(\nu; 4a)} \wedge \frac{1}{M(b, \nu; 4a)}.$$

Since $K(\nu_0; 4a) \vee G(\nu; 4a) \vee M(b, \nu; 4a) \asymp |\psi|^*((4a)^{-1}) \asymp |\psi|^*(a^{-1})$, see Theorem 2.5 and Lemma 2.3, the estimate follows in the second case.

The first case needs a further distinction of cases: If $K(\nu_0; 4a) \leq cM(b, \nu; 4a)$, we have $K(\nu_0; 4a)/M(b, \nu; 4a)^2 \leq c/M(b, \nu; 4a)$, and we are back in Case 2. If $K(\nu_0; 4a) > cM(b, \nu; 4a)$, we have $\mathbb{P}(X_t^* \leq a) \leq K(\nu_0; 4a)^{-1} \leq cM(b, \nu; 4a)^{-1}$, which puts us again in Case 2. This finishes the argument. \square

We can apply the technique used in the proof of Theorem 5.13 to get maximal tail estimates for stochastic integrals driven by a Lévy process. The following result recovers a result by Joulin [69] and Giné & Marcus [48] on stochastic integrals driven by stable Lévy processes.

Corollary 5.14. *Let $(X_t)_{t \geq 0}$ be a Lévy process with filtration $(\mathcal{F}_t)_{t \geq 0}$ and $(H_t)_{t \geq 0}$ be an \mathcal{F}_t -adapted càdlàg process such that $\int_0^t \mathbb{E}(H_s^2) ds < \infty$ for all $t \geq 0$. There exists a constant $c \in (0, \infty)$ such that for every $a, \ell > 0$ and $t \geq 0$ the following tail estimate for the stochastic integral $H \bullet X_t = \int_0^t H_{s-} dX_s$ holds:*

$$\mathbb{P}((H \bullet X)_t^* > \ell) \leq c \left(t + \frac{a^2}{\ell^2} t + \frac{a^2}{\ell^2} \int_0^t \mathbb{E}(H_s^2) ds \right) |\psi|^*(a^{-1}). \quad (5.14)$$

Proof. Fix $a > 0$ and write $X_t = (Y_t - \mathbb{E}Y_t) + (J_t + \mathbb{E}Y_t)$ where $J_t = J_t(a) = \sum_{s \leq t} \Delta X_s \mathbb{1}_{(2a, \infty)}(|\Delta X_s|)$ is the Lévy process containing all jumps of X_t whose absolute value is larger than $2a$. The process $\bar{Y}_t := Y_t - \mathbb{E}Y_t$ is both a Lévy process and a square-integrable martingale.

In view of our assumptions, the stochastic integral $H \bullet X_t = \int_0^t H_{s-} dX_s$ is well-defined. We will now estimate the tail of $(H \bullet X)_t^*$:

$$\begin{aligned} \mathbb{P}((H \bullet X)_t^* > \ell) &\leq \mathbb{P}((H \bullet X)_t^* > \ell, X_t^* \leq a) + \mathbb{P}(X_t^* > a) \\ &\leq \mathbb{P}((H \bullet Y)_t^* > \ell, X_t^* \leq a) + ct|\psi|^*(a^{-1}). \end{aligned}$$

The last estimate uses that $X_s = Y_s$ for all $s \leq t$ on the set $\{X_t^* \leq a\}$, since the latter assumption excludes all jumps larger than $2a$; we have also used Theorem 5.13 to estimate the tail of X_t^* . Denote the first expression on the right-hand side by P . Now we use the Markov inequality and the elementary estimate $(A + B)^2 \leq 2A^2 + 2B^2$ to get for any $\ell > 0$

$$\begin{aligned} P &\leq \mathbb{P}((H \bullet Y)_t^* > \ell) \\ &\leq \frac{2}{\ell^2} \mathbb{E} \left([(H \bullet \bar{Y})_t^*]^2 \right) + \frac{2}{\ell^2} (\mathbb{E}Y_1)^2 \mathbb{E} \left(\left[\int_0^t H_s ds \right]^2 \right). \end{aligned}$$

Applying Doob's maximal inequality (Theorem 4.4 with $p = 2$), Itô's isometry – recall that $[\bar{Y}]_t = \mathbb{E}[\bar{Y}_t^2] = t\mathbb{E}[\bar{Y}_1^2]$ – and Jensen's inequality, yields

$$\begin{aligned}
P &\leq \frac{8}{\ell^2} \mathbb{E} \left(\int_0^t H_s^2 d[\bar{Y}]_s \right) + \frac{2t}{\ell^2} (\mathbb{E}Y_1)^2 \mathbb{E} \left(\int_0^t H_s^2 ds \right) \\
&= \frac{8}{\ell^2} \mathbb{E} \left(\int_0^t H_s^2 ds \right) \mathbb{E} (\bar{Y}_1^2) + \frac{2t}{\ell^2} (\mathbb{E}Y_1)^2 \mathbb{E} \left(\int_0^t H_s^2 ds \right) \\
&= \frac{8}{\ell^2} \left(\mathbb{E} (\bar{Y}_1^2) + t (\mathbb{E}Y_1)^2 \right) \int_0^t \mathbb{E} (H_s^2) ds.
\end{aligned}$$

We can now apply Theorem 2.5 to bound the expressions $\ell^{-2} \mathbb{E} (\bar{Y}_1^2)$ and $\ell^{-2} (\mathbb{E}Y_1)^2$ by $\ell^{-2} a^2 |\psi|^* ((2a)^{-1})$. Finally, an application of Lemma 2.3 completes the proof, yielding (5.14). \square

Remark 5.15. There are quite a few variations possible in Corollary 5.14.

a) If $(X_t)_{t \geq 0}$ is symmetric, $\mathbb{E}Y_1 = 0$, and the term $a^2 \ell^{-2} t$ inside the parentheses on the right-hand side of (5.14) can be omitted.

b) If $(X_t)_{t \geq 0}$ is a symmetric α -stable Lévy process, then $|\psi|^*(a^{-1}) = a^{-\alpha}$; in particular, (5.14) becomes estimate of Joulin [69, Eq. (2.11)] and Giné & Marcus [48], if we use $a := \ell / \int_0^t \mathbb{E}(H_s) ds$.

c) We can obtain L^p -estimates, if we use the Markov inequality with an L^p -norm (rather than an L^2 -norm) when estimating the probability P , and combine this with the Burkholder–Davis–Gundy inequalities (Theorem 4.21). Further variants can be found in Joulin [69].

A typical application of the maximal estimates of Theorem 5.13 are Pruitt's estimates [140] for the speed of a Lévy process. Recall that we have for the stopping time $\sigma_r := \inf \{t > 0 \mid |X_t| > r\}$, $r > 0$, and $t > 0$ the inclusions

$$\{X_t^* < r\} \subseteq \{\sigma_r > t\} \subseteq \{X_t^* \leq r\} \subseteq \{\sigma_r \geq t\}.$$

Corollary 5.16. *Let $(X_t)_{t \geq 0}$ be a real-valued Lévy process and denote by σ_r the first exit time from the open ball (interval) $B_r(0)$. Then, there are absolute constants $c, c' > 0$ such that*

$$\frac{c}{|\psi|^*(r^{-1})} \leq \mathbb{E}^0 [\sigma_r] \leq \frac{c'}{|\psi|^*(r^{-1})}, \quad r > 0. \quad (5.15)$$

Proof. We begin with a slight improvement of the second maximal estimate in (5.13). The idea for this is already present in the proof of Theorem 5.13. We use the fact that $(X_s)_{s \leq t/2}$ and $(X_{u+t/2} - X_{t/2})_{u \geq 0}$ are iid Lévy processes to get

$$\begin{aligned}
\mathbb{P}^0 (X_t^* \leq a) &= \mathbb{P}^0 \left(X_{t/2}^* \leq a, \sup_{t/2 \leq s \leq t} |X_s| \leq a \right) \\
&\leq \mathbb{P}^0 \left(X_{t/2}^* \leq a, \sup_{u \leq t/2} |X_{u+t/2} - X_{t/2}| \leq 2a \right) \\
&\leq \mathbb{P}^0 \left(X_{t/2}^* \leq 2a \right)^2 \\
&\leq \frac{c}{t^2 (|\psi|^*((2a)^{-1}))^2}.
\end{aligned}$$

The argument $(2a)^{-1}$ can be replaced by a^{-1} using Lemma 2.3. By the layer cake formula we have

$$\mathbb{E}^0 [\sigma_r] = \int_0^\infty \mathbb{P}^0 (\sigma_r > t) dt \leq \int_0^\infty \mathbb{P}^0 (X_t^* \leq r) dt.$$

For the upper estimate we use the (just improved) second inequality of (5.13)

$$\begin{aligned} \mathbb{E}^0 [\sigma_r] &\leq x + \int_x^\infty \mathbb{P}^0 (X_t^* \leq r) dt \\ &\leq x + c \int_x^\infty \frac{dt}{t^2 (|\psi|^*(r^{-1}))^2} \\ &= x + \frac{1}{x} \frac{c}{(|\psi|^*(r^{-1}))^2}. \end{aligned}$$

Minimizing in x gives the upper bound in (5.15). The lower bound follows in a similar fashion, now with the (original version of the) first maximal inequality in (5.13):

$$\begin{aligned} \mathbb{E}^0 [\sigma_r] &= \int_0^\infty \mathbb{P}^0 (\sigma_r \geq t) dt \geq \int_0^x \mathbb{P}^0 (X_t^* \leq r) dt \\ &= x - \int_0^x \mathbb{P}^0 (X_t^* > r) dt \\ &\geq x - c \int_0^x t |\psi|^*(r^{-1}) dt \\ &= x - \frac{c}{2} x^2 |\psi|^*(r^{-1}). \end{aligned}$$

Again we can optimize in x to get the lower bound. □

Here is a typical application of the estimates in Corollary 5.16, which is needed to derive Harnack inequalities for Lévy processes, see Song & Vondraček [164].

Corollary 5.17. *Let $(X_t^z)_{t \geq 0}$ be a Lévy process with starting point $X_0 = z$, and denote by \mathbb{P}^z the corresponding probability measure and by $\sigma_{B_r(x)}^z$ the first exit time from the open ball $B_r(x)$ with radius $r > 0$ and centre x . There is a constant $C \in (0, \infty)$ such that*

$$\sup_{z \in B_r(x)} \mathbb{E}^z \left[\sigma_{B_r(x)}^z \right] \leq C \inf_{y \in B_{r/2}(x)} \mathbb{E}^y \left[\sigma_{B_r(x)}^y \right], \quad x \in \mathbb{R}. \quad (5.16)$$

Proof. Write $X_t := X_t^0$ for the Lévy process with starting point $X_0 = 0$. For a Lévy process, $(X_t^z)_{t \geq 0}$ is obtained by a simple shift: $X_t^z = X_t + z$. Since $B_r(x) \subseteq B_{2r}(z)$ for all $z \in B_r(x)$ and $B_{r/2}(y) \subseteq B_r(x)$ for all $y \in B_{r/2}(x)$, we see that for any $x \in \mathbb{R}$

$$\forall y \in B_{r/2}(x), \forall z \in B_r(x) : \quad \sigma_{B_{r/2}(y)}^y \leq \sigma_{B_r(x)}^y \quad \text{and} \quad \sigma_{B_r(x)}^z \leq \sigma_{B_{2r}(z)}^z.$$

Because of $\mathbb{E}^y [\sigma_{B_s(y)}^y] = \mathbb{E}^0 [\sigma_{B_s(0)}^0]$, we can apply Corollary 5.16 to get

$$\sup_{z \in B_r(x)} \mathbb{E}^z \left[\sigma_{B_r(x)}^z \right] \leq \sup_{z \in B_r(x)} \mathbb{E}^z \left[\sigma_{B_{2r}(z)}^z \right] \leq \frac{c}{|\psi|^* \left(\frac{1}{2r} \right)}$$

as well as

$$\frac{c'}{|\psi|^* \left(\frac{2}{r} \right)} \leq \inf_{y \in B_{r/2}(x)} \mathbb{E}^y \left[\sigma_{B_{r/2}(y)}^y \right] \leq \inf_{y \in B_{r/2}(x)} \mathbb{E}^y \left[\sigma_{B_r(x)}^y \right].$$

Using Lemma 2.3, we can compare $|\psi|^* \left(\frac{2}{r} \right) \asymp |\psi|^* \left(\frac{1}{2r} \right)$, leading to (5.16). \square

5.3. Generalized moments of Lévy processes

Traditionally expressions of the form $\mathbb{E}[g(X_t)]$ are called *generalized moments*. The minimum requirement on $g: \mathbb{R}^d \rightarrow [0, \infty)$ is measurability, but to get satisfactory existence (finiteness) results, we need that g is *locally bounded* and *submultiplicative*, i.e.

$$\forall r > 0 : \sup_{|x| \leq r} g(x) < \infty \quad \text{and} \quad \forall x, y \in \mathbb{R}^d : g(x + y) \leq g(x)g(y). \quad (5.17)$$

Scholium 5.18 (on submultiplicative functions). Sometimes the submultiplicativity property is stated as $g(x + y) \leq cg(x)g(y)$ for some constant $c \geq 1$; changing $g \rightsquigarrow c \cdot g$ reduces things to the present definition.

Regularity By a mollifier argument, see Berger *et al.* [10], we can always modify g in such a way, that it is smooth and its derivatives $|\partial^\alpha g(x)|$ are dominated by $c_\alpha g(x)$ for a suitable constant $c_\alpha \in (0, \infty)$. Fix g . Let $\phi = \phi_g \in C^\infty(\mathbb{R}^d)$ such that $\phi \geq 0$, $\text{supp } \phi = \overline{B}_1(0)$ and $\int \phi(z) dz = \sup_{|x| \leq 1} g(x) = c_g$. From

$$g * \phi(x + y) = \int g(x + y - z)\phi(z) dz \leq \int g(y - z)\phi(z) dz \cdot g(x)$$

in combination with

$$c_g g(x) = \int g(x - z + z)\phi(z) dz \leq \sup_{|z| \leq 1} g(z) \int g(x - z)\phi(z) dz = c_g g * \phi(x)$$

we get that $g * \phi$ is submultiplicative and, taking $y = 0$, that $g * \phi \asymp g$. It is now easy to see that $\partial^\alpha(g * \phi) = g * (\partial^\alpha \phi)$ can be bounded by $\|g \cdot \partial^\alpha \phi\|_{L^1} \cdot g(x)$.

Taking a convolution does not only regularize the function g , but it gives additional structure, which allows one to get such kind of bounds for the derivatives. For example, in dimension one $g(x) = e^x(2 + \sin(x^2))$ is smooth and does *not* satisfy the above bounds, see Martin [122].

Boundedness Since $\gamma(r) := \sup_{|x| \leq r} g(x)$ satisfies

$$\gamma(r + s) = \sup_{|z| \leq r+s} g(z) = \sup_{|x| \leq r, |y| \leq s} g(x + y) \leq \gamma(r)\gamma(s),$$

it is submultiplicative and increasing. Without loss of generality, γ is smooth, so integrating up the estimate $\gamma'(r) \leq c\gamma(r)$ yields $\log \gamma(r) \leq c_0 r + c_1$. Thus, $g(x)$ grows at most exponentially. For an alternative proof see Sato [143, Lem. 25.5].

Examples If g is submultiplicative, so are – up to a multiplicative constant – $x \mapsto g(x) + 1$ and $x \mapsto \max\{g(x), 1\}$ and $g^\alpha(x)$ for any $\alpha > 0$. Typical examples of submultiplicative functions are $x \mapsto \max\{|x|, 1\}$, $x \mapsto \exp [c|x|^\beta]$, $\beta \in (0, 1]$, $\log(|x| + e)$, $\log(\log(x + e^e))$ and all combinations with coordinate projections, see Sato [143, Prop. 25.4].

Since we are only interested in estimating $\mathbb{E}[g(X_t)]$, we may always switch between g , $\phi * g$ and $\phi * g + 1$, i.e. assume that g is as good as we need it to be.

A classical result in the theory of Lévy processes, due to Kruglov [84] (see also Sato [143, Thm. 25.3]), shows that the generalized moment of a Lévy process with triplet (b, Q, ν) satisfies

$$\mathbb{E}[g(X_t)] < \infty \iff \int_{|y| \geq 1} g(y) \nu(dy) < \infty.$$

The right hand side does not depend on t , and so we may read the left hand side to be true for *some* or for *all* $t > 0$. In the following theorem we will see that this equivalence remains true for $\sup_{s \leq t} g(X_s)$, i.e. we have a maximal inequality. The following statement is taken from Berger *et al.* [10, Thm. 3], but most of the equivalences were known earlier, see Remark 5.20 for a discussion.

Theorem 5.19. *Let $(X_t)_{t \geq 0}$ be a d -dimensional Lévy process with characteristic exponent ψ and triplet (b, Q, ν) . Moreover, $g: \mathbb{R}^d \rightarrow [0, \infty)$ is a locally bounded submultiplicative function, and \mathcal{T} is the family of all stopping times for the process $(X_t)_{t \geq 0}$. Then, the maximal moment is finite*

(M1) $\mathbb{E}[\sup_{s \leq t} g(X_s)] < \infty$ for some (hence, all) $t > 0$,

if, and only if, one of the following (a fortiori equivalent) conditions (M2)–(S9) is satisfied.

Siebert’s moment conditions:

(M2) $\mathbb{E}[g(X_t)] < \infty$ for some (hence, all) $t > 0$;

(M3) $\{g(X_{\sigma \wedge t})\}_{\sigma \in \mathcal{T}}$ is uniformly integrable for some (hence, all) $t > 0$, i.e.

$$\lim_{R \rightarrow \infty} \sup_{\sigma \in \mathcal{T}} \int_{g(X_\sigma) > R} g(X_\sigma) d\mathbb{P} = 0;$$

(M4) $\sup_{\sigma \in \mathcal{T}} \mathbb{E}[g(X_{\sigma \wedge t})] < \infty$ for some (hence, all) $t > 0$;

(M5) [this requires, in addition, that $g(x) = g(|x|)$ and $r \mapsto g(r)$ is increasing]

$$\mathbb{E}[g(\sup_{s \leq t} |X_t|)] < \infty \text{ for some (hence, all) } t > 0.$$

Kruglov’s condition in terms of the triplet:

(T6) $\int_{|y| \geq 1} g(y) \nu(dy) < \infty$.

Hulanicki’s conditions in terms of the semigroup and the generator:

(S7) The adjoint semigroup $P_t^* f(x) := \mathbb{E}[f(x - X_t)]$ is a strongly continuous semigroup on the weighted Lebesgue space $L^1(\mathbb{R}^d; g(x) dx)$;

(S8) The adjoint infinitesimal generator $(A^*\phi)(x) := (A\phi)(-x)$ satisfies $A^*\phi \in L^1(\mathbb{R}^d, g(x) dx)$ for all $\phi \in C_c^\infty(\mathbb{R}^d)$;

(S9) There is a non-negative $0 \neq \phi \in C_c^\infty(\mathbb{R}^d)$ such that $A^*\phi \in L^1(\mathbb{R}^d; g(x) dx)$.

If one (hence, all) of these conditions is satisfied, then there is some constant $c > 0$ such that

$$\mathbb{E}^x \left[\sup_{s \leq t} g(X_s) \right] \leq g(x)e^{ct}, \quad t > 0.$$

Proof. Without loss of generality we can assume that g is of the form $g * \phi$ and $g \geq 1$, see Scholium 5.18.

1° We begin by showing that $S(t) := \mathbb{E} [\sup_{s \leq t} g(X_s)] < \infty$ for *some* t implies $S(t) < \infty$ for *all* t . Since $t \mapsto S(t)$ is increasing, the claim follows if we can show that $S(2t) < \infty$ if $S(t) < \infty$. We have used a very similar argument in the proof of Corollary 5.16. Since $(X_s)_{s \leq t}$ and $(X_{u+t} - X_t)_{u \geq 0}$ are iid Lévy processes, we have

$$\begin{aligned} S(2t) &\leq S(t) + \mathbb{E} \left[\sup_{u \leq t} g(X_{u+t} - X_t + X_t) \right] \\ &\leq S(t) + \mathbb{E} \left[\sup_{u \leq t} g(X_{u+t} - X_t) g(X_t) \right] \\ &\leq S(t) + \mathbb{E}[g(X_t)] \cdot \mathbb{E} \left[\sup_{u \leq t} g(X_{u+t} - X_t) \right] \\ &\leq S(t) + S(t)^2. \end{aligned}$$

2° If $\mathbb{E}[g(X_t)]$ is finite for *some* $t > 0$, then it is finite for *all* $s < t$, too. Indeed, using submultiplicativity in the form $g(y) \geq g(x)/g(x-y)$ and the independent increments property of a Lévy process, we get

$$\infty > \mathbb{E}[g(X_t)] \geq \mathbb{E} \left[\frac{g(X_s)}{g(X_s - X_t)} \right] = \mathbb{E}[g(X_s)] \cdot \underbrace{\mathbb{E} \left[\frac{1}{g(X_s - X_t)} \right]}_{\in (0,1]}.$$

Since $1 \leq g(x) < \infty$, we have $0 < 1/g(x) \leq 1$, and the assertion follows.

3° Assume that (M2) holds for *some* t , hence for all sufficiently small t . We show that (M1) holds for *some* $T > 0$. For $a, b > 0$ we define a stopping time

$$\sigma := \sigma_{a,b} := \inf \{ s \mid g(X_s) > g(0)e^{a+b} \}$$

and observe that the submultiplicativity of g gives

$$\begin{aligned} \mathbb{P}(g(X_t) > e^a) &\geq \mathbb{P}(g(X_t) > e^a, g(X_\sigma) \geq g(0)e^{a+b}, \sigma \leq t) \\ &\geq \mathbb{P}(g(X_\sigma - X_t) < g(0)e^b, g(X_\sigma) \geq g(0)e^{a+b}, \sigma \leq t). \end{aligned}$$

The strong Markov property yields (for all $t \leq T$, T will be determined in the following step)

$$\begin{aligned} \mathbb{P}(g(X_t) > e^a) &\geq \int_{\sigma \leq t} \mathbb{P}(g(-X_{t-\sigma(\omega)}) < g(0)e^b) \mathbb{P}(d\omega) \\ &\geq \inf_{r \leq t} \mathbb{P}(g(-X_r) < g(0)e^b) \cdot \mathbb{P}(\sigma \leq t) \\ &\geq \frac{1}{2} \mathbb{P}(\sup_{r \leq t} g(X_r) > cg(0)e^{a+b}). \end{aligned}$$

In the last estimate we use $\{\sup_{s \leq t} g(X_s) > g(0)e^{a+b}\} \subseteq \{\sigma \leq t\}$. The factor $\frac{1}{2}$ comes from the fact that g is locally bounded and $\lim_{t \rightarrow 0} \mathbb{P}(|X_t| > \epsilon) = 0$ (continuity in probability), which shows that there is some $T > 0$ such that

$$\mathbb{P}(g(-X_t) < g(0)e^b) \geq \frac{1}{2} \quad \text{for all } t \leq T.$$

Setting $e^\gamma := g(0)e^b$ we have shown

$$\mathbb{P}(\sup_{s \leq T} g(X_s) > e^{a+\gamma}) \leq 2\mathbb{P}(g(X_T) > e^a).$$

Now a simple variant of the layer-cake argument yields

$$\mathbb{E} \left[\sup_{s \leq T} g(X_s) \right] \leq e^\gamma + 2e^\gamma \mathbb{E}[g(X_T)].$$

Using Step 2^o, we see that (M1) holds for this particular $T > 0$, then by Step 1^o for all $T > 0$. The latter trivially implies (M2) for all $t > 0$. Thus, we have established that (M1) \Leftrightarrow (M2).

4^o The uniform integrability of $\{g(X_{\sigma \wedge t})\}_{\sigma \in \mathcal{T}}$ (for some or all $t > 0$) follows immediately from the uniform bound $g(X_{\sigma \wedge t}) \leq \sup_{s \leq t} g(X_s) \in L^1(\mathbb{P})$ (for some or all $t > 0$).

Since uniformly integrable families are L^1 -bounded, it is clear that (M3) implies (M4) – both for some $t > 0$ or for all $t > 0$. Finally, (M4) trivially gives (M2).

Summing up, we have shown (M1) \Rightarrow (M3) \Rightarrow (M4) \Rightarrow (M2) \Rightarrow (M1).

5^o Assume, in addition, that $g(x) = g(|x|)$ and that $g(r)$ is increasing. By monotonicity, $\sup_{s \leq t} g(|X_s|) \leq g(\sup_{s \leq t} |X_s|)$. On the other hand, by right continuity, there is a random time $\tau \in [0, t]$ such that $|\sup_{s \leq t} |X_s(\omega)| - |X_\tau(\omega)|| \leq 1$. Consequently,

$$g\left(\sup_{s \leq t} |X_s|\right) \leq g\left(\sup_{s \leq t} |X_s| - |X_\tau|\right) g(|X_\tau|) \leq \sup_{|y| \leq 1} g(|y|) \sup_{s \leq t} g(|X_s|).$$

The equivalence (M1) \Leftrightarrow (M5) now follows from the estimates $\sup_{s \leq t} g(|X_s|) \asymp g(\sup_{s \leq t} |X_s|)$.

6° Now we show (T6) \Rightarrow (M2). For this, we use that g is smooth (and has a convolution structure) so that we can insert g into the formula (5.5) of the generator A of $(X_t)_{t \geq 0}$. The differential operator part can be estimated using $|\partial^\alpha g(x)| \leq c_\alpha g(x)$, see Scholium 5.18. For the second integral expression we first apply Taylor's formula, $|g(x+y) - g(x) - y \cdot \nabla g(x)| \leq c|y|^2 |\nabla^2 g(x + \theta y)|$ with $\theta \in (0, 1)$ and for all $|y| < 1$, and use again $|\partial^\alpha g(x + \theta y)| \leq c_\alpha g(x + \theta y) \leq c_\alpha g(x) \sup_{|z| \leq 1} g(z)$. For the first integral term of (5.5) we have $g(x+y) - g(x) \leq g(x)(g(y) - 1)$. Altogether,

$$Ag(x) \leq c(b, Q, \nu|_{B_1(0)}) \cdot g(x) + \int_{|y| \geq 1} (g(y) - 1) \nu(dy) \cdot g(x).$$

The last integral is finite due to (T6). By Itô's formula, $M_t = g(X_t) - g(x) - \int_0^t Ag(X_s) ds$ is a local martingale. If $(\sigma_n)_{n \in \mathbb{N}}$ is a sequence of stopping times such that $M_{t \wedge \sigma_n}$ is a martingale, we get

$$\mathbb{E}^x [g(X_{t \wedge \sigma_n})] - g(x) = \int_0^t \mathbb{E}^x [Ag(X_{s \wedge \sigma_n})] ds \leq c \int_0^t \mathbb{E}^x [g(X_{s \wedge \sigma_n})] ds.$$

Now we can apply Gronwall's lemma to conclude

$$\mathbb{E}^x [g(X_{t \wedge \sigma_n})] \leq g(x)e^{ct}.$$

Letting $n \rightarrow \infty$, Fatou's lemma proves (M2) as well as the exponential bound for the generalized moment.

7° In order to see (M1) \Rightarrow (T6) we write $X_t = Y_t + J_t$ where J_t is the compound Poisson process comprising all jumps whose absolute height exceeds 1; we have used this (in dimension 1) in the proof of Theorem 5.13, see also Scholium 2.1. In particular, $J_t = H_1 + \dots + H_{N_t}$, where the jump heights H_k are iid random variables with law $\nu(dy \cap B_1^c(0)) / \nu(B_1^c(0))$ and $(N_t)_{t \geq 0}$ is an independent Poisson process with intensity $\nu(B_1^c(0))$.

Now we can argue as in Step 2°, using the subadditivity of g and the independence of Y_t and J_t :

$$\infty > \mathbb{E} [g(X_t)] = \mathbb{E} [g(Y_t + J_t)] \geq \mathbb{E} \left[\frac{g(J_t)}{g(-Y_t)} \right] = \mathbb{E} [g(J_t)] \cdot \underbrace{\mathbb{E} \left[\frac{1}{g(-Y_t)} \right]}_{\in (0,1] \text{ as } g(y) \geq 1}.$$

Moreover, because of the structure of J_t , we see

$$\begin{aligned} \mathbb{E} [g(J_t)] &= \mathbb{E} \left[\sum_{n=0}^{\infty} g(H_1 + \dots + H_n) \mathbb{P}(N_t = n) \right] \\ &\geq \mathbb{E} [g(H_1)] \mathbb{P}(N_t = 1) \\ &= e^{-t\nu(B_1^c(0))} \int_{|y| \geq 1} g(y) \frac{\nu(dy)}{\nu(B_1^c(0))}, \end{aligned}$$

and (T6) follows.

This completes the proof of the equivalence of (M1)–(T6). For Hulanicki’s conditions, we need to show $(M3) \Rightarrow (S7) \Rightarrow (M2)$ and $(T6) \Rightarrow (S8) \Rightarrow (S9) \Rightarrow (M2)$. Since this requires semigroup methods, we omit the proofs, and refer to Berger *et al.* [10]. \square

Remark 5.20. Let us be a bit more precise in naming the above conditions. Siebert [161] showed the equivalence $(M1) \Leftrightarrow (M2)$. Since it is not a long way from there to the conditions (M3), (M4) in this group (as observed by Berger *et al.* [10]), we feel that the name-tag is justified. The last condition (M5) in this group is inspired by Kwapien & Woyczyński [99, Prop. 8.4.1] and Sato [143, Thm. 25.18], but our almost trivial new proof reveals that it properly belongs here. We have already mentioned Kruglov [84] for $(M2) \Leftrightarrow (T6)$, while the equivalence of (M2), (S8) and (S9) is due to Hulanicki [61]; morally, (S7) also belongs here, it is from Berger *et al.* [10].

The estimates of Theorem 5.19 (partially) extend to larger classes of processes, notably to additive processes (see the following remark) and Lévy-type processes, see Kühn [88].

Remark 5.21. A ‘Lévy process without stationary increments’, i.e. a stochastic process with independent increments (L2) or (L2’) and stochastic continuity (L3) is called an *additive process*¹³. Essentially, the Lévy–Khintchine formula (5.3) and Theorem 5.6 remain valid with t -dependent Lévy triplets. This means, in particular, that the equivalence of (M1)–(T6) in Theorem 5.19 remain valid – even with almost the same proof, see Berger *et al.* [10, Thm. 11]; the ‘if we have an estimate for some t , we have it for all t ’ assertion, however, breaks down for additive processes.

6. Lévy-type processes

In this section, we study maximal inequalities for so-called *Lévy-type processes*. Intuitively,

a Lévy-type process behaves locally like a Lévy process, which means that the Lévy triplet may change, depending on the current position of the process. Once started at a point $x_0 \in \mathbb{R}^d$, the process behaves for an ‘infinitesimally small’ time Δt like a Lévy process with Lévy triplet $(b(x_0), Q(x_0), \nu(x_0, dy))$ depending on the starting point x_0 , then for an ‘infinitesimally small’ time like a Lévy process with Lévy triplet $(b(X_{\Delta t}), Q(X_{\Delta t}), \nu(X_{\Delta t}, dy))$, and so on.

The corresponding family $(b(x), Q(x), \nu(x, dy))$, $x \in \mathbb{R}^d$, of Lévy triplets characterizes many distributional properties and sample path properties of the Lévy-type process. Throughout, we assume that $x \mapsto (b(x), Q(x), \nu(x, B))$ is measurable for every Borel set $B \subset \mathbb{R}^d \setminus \{0\}$.

¹³In fact, Lévy himself considered mostly this class of processes and Itô calls them – correctly – Lévy processes, contrary to today’s usage.

6.1. Getting started: Basics on Lévy-type processes

Let $(X_t)_{t \geq 0}$ be a strong Markov process with respect to a family of probability measures $(\mathbb{P}^x)_{x \in \mathbb{R}^d}$, and assume that $(X_t)_{t \geq 0}$ has càdlàg sample paths. Recall that the underlying filtered probability space $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F}_t)$ satisfies the usual conditions, i.e. $(\Omega, \mathcal{A}, \mathbb{P})$ is complete, \mathcal{F}_0 contains all \mathbb{P} -null sets and $(\mathcal{F}_t)_{t \geq 0}$ is right-continuous. We call $(X_t)_{t \geq 0}$ a *Lévy-type process* if for every $x \in \mathbb{R}^d$ and $f \in C_c^\infty(\mathbb{R}^d)$, the process

$$M_t^f := f(X_t) - f(X_0) - \int_0^t Af(X_s) ds$$

is a martingale with respect to \mathbb{P}^x for some *Lévy-type operator* A , i.e. an operator of the form

$$\begin{aligned} Af(x) &= b(x) \cdot \nabla f(x) + \frac{1}{2} \operatorname{tr} (Q(x) \cdot \nabla^2 f(x)) \\ &\quad + \int_{\mathbb{R}^d \setminus \{0\}} (f(x+y) - f(x) - \nabla f(x) \cdot y \mathbf{1}_{(0,1)}(|y|)) \nu(x, dy); \end{aligned} \quad (6.1)$$

here, $(b(x), Q(x), \nu(x, dy))$ is a family of Lévy triplets, the so-called (*infinitesimal*) *characteristics*. As before, we assume that $x \mapsto (b(x), Q(x), \nu(x, B))$ is measurable for every Borel set $B \subseteq \mathbb{R}^d \setminus \{0\}$. In other words, the Lévy-type process $(X_t)_{t \geq 0}$ is a solution to the $(A, C_c^\infty(\mathbb{R}^d))$ -martingale problem for the Lévy-type operator A ; see Ethier & Kurtz [41] for further information on the martingale problem and its connection to Markov processes. Note that, by Taylor's formula, the pointwise expression (6.1) makes sense for any $f \in C_b^2(\mathbb{R}^d)$. Using standard rules for the Fourier transform, it is not difficult to see that $A|_{C_c^\infty(\mathbb{R}^d)}$ can be written as a pseudo-differential operator

$$Af(x) = - \int_{\mathbb{R}^d} q(x, \xi) e^{ix \cdot \xi} \widehat{f}(\xi) d\xi, \quad f \in C_c^\infty(\mathbb{R}^d), \quad x \in \mathbb{R}^d,$$

where $\widehat{f}(\xi) = (2\pi)^{-d} \int_{\mathbb{R}^d} f(x) e^{-ix \cdot \xi} dx$ is the Fourier transform of f and

$$q(x, \xi) = -ib(x) \cdot \xi + \frac{1}{2} \xi \cdot Q(x) \xi + \int_{y \neq 0} (1 - e^{iy \cdot \xi} + iy \cdot \xi \mathbf{1}_{(0,1)}(|y|)) \nu(x, dy) \quad (6.2)$$

is the *symbol*. For fixed $x \in \mathbb{R}^d$, the mapping $\xi \mapsto q(x, \xi)$ is the characteristic exponent of a Lévy process. In particular, $\xi \mapsto q(x, \xi)$ is continuous and negative definite (in the sense of Schoenberg), cf. Sections 2.1 and 5.1. This implies that $\xi \mapsto \sqrt{|q(x, \xi)|}$ is subadditive, i.e.

$$\sqrt{|q(x, \xi + \eta)|} \leq \sqrt{|q(x, \xi)|} + \sqrt{|q(x, \eta)|}, \quad x, \xi, \eta \in \mathbb{R}^d, \quad (6.3)$$

see e.g. Schilling [148, Thm. 6.2] or the proof of Lemma 2.3. In the following, we will always assume that the symbol $(x, \xi) \mapsto q(x, \xi)$ is locally bounded. In terms of the characteristics this means that

$$\sup_{x \in K} \left(|b(x)| + |Q(x)| + \int_{y \neq 0} \min \{1, |y|^2\} \nu(x, dy) \right) < \infty \quad (6.4)$$

for any compact set $K \subseteq \mathbb{R}^d$. In particular, the local boundedness of q entails that $\|Af\|_\infty < \infty$ for any $f \in C_c^\infty(\mathbb{R}^d)$. Let us give some important examples of Lévy-type processes.

Example 6.1. 1. *Lévy processes:* If $(X_t)_{t \geq 0}$ is a Lévy process with characteristic exponent ψ and Lévy triplet (b, Q, ν) , then $(X_t)_{t \geq 0}$ is a Lévy-type process with symbol $q(x, \xi) = \psi(\xi)$ and characteristics (b, Q, ν) . In fact, there is a one-to-one correspondence between *Lévy triplets, characteristic exponents (symbols) and Lévy processes*.

2. *Feller processes:* Let $(X_t)_{t \geq 0}$ be a Feller process. If the domain of its (strong) infinitesimal generator contains the test functions $C_c^\infty(\mathbb{R}^d)$, then a theorem by Courrège and van Waldenfels shows that $(X_t)_{t \geq 0}$ is a Lévy-type process, cf. Böttcher *et al.* [14]. For a thorough study of Feller processes and their connection to Lévy-type operators we refer to the monographs Böttcher *et al.* [14], Hoh [58] and Jacob [66].

Although a Feller process always gives an x -dependent symbol $q(x, \xi)$ and an x -dependent Lévy triplet $(b(x), Q(x), \nu(x, dy))$, the converse does not hold: there are triplets and symbols which are not connected with a *Feller* process, cf. Böttcher *et al.* [14, Ex. 2.26] and van Casteren [175, p. 156]. All of these examples still lead to (deterministic) processes; we are not aware of examples of symbols of the form (6.2) which are not connected with a (stochastic or deterministic) process at all.

3. *Solutions of Lévy-driven SDEs:* Let $(L_t)_{t \geq 0}$ be a Lévy process with characteristic exponent ψ and σ a continuous bounded function. If the stochastic differential equation (SDE)

$$dX_t = \sigma(X_{t-}) dL_t, \quad X_0 = x,$$

has a unique weak solution, then $(X_t)_{t \geq 0}$ is a Lévy-type process with symbol $q(x, \xi) = \psi(\sigma(x)^\top \xi)$, where ψ is the characteristic exponent of $(L_t)_{t \geq 0}$, cf. Schilling & Schnurr [153]. The assumption on the boundedness of σ and the uniqueness of the solution can be relaxed, cf. Kühn [89, 92].

4. *(Compound) pseudo Poisson processes:* William Feller [43, Ch. X.1] calls Lévy-type processes with a triplet of the form $(0, 0, \nu(x, dy))$ such that $\lambda_0 := \sup_{x \in \mathbb{R}^d} \nu(x, \mathbb{R}^d \setminus \{0\}) < \infty$ (compound) pseudo Poisson processes. The condition on the Lévy kernel means that we have only finite jump activity such that the sample paths of the process are piecewise constant with exponential holding times between any two jumps. It is not difficult to construct such processes in a pathwise manner, see Feller *loc. cit.* or Ethier & Kurtz [41, Ch. 4.§2]. Another way to understand (and, indeed to construct) such processes is to start with an arbitrary time-homogeneous Markov chain $(Y_n)_{n \in \mathbb{N}}$ with transition kernel $p(x, dy) := \lambda_0^{-1} \nu(x, dy)$, and to time-change it using an independent Poisson process $(N_t)_{t \geq 0}$ with intensity λ_0 , i.e. $X_t := Y_{N_t}$.

5. *Stable-like processes*: A Feller process whose generator is of the form $(-\Delta)^{\alpha(x)}$ for some continuous function $\alpha: \mathbb{R}^d \rightarrow [0, 2]$ is a stable-like process. Intuitively this is a stable process which changes its stability behaviour ‘on the go’. This class of processes was proposed by Bass [4] who also proved the existence in dimension $d = 1$. Rigorous results in higher dimensions are due to Hoh [59], Jacob & Leopold [67] and using a parametric construction e.g. in Knopova *et al.* [76], Kolokoltsov [83] or Kühn [87], see also the references therein.
6. *Processes connected with Dirichlet forms*: In Schilling & Uemura [154] criteria are given in terms of Dirichlet forms such that a negative definite symbol $q(x, \xi)$ gives rise to a Dirichlet form and, thus, a stochastic process. This process need not be a Feller process. In Schilling & Wang [156] this discussion is extended to not necessarily symmetric semi-Dirichlet forms.

The symbol plays a key role for the probability estimates which we are going to derive. It is therefore natural to ask how the symbol can be calculated given the stochastic process. If $(X_t)_{t \geq 0}$ is a Lévy process, then the Lévy–Khintchine formula yields

$$-\psi(\xi) = \lim_{t \rightarrow 0} \frac{e^{-t\psi(\xi)} - 1}{t} = \lim_{t \rightarrow 0} \frac{\mathbb{E}^x [e^{i\xi(X_t - x)}] - 1}{t}.$$

Although the Lévy–Khintchine formula breaks down for Lévy-type processes, i.e. $\mathbb{E}^x e^{i\xi(X_t - x)} \neq e^{-tq(x, \xi)}$ in general, the identity

$$-q(x, \xi) = \lim_{t \rightarrow 0} \frac{\mathbb{E}^x e^{i\xi(X_t - x)} - 1}{t} \tag{6.5}$$

remains valid for a wide class of Lévy-type processes.

Proposition 6.2. *Let $(X_t)_{t \geq 0}$ be a Lévy-type process with symbol $q(x, \xi)$ and characteristics $(b(x), Q(x), \nu(x, dy))$. Assume that q has bounded coefficients, i.e.*

$$|q(x, \xi)| \leq C(1 + |\xi|^2), \quad x, \xi \in \mathbb{R}^d,$$

for some constant $C > 0$.¹⁴ If $x \mapsto q(x, \xi)$ is continuous, then (6.5) holds.

The key step in the proof is the identity

$$\mathbb{E}^x (e_\xi(X_t)) - e_\xi(x) = \mathbb{E}^x \left(\int_0^t (Ae_\xi)(X_s) ds \right) \tag{6.6}$$

for the function $e_\xi(x) := e^{i\xi \cdot x}$. The definition of a Lévy-type process entails that

$$\mathbb{E}^x (e_\xi \chi_k)(X_t) - (e_\xi \chi_k)(x) = \mathbb{E}^x \left(\int_0^t (A(e_\xi \chi_k))(X_s) ds \right)$$

¹⁴Equivalently, $\sup_{x \in \mathbb{R}^d} (|b(x)| + |Q(x)| + \int \min\{1, |y|^2\} \nu(x, dy)) < \infty$.

for $\chi_k(x) := \chi(x/k)$ and $\chi \in C_c^\infty(\mathbb{R}^d)$ with $\mathbb{1}_{B_1(0)} \leq \chi \leq \mathbb{1}_{B_2(0)}$. In order to get (6.6), a limiting argument is then needed to pass to the limit $k \rightarrow \infty$, see Böttcher *et al.* [14, Proof Thm. 2.36] for details. Once we have (6.6) at our disposal, we note that $(Ae_\xi)(x) = -e_\xi(x)q(x, \xi)$. Using the boundedness of the coefficients and the continuity of $x \mapsto q(x, \xi)$, the dominated convergence theorem yields the assertion.

Remark 6.3. There is a variant of Proposition 6.2 for Lévy-type processes with unbounded coefficients. If the symbol q is locally bounded and continuous with respect to the x -variable, then

$$-q(x, \xi) = \lim_{t \rightarrow 0} \frac{\mathbb{E}^x \left[e^{i\xi(X_{t \wedge \tau_r^x} - x)} \right] - 1}{t}, \tag{6.7}$$

where $\tau_r^x = \inf \{t \geq 0 \mid |X_t - x| > r\}$ is the first exit time from the closed ball $\overline{B}_r(x)$. The proof is similar to the one above, combined with a standard stopping argument, see also Böttcher *et al.* [14, Cor. 2.39]. This technique was developed by Schnurr [157, Ch. 4.3] and Schilling & Schnurr [153].

6.2. Maximal tail estimates for Lévy-type processes

Let $(X_t)_{t \geq 0}$ be a Lévy-type process. We will now study probability estimates for the running supremum $\sup_{s \leq t} |X_s - X_0|$, that is, we are interested in lower and upper bounds for the probability $\mathbb{P}^x(\sup_{s \leq t} |X_s - x| \geq r)$. If we denote by

$$\tau_r^x = \inf \{t \geq 0 \mid X_t \notin \overline{B}_r(x)\} \tag{6.8}$$

the first exit time from the closed ball $\overline{B}_r(x)$, then

$$\{\tau_r^x < t\} \subseteq \left\{ \sup_{s \leq t} |X_s - x| > r \right\} \subseteq \{\tau_r^x \leq t\} \subseteq \left\{ \sup_{s \leq t} |X_s - x| \geq r \right\}, \tag{6.9}$$

and so the results can be read equivalently as estimates for $\mathbb{P}^x(\tau_r^x \leq t)$. Let us point out that (6.9) remains valid if we replace t by a random time.

Theorem 6.4. *Let $(X_t)_{t \geq 0}$ be a Lévy-type process with symbol $q(x, \xi)$. There is a constant $C > 0$ such that*

$$\mathbb{P}^x \left(\sup_{s \leq \sigma} |X_s - x| \geq r \right) \leq C \mathbb{E}^x [\sigma] \sup_{|y-x| \leq r} \sup_{|\xi| \leq r^{-1}} |q(y, \xi)| \tag{6.10}$$

for all $x \in \mathbb{R}^d$, $r > 0$ and any a.s. finite stopping time σ , i.e. $\mathbb{P}^x(\sigma < \infty) = 1$. In particular,

$$\mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \geq r \right) \leq C t \sup_{|y-x| \leq r} \sup_{|\xi| \leq r^{-1}} |q(y, \xi)|. \tag{6.11}$$

We give a direct proof using martingale methods which goes back to Schilling [144]. This approach will allow us to establish a localized version of Theorem 6.4. A much shorter proof based on Lévy's truncation inequality (see e.g. the upper bound in Lemma 2.12, or Schilling [152, Thm. 30.1] for a textbook version) can be found in Kühn [87, Thm. 1.29]. Using similar methods as [144], Schnurr [158, Prop. 3.10] has established an analogous estimate for homogeneous diffusions with jumps (in the sense of Jacod & Shiryaev [68, Def. III.2.18]). The proof of Schnurr is based on the triplet b , Q and ν of (6.2) and it completely avoids the *generator*; this means that the Markovianity of the process is not needed.

Proof of Theorem 6.4. The second assertion (6.11) follows from (6.10).

1° We first establish (6.10) with ' $> r$ ' instead of ' $\geq r$ ' on the left hand side. In view of (6.9), it suffices to show that

$$\mathbb{P}^x(\tau_r^x \leq \sigma) \leq C \mathbb{E}^x[\sigma] \sup_{|y-x| \leq r} \sup_{|\xi| \leq r^{-1}} |q(y, \xi)|.$$

Because of monotone convergence, we need to consider only bounded stopping times, i.e. $\mathbb{P}^x(\sigma \leq M) = 1$ for some $M > 0$. Fix $x \in \mathbb{R}^d$, $r > 0$, and pick $u \in C_c^\infty(\mathbb{R}^d)$ such that $u(0) = 1$, $0 \leq u \leq 1$ and $\text{supp } u \subseteq B_1(0)$. If we set $u_r^x(y) := u((y-x)/r)$, then $M_t = u_r^x(X_t) - \int_0^t Au_r^x(X_s) ds$ is a martingale, and so the optional sampling theorem yields $\mathbb{E}^x M_0 = \mathbb{E}^x M_{\sigma \wedge \tau_r^x}$, i.e.

$$\mathbb{E}^x [u_r^x(X_{\sigma \wedge \tau_r^x})] - 1 = \mathbb{E}^x \left(\int_{[0, \sigma \wedge \tau_r^x)} Au_r^x(X_s) ds \right).$$

Thus,

$$\begin{aligned} \mathbb{P}^x(\tau_r^x \leq \sigma) &\leq 1 - \mathbb{E}^x [u_r^x(X_\sigma)] = -\mathbb{E}^x \left(\int_{[0, \sigma \wedge \tau_r^x)} Au_r^x(X_s) ds \right) \\ &= -\mathbb{E}^x \left(\int_{[0, \sigma \wedge \tau_r^x)} \mathbb{1}_{\{|X_s-x| < r\}} Au_r^x(X_s) ds \right). \end{aligned}$$

Since we have

$$\begin{aligned} -Au_r^x(y) &= \int_{\mathbb{R}^d} e^{iy \cdot \xi} q(y, \xi) \widehat{u_r^x}(\xi) d\xi = e^{-ix \cdot \xi} \cdot r^d \int_{\mathbb{R}^d} e^{iy \cdot \xi} q(y, \xi) \widehat{u}(r\xi) d\xi \\ &= e^{-ix \cdot \xi} \int_{\mathbb{R}^d} e^{iy \cdot \xi} q(y, r^{-1}\xi) \widehat{u}(\xi) d\xi \end{aligned}$$

for all $y \in \mathbb{R}^d$, we find that

$$\mathbb{P}^x(\tau_r^x \leq \sigma) \leq \mathbb{E}^x \left[\int_{[0, \sigma \wedge \tau_r^x)} \int_{\mathbb{R}^d} |q(X_s, r^{-1}\xi)| |\widehat{u}(\xi)| d\xi \mathbb{1}_{\{|X_s-x| < r\}} ds \right]. \quad (6.12)$$

Using the growth estimate $\sup_{x \in K} |q(x, \xi)| \leq c_K(1 + |\xi|^2)$ for q with the constant $c_K := 2 \sup_{x \in K} \sup_{|\eta| \leq 1} |q(x, \eta)|$ for any compact set $K \subseteq \mathbb{R}^d$, cf. Böttcher *et al.* [14, Thm. 2.31], we get

$$\begin{aligned} \int_{\mathbb{R}^d} \sup_{|y-x| \leq r} |q(y, r^{-1}\xi)| |\widehat{u}(\xi)| d\xi &\leq 2 \sup_{|y-x| \leq r} \sup_{|\eta| \leq 1/r} |q(y, \eta)| \int_{\mathbb{R}^d} (1 + |\xi|^2) |\widehat{u}(\xi)| d\xi \\ &\leq C \sup_{|y-x| \leq r} \sup_{|\eta| \leq 1/r} |q(y, \eta)| \end{aligned}$$

for the absolute constant $C := 2 \int_{\mathbb{R}^d} (1 + |\xi|^2) |\widehat{u}(\xi)| d\xi < \infty$. In summary,

$$\mathbb{P}^x (\tau_r^x \leq \sigma) \leq C \mathbb{E}^x [\sigma \wedge \tau_r^x] \sup_{|y-x| \leq r} \sup_{|\eta| \leq 1/r} |q(y, \eta)|. \tag{6.13}$$

2° From

$$\mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \geq r \right) = \lim_{n \rightarrow \infty} \mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| > r - \frac{1}{n} \right)$$

and Step 1° we get

$$\begin{aligned} \mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \geq r \right) &\leq C \mathbb{E}^x [\sigma] \lim_{n \rightarrow \infty} \sup_{|y-x| \leq r-1/n} \sup_{|\xi| \leq 1/(r-1/n)} |q(y, \xi)| \\ &\leq C \mathbb{E}^x [\sigma] \sup_{|y-x| \leq r} \sup_{|\xi| \leq 2/r} |q(y, \xi)|. \end{aligned}$$

Since $\xi \mapsto \sqrt{|q(y, \xi)|}$ is subadditive, cf. (6.3), we have $|q(y, 2\xi)| \leq 4|q(y, \xi)|$, and so

$$\mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \geq r \right) \leq 4C \mathbb{E}^x [\sigma] \sup_{|y-x| \leq r} \sup_{|\xi| \leq 1/r} |q(y, \xi)|. \quad \square$$

As an immediate consequence, we obtain the following lower bound for the expectation of stopping times. This is also an alternative way to derive the lower bound in (5.15) for Lévy processes.

Corollary 6.5. *Let $(X_t)_{t \geq 0}$ be a Lévy-type process with symbol $q(x, \xi)$. There is an absolute constant $C > 0$ such that*

$$\mathbb{E}^x [\sigma] \geq C \sup_{r > 0} \frac{\mathbb{P}^x (\tau_r^x \leq \sigma < \infty)}{\sup_{|y-x| \leq r} \sup_{|\xi| \leq 1/r} |q(y, \xi)|}$$

for all $x \in \mathbb{R}^d$ and any stopping time σ . In particular,

$$\mathbb{E}^x [\tau_r^x] \geq C \frac{\mathbb{P}^x (\tau_r^x < \infty)}{\sup_{|y-x| \leq r} \sup_{|\xi| \leq 1/r} |q(y, \xi)|}$$

Proof. From (6.13) with $\sigma \rightsquigarrow \sigma \wedge t$, we see that

$$\mathbb{P}^x (\tau_r^x \leq \sigma \wedge t) \leq C \mathbb{E}^x [t \wedge \sigma \wedge \tau_r^x] \sup_{|y-x| \leq r} \sup_{|\eta| \leq 1/r} |q(y, \eta)|.$$

Letting $t \rightarrow \infty$ yields

$$\mathbb{P}^x (\tau_r^x \leq \sigma < \infty) \leq C \mathbb{E}^x [\sigma] \sup_{|y-x| \leq r} \sup_{|\eta| \leq 1/r} |q(y, \eta)|$$

for any $r > 0$. Rearranging the terms gives the first assertion, and then the second one is immediate if we take $\sigma := \tau_r^x$. \square

Next we present a localized version of the maximal inequality, which deals with the small-time asymptotics of $\mathbb{P}^x (\sup_{s \leq t} |X_s - x| > r)$. It has been established in Kühn & Schilling [96] for Feller processes, but the proof remains valid for the wider class of Lévy-type processes.

Theorem 6.6. *Let $(X_t)_{t \geq 0}$ be a Lévy-type process with symbol $q(x, \xi)$. Assume that $x \mapsto q(x, \xi)$ is continuous for any $\xi \in \mathbb{R}^d$. There is an absolute constant $C > 0$ such that*

$$\limsup_{t \rightarrow 0} \frac{1}{t} \mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \geq r \right) \leq C \sup_{|\xi| \leq 1/r} |q(x, \xi)| \tag{6.14}$$

for any $x \in \mathbb{R}^d$ and $r > 0$.

Remark 6.7. The symbol q is continuous with respect to the x -variable if, and only if, the pseudo-differential operator A with symbol q satisfies $Af \in C(\mathbb{R}^d)$ for any $f \in C_c^\infty(\mathbb{R}^d)$. There is also a characterization in terms of the triplets $(b(x), Q(x), \nu(x, dy))$, see Kühn & Kunze [95, Thm. B.1]. If $(X_t)_{t \geq 0}$ is a Feller process and the domain of the (strong) infinitesimal generator contains $C_c^\infty(\mathbb{R}^d)$, then $x \mapsto q(x, \xi)$ is continuous if, and only if, $x \mapsto q(x, 0)$ is continuous; if we have a conservative Feller process, then this is always satisfied since $q(x, 0) = 0$, cf. Schilling [146, Thm. 4.4] or Böttcher *et al.* [14, Thm. 2.30].

Proof of Theorem 6.6. Let u be the function from the proof of Theorem 6.4. From (6.12) with $\sigma \equiv t$, we see that

$$\mathbb{P}^x (\tau_r^x \leq t) \leq t \mathbb{E}^x \left(\int_{\mathbb{R}^d} \sup_{s < t \wedge \tau_r^x} |q(X_s, r^{-1}\xi)| |\widehat{u}(\xi)| d\xi ds \right).$$

Since $X_s \in \overline{B}_r(x)$ for all $s < t \wedge \tau_r^x$, there is a constant $C = C(r, x)$ such that

$$\sup_{s < t \wedge \tau_r^x} |q(X_s, r^{-1}\xi)| \leq C(1 + |\xi|^2) |\widehat{u}(\xi)| \in L^1(d\xi)$$

for all $t \geq 0$. Moreover, the continuity of $x \mapsto q(x, \xi)$ entails that

$$\lim_{t \rightarrow 0} \sup_{s < t \wedge \tau_r^x} |q(X_s, r^{-1}\xi)| |\widehat{u}(\xi)| = |q(x, r^{-1}\xi)| |\widehat{u}(\xi)|.$$

Applying dominated convergence yields

$$\limsup_{t \rightarrow 0} \frac{1}{t} \mathbb{P}^x (\tau_r^x \leq t) \leq \int_{\mathbb{R}^d} |q(x, r^{-1}\xi)| |\widehat{u}(\xi)| d\xi.$$

Taking into account (6.9) and the estimate

$$|q(x, r^{-1}\xi)| \leq 2 \sup_{|\eta| \leq 1/r} |q(x, \eta)|(1 + |\xi|^2)$$

gives

$$\limsup_{t \rightarrow 0} \frac{1}{t} \mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| > r \right) \leq C \sup_{|\xi| \leq 1/r} |q(x, \xi)|.$$

In order to get (6.14) with ' $\geq r$ ', we can almost literally follow the strategy used in Step 2° of the proof of Theorem 6.4. \square

Example 6.8. Let $(X_t)_{t \geq 0}$ be a process of variable order, i.e. a Lévy-type process with symbol $q(x, \xi) = |\xi|^{\alpha(x)}$ for a continuous function $\alpha: \mathbb{R}^d \rightarrow (0, 2]$, cf. Example 6.1 No. 5. Then Theorem 6.4 and Theorem 6.6 show that

$$\mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \geq r \right) \leq C t r^{-\alpha_*(x, r)}$$

and

$$\limsup_{t \rightarrow 0} \frac{1}{t} \mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \geq r \right) \leq C r^{-\alpha(x)}$$

for $r > 0$ and some absolute constant $C > 0$, where $\alpha_*(x, r) = \min_{|z-x| \leq r} \alpha(z)$.

Example 6.9. Let $(X_t)_{t \geq 0}$ be the unique weak solution to an SDE

$$dX_t = \sigma(X_{t-}) dL_t,$$

for a one-dimensional Lévy process $(L_t)_{t \geq 0}$ and $\sigma \in C_b(\mathbb{R})$. The symbol of $(X_t)_{t \geq 0}$ is $q(x, \xi) = \psi(\sigma(x)\xi)$ where ψ is the characteristic exponent of $(L_t)_{t \geq 0}$, cf. Example 6.1, and so Theorems 6.4 and 6.6 yield

$$\mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \geq r \right) \leq C t \sup_{|y-x| \leq r} \sup_{|\xi| \leq r^{-1}} |\psi(\sigma(y)\xi)|$$

and

$$\limsup_{t \rightarrow 0} \frac{1}{t} \mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \geq r \right) \leq C |\psi(\sigma(x)\xi)|.$$

If σ satisfies a growth estimate $|\sigma(x)| \leq M(1 + |x|^\gamma)$ for some constants $\gamma \geq 0$, $M > 0$, then we get

$$\mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \geq r \right) \leq C' t \sup_{|\eta| \leq r^{\gamma-1}} |\psi(\eta)| = |\psi|^*(r^{\gamma-1})$$

for $r \gg 1$.

So far, we have established upper bounds for the tails of the maximum. Next we turn to lower bounds. Lower bounds for $\mathbb{P}^x(\sup_{s \leq t} |X_s - x| > r)$ are of course equivalent to upper bounds for the probability $\mathbb{P}^x(\sup_{s \leq t} |X_s - x| \leq r)$ of the complementary event. Since we know from (6.9) that

$$\mathbb{P}^x(\tau_r^x \geq t) \geq \mathbb{P}^x\left(\sup_{s \leq t} |X_s - x| \leq r\right) \geq \mathbb{P}^x(\tau_r^x > t) \geq \mathbb{P}^x\left(\sup_{s \leq t} |X_s - x| < r\right),$$

this means that the estimates can also be stated as tail estimates for the first exit time τ_r^x .

We start with a bound in terms of the characteristics $(b(x), Q(x), \nu(x, dy))$. The basic idea is quite simple: the process $(X_t)_{t \geq 0}$ leaves the ball $B_r(x)$ centred at the starting point x as soon as a jump of modulus larger than $2r$ occurs. Thus we can estimate the exit probability by considering only the compound pseudo Poisson process (see Example 6.1 No. 4) accounting for the jumps of $(X_t)_{t \geq 0}$ exceeding absolute size $2r$. The likelihood that a jump of size $> 2r$ happens has a lower bound which is proportional to $\inf_z \nu(z, \{y \in \mathbb{R}^d \mid |y| > 2r\})$.

Proposition 6.10. *Let $(X_t)_{t \geq 0}$ be a Lévy-type process with infinitesimal characteristics $(b(x), Q(x), \nu(x, dy))$, $x \in \mathbb{R}^d$. If σ is a finite stopping-time, then*

$$\mathbb{P}^x\left(\sup_{s \leq \sigma} |X_s - x| > r\right) \geq \mathbb{E}^x[\tau_r^x \wedge \sigma] \cdot G(x, 2r) \quad (6.15)$$

for all $x \in \mathbb{R}^d$ and $r > 0$, where¹⁵

$$G(x, r) := \inf_{|z-x| \leq r} \nu(z, \{y \in \mathbb{R}^d \mid |y| > r\}). \quad (6.16)$$

In particular,

$$\mathbb{P}^x\left(\sup_{s \leq t} |X_s - x| > r\right) \geq \frac{t G(x, 2r)}{1 + t G(x, 2r)}. \quad (6.17)$$

Note that (6.17) is equivalent to

$$\mathbb{P}^x\left(\sup_{s \leq t} |X_s - x| \leq r\right) \leq \frac{1}{1 + t G(x, 2r)}; \quad (6.18)$$

this bound will be improved in Corollary 6.11 below.

Proof of Proposition 6.10. Fix $x \in \mathbb{R}^d$, $\epsilon > 0$ and $r > 0$. By monotone convergence, it suffices to prove the assertion for bounded stopping times σ . Pick a function $u \in C_c^\infty(\mathbb{R}^d)$ such that $\mathbb{1}_{B_{r+\epsilon}(x)} \leq u \leq \mathbb{1}_{B_{r+2\epsilon}(x)}$. Denote by τ_r^x the first exit time from the ball $\bar{B}_r(x)$. As $u(X_t) = 1$ for $t < \tau_r^x$, we find from Dynkin's formula that

¹⁵Our notation is chosen to be in-line with the truncated and censored second moments which were introduced on page 2.3, Eq. (2.3).

$$\begin{aligned} \mathbb{P}^x(\tau_r^x > \sigma) &= \mathbb{E}^x[u(X_\sigma)\mathbf{1}_{\{\tau_r^x > t\}}] \\ &\leq \mathbb{E}^x[u(X_{\sigma \wedge \tau_r^x})] = 1 + \mathbb{E}^x\left[\int_{(0, \sigma \wedge \tau_r^x)} Au(X_s) ds\right]. \end{aligned}$$

For $z \in B_{r+\epsilon}(x)$ we have $u(z) = 1$, $\nabla u(z) = 0$ and $\nabla^2 u(z) = 0$, so that

$$Au(z) = \int_{y \neq 0} (u(z+y) - 1) \nu(z, dy), \quad z \in B_{r+\epsilon}(x).$$

Since the integrand is negative and $u = 0$ outside the ball $B_{r+2\epsilon}(x)$, it follows that

$$Au(z) \leq \int_{|(z+y)-x| \geq r+2\epsilon} (u(z+y) - 1) \nu(z, dy) \leq - \int_{|y| \geq 2r+3\epsilon} \nu(z, dy)$$

for all $z \in B_{r+\epsilon}(x)$. By definition, $X_s \in \bar{B}_r(x)$ for $s < \tau_r^x$, and so we may use the bound for $z = X_s$, $s < \tau_r^x$, to find that

$$\begin{aligned} \mathbb{P}^x(\tau_r^x > \sigma) &\leq 1 - \mathbb{E}^x\left[\int_{(0, \sigma \wedge \tau_r^x)} \nu(X_s, \{y \in \mathbb{R}^d \mid |y| \geq 2r + 3\epsilon\}) ds\right] \\ &\xrightarrow{\epsilon \downarrow 0} 1 - \mathbb{E}^x\left[\int_{(0, \sigma \wedge \tau_r^x)} \nu(X_s, \{y \in \mathbb{R}^d \mid |y| \geq 2r\}) ds\right]. \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{P}^x(\tau_r^x > \sigma) &\leq 1 - \mathbb{E}^x[\tau_r^x \wedge \sigma] \inf_{|z-x| \leq r} \nu(z, \{y \in \mathbb{R}^d \mid |y| > 2r\}). \quad (6.19) \\ &\leq 1 - \mathbb{E}^x[\tau_r^x \wedge \sigma] G(x, 2r). \end{aligned}$$

Since $\{\tau_r^x \geq \sigma\} = \bigcap_n \{\tau_r^x > \sigma - 1/n\}$, a standard continuity-of-measure argument shows that the same estimate holds for $\mathbb{P}^x(\tau_r^x \geq \sigma)$. In particular, by (6.9),

$$\mathbb{P}^x\left(\sup_{s \leq \sigma} |X_s - x| > r\right) \geq \mathbb{E}^x[\tau_r^x \wedge \sigma] G(x, 2r).$$

For $\sigma \equiv t$ we use in (6.19) the elementary estimate $\mathbb{E}^x[t \wedge \tau_r^x] \geq t\mathbb{P}^x(\tau_r^x \geq t)$ to see

$$\mathbb{P}^x(\tau_r^x \geq t) \leq 1 - t\mathbb{P}^x(\tau_r^x \geq t) G(x, 2r),$$

i.e.

$$\mathbb{P}^x\left(\sup_{s \leq t} |X_s - x| \leq r\right) \leq \mathbb{P}^x(\tau_r^x \geq t) \leq \frac{1}{1 + tG(x, 2r)}.$$

In other words,

$$\mathbb{P}^x\left(\sup_{s \leq t} |X_s - x| > r\right) \geq \frac{tG(x, 2r)}{1 + tG(x, 2r)}. \quad \square$$

Let $N_t := \#\{s \leq t \mid |\Delta X_s| > 2r\}$ be the process counting the jumps having modulus larger than $2r$. Clearly, $\{\sup_{s \leq t} |X_s - x| \leq r\} \subseteq \{N_t = 0\}$. For the particular case that $(X_t)_{t \geq 0}$ is a Lévy process, it is known that $(N_t)_{t \geq 0}$ is a Poisson process with intensity $\lambda = \nu(\{y \in \mathbb{R}^d \mid |y| > 2r\})$, where ν denotes the Lévy measure, see e.g. Schilling [148, Ch. 9]. Consequently,

$$\begin{aligned} \mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \leq r \right) &\leq \mathbb{P}^x(N_t = 0) \\ &= e^{-\lambda t} = \exp[-t\nu(\{y \in \mathbb{R}^d \mid |y| > 2r\})] \end{aligned}$$

for a Lévy process $(X_t)_{t \geq 0}$. If $(X_t)_{t \geq 0}$ is a general Lévy-type process, $(N_t)_{t \geq 0}$ is no longer a Poisson process (but a pseudo Poisson process), but the following result shows that we can still expect a similar estimate in terms of the jump intensity.

Corollary 6.11. *Let $(X_t)_{t \geq 0}$ be a Lévy-type process. Then*

$$\mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \leq r \right) \leq \exp(-tG(x, 5r)), \quad x \in \mathbb{R}^d, r > 0, t > 0 \quad (6.20)$$

and

$$\mathbb{E}^x [\tau_r^x] \leq \frac{1}{G(x, 2r)}, \quad x \in \mathbb{R}^d, r > 0, \quad (6.21)$$

for the function G defined in (6.16).

Proof. For fixed $x \in \mathbb{R}^d$, $r > 0$, $n \in \mathbb{N}$ and $t > 0$, set $t_i := \frac{i}{n}t$, then

$$\begin{aligned} \mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \leq r \right) \\ \leq \mathbb{P}^x \left(\bigcap_{i=0}^{n-1} \left\{ |X_{t_i} - x| \leq r, \sup_{s \leq t_{i+1} - t_i} |X_{t_i+s} - X_{t_i}| \leq 2r \right\} \right). \end{aligned}$$

Applying the Markov property and using that, by Proposition 6.10,

$$\mathbb{P}^y \left(\sup_{s \leq \rho} |X_s - y| \leq 2r \right) \leq \frac{1}{1 + \rho G(y, 4r)} \leq \frac{1}{1 + \rho G(x, 5r)}$$

for any $y \in \overline{B}_r(x)$ and $\rho > 0$, we get

$$\mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \leq r \right) \leq \left(\frac{1}{1 + \frac{t}{n} G(x, 5r)} \right)^n \xrightarrow{n \rightarrow \infty} \exp[-tG(x, 5r)],$$

which proves (6.20). For (6.21), we note that by (6.19) with $\sigma \equiv t$

$$\mathbb{E}^x [\tau_r^x \wedge t] \leq \frac{1 - \mathbb{P}^x(\tau_r^x > t)}{G(x, 2r)} \leq \frac{1}{G(x, 2r)}.$$

Letting $t \rightarrow \infty$ using Fatou's lemma gives (6.21). \square

The rate in (6.20) is optimal for a wide class of jump process. However, the estimate uses only the tails of the Lévy measure, and therefore not the full information is taken into account, as the following example shows. Recall that we have already seen a similar phenomenon in connection with Lévy processes, see Theorem 5.13 and its proof.

Example 6.12. Let $(X_t)_{t \geq 0}$ be an α -stable Lévy process for $\alpha \in (0, 1)$. Taylor [172] has shown that the bounds for $\mathbb{P}(\sup_{s \leq t} |X_s| \leq r)$ depend substantially on the support of the Lévy measure ν : if the spectral measure (i.e. the spherical part of ν) is concentrated in a hemisphere $\{y \in \mathbb{S}^{d-1} \mid y_i \geq 0\}$ for some $i \in \{1, \dots, d\}$, then

$$\mathbb{P}\left(\sup_{s \leq t} |X_s| \leq r\right) \asymp \exp\left(-ct^{1/(1-\alpha)}r^{-\alpha/(1-\alpha)}\right) \tag{6.22}$$

for small $r > 0$, where the constants c in the lower and upper bound may differ. For all other stable processes, it holds that

$$\mathbb{P}\left(\sup_{s \leq t} |X_s| \leq r\right) \asymp \exp(-ctr^{-\alpha}) \tag{6.23}$$

for small $r > 0$, again with possibly different constants in the lower and upper bound. Corollary 6.11 yields the estimate $\mathbb{P}(\sup_{s \leq t} |X_s| \leq r) \leq \exp[-ctr^{-\alpha}]$ for any stable process, i.e. it does not capture the information on the support of the Lévy measure.

Theorem 6.13. *Let $(X_t)_{t \geq 0}$ be a Lévy-type process with symbol $q(x, \xi)$. Then*

$$\mathbb{P}^x\left(\sup_{s \leq t} |X_s - x| \leq r\right) \leq \frac{c}{1 + t \sup_{|\xi| \leq 1/(rk(x,r))} \inf_{|y-x| \leq r} \operatorname{Re} q(y, \xi)}, \tag{6.24}$$

where $c := 4/\cos\sqrt{\frac{2}{3}}$ and¹⁶

$$k(x, r) := \inf \left\{ k \geq \left(\arccos\sqrt{\frac{2}{3}}\right)^{-1} \mid \inf_{|\xi| \leq 1/(kr)} \inf_{|y-x| \leq r} \frac{\operatorname{Re} q(y, \xi)}{|\xi| |\operatorname{Im} q(y, \xi)|} \geq 2r \right\}.$$

The estimate (6.24) was first shown in Schilling [144, Lem. 6.3] for Feller processes whose symbol satisfy a sector condition, see also Böttcher *et al.* [14, Thm. 5.5].¹⁷ It turns out that the proof goes through for the wider class of Lévy-type processes without any major modifications, and so we do not present the proof in this exposition.

¹⁶The definition of $k(x, r)$ in Böttcher *et al.* [14, Theorem 5.5] has a misprint which we corrected here: $\sup \inf$ should read $\inf \inf$.

¹⁷The additive constant ‘1+’ in the denominator in (6.24) comes from the trivial observation that a probability is less than 1, i.e. taking the minimum with 1 in $\mathbb{P}^x(\dots) \leq \frac{c}{p}$, we get $\frac{c}{p} \wedge 1 \leq c'(1+p)^{-1}$.

The definition of $k(x, r)$ is tailor-made in such a way that

$$\operatorname{Re} q(y, \xi) \geq 2r|\xi| |\operatorname{Im} q(y, \xi)|$$

for all $|\xi| \leq 1/(rk(x, r))$ and $|y - x| \leq r$ – this is exactly what is needed for some technical estimates in the proof of Theorem 6.13. If the symbol q satisfies the *sector condition*, i.e.

$$|\operatorname{Im} q(x, \xi)| \leq C \operatorname{Re} q(x, \xi), \quad x, \xi \in \mathbb{R}^d, \quad (6.25)$$

for some constant $C > 0$, then $k(x, r) = k_0 := \max \left\{ 2C, (\arccos \sqrt{2/3})^{-1} \right\}$, and so we arrive at the following statement.

Corollary 6.14. *Let $(X_t)_{t \geq 0}$ be a Lévy-type process with symbol $q(x, \xi)$ satisfying the sector condition. Then there is a constant $c > 0$ depending only on the sector constant such that*

$$\mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \leq r \right) \leq \frac{c}{1 + t \sup_{|\xi| \leq 1/(rk_0)} \inf_{|y-x| \leq r} \operatorname{Re} q(y, \xi)} \quad (6.26)$$

for all $x \in \mathbb{R}^d$, $r > 0$ and $t > 0$.

Using an iteration argument based on the Markov property – similar to the proof of Corollary 6.11 –, we can improve the above estimates to obtain an exponential bound.

Corollary 6.15. *Let $(X_t)_{t \geq 0}$ be a Lévy-type process with symbol $q(x, \xi)$. Then*

$$\mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| \leq r \right) \leq \exp \left[-ct \sup_{|\xi| \leq 1/(2rk^*(x, r))} \inf_{|y-x| \leq 3r} \operatorname{Re} q(y, \xi) \right]$$

for $c := \cos \sqrt{2/3}/(4e)$ and

$$k^*(x, r) := \inf \left\{ k \geq \left(\arccos \sqrt{2/3} \right)^{-1} \mid \inf_{|\xi| \leq 1/(2kr)} \inf_{|y-x| \leq 3r} \frac{\operatorname{Re} q(y, \xi)}{|\xi| |\operatorname{Im} q(y, \xi)|} \geq 4r \right\}.$$

Moreover,

$$\mathbb{E}^x [\tau_r^x] \leq \frac{C}{\sup_{|\xi| \leq 1/(2rk^*(x, r))} \inf_{|y-x| \leq 3r} \operatorname{Re} q(y, \xi)}.$$

If the symbol q satisfies the sector condition (6.25), then $k^*(x, r)$ may be replaced by $k_0 = \max \left\{ 2C, (\arccos \sqrt{2/3})^{-1} \right\}$.

Remark 6.16. Maximal inequalities and, in particular, the estimates on the mean exit time have interesting applications in the study of path properties and distributional properties of Lévy-type processes. For example, one can prove

- *Harnack inequalities* for Lévy-type generators following the strategy of Bass & Levin [6] – see also Song & Vondraček [164] for a systematic exposition of this method and the comments in Example 5.17; this method is explained in Schilling & Uemura [154, §5].
- *tightness estimates* for Lévy-type process $(X_t)_{t \geq 0}$ which can be applied to show the continuity of expressions of the form $x \mapsto \mathbb{E}^x [f(X_{\tau_U})]$ for exit times τ_U , the (*strong*) *Feller property* for certain Lévy-type processes, see Schilling & Wang [155], and the existence of *Markovian selections* for martingale problems, see Kühn [92].
- *moment estimates* for Lévy-type processes using the layer-cake formula (Lemma 2.20), see Deng & Schilling [30], Kühn [88] and Theorem 5.19 for an illustration of the method in the Lévy case.
- results on *upper and lower functions* and *small-time asymptotics* for Lévy-type processes as in Knopova & Schilling [77], Kühn [94], Kühn & Schilling [96].
- estimates on the *Hausdorff dimension* of the image of a Feller process Schilling [145], Knopova *et al.* [78].
- results on the *p-variation* of the sample paths, see Manstavičius [117], Manstavičius & Schnurr [118].
- results on *random time changes* of Lévy-type processes, cf. Krühner & Schnurr [85], Kühn [90].

7. Markov processes

Let us resume the discussion started in Chapter 3 and turn to a situation where we want to make as few assumptions on the stochastic process as possible. Throughout, $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F}_t)$ is a filtered probability space which satisfies the usual conditions, i.e. it is complete and the filtration is right continuous and \mathcal{F}_0 contains all \mathbb{P} -null sets. Let $(X_t)_{t \geq 0}$ be a strong Markov process taking values in \mathbb{R}^d . We will consider only temporally homogeneous processes; in this case, $p_t(x, dy) = \mathbb{P}^x(X_t \in dy)$ is the transition function, and we write $\mathbb{E}^x(\dots) = \int_{\Omega} \dots d\mathbb{P}^x$ for the corresponding expectation. We assume that $t \mapsto X_t(\omega)$ is, almost surely for all \mathbb{P}^x , right continuous with finite left limits (càdlàg). The standard argument how one can construct a càdlàg modification can be found in Revuz & Yor [141, Thm. III.(2.7)] (for Feller processes, i.e. Markov processes whose transition semigroup $P_t u(x) := \mathbb{E}^x u(X_t) = \int u(y) p_t(x, dy)$ preserves the family C_∞ of continuous functions vanishing at infinity) or in Sharpe [159, Thm. I.(2.7)] (for processes satisfying Meyer's first *hypothèse droite* (HD1)).

Definition 7.1. A (strong) Markov process $(X_t)_{t \geq 0}$ is said to be *uniformly stochastically continuous* if

$$\forall \epsilon > 0, \forall t \geq 0 : \limsup_{s \rightarrow t} \sup_{x \in \mathbb{R}^d} \mathbb{P}^x(|X_s - X_t| > \epsilon) = 0. \quad (7.1)$$

Because of the Markov property, it is enough to require (7.1) for $t = 0$.

Since we do not assume further structural properties on a Markov process, we will get only a few results on the maximal process X_t^* . The first result is a version of Etemadi's inequality, see Theorem 5.11.

Theorem 7.2 (Etemadi-type maximal inequality). *Assume that $(X_t)_{t \geq 0}$ is a strong Markov process. For all $a > 0$ and $x \in \mathbb{R}^d$ one has*

$$\mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| > 3a \right) \leq 2 \sup_{s \leq t} \sup_{y \in B_{3a}^c(x)} \mathbb{P}^y(|X_s - y| > a). \quad (7.2)$$

Proof. Denote by $\tau := \tau_{3a}^x := \inf \{s > 0 \mid |X_s - x| > 3a\}$ the first exit time from the ball $\bar{B}_{3a}(x)$. Because of the inclusion $\{\sup_{s \leq t} |X_s - x| > 3a\} \subseteq \{\tau \leq t\}$ we have

$$\begin{aligned} \mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| > 3a \right) &\leq \mathbb{P}^x(\tau \leq t, |X_t - x| > a) + \mathbb{P}^x(\tau \leq t, |X_t - x| \leq a) \\ &\leq \mathbb{P}^x(|X_t - x| > a) + \mathbb{P}^x(\tau \leq t, |X_t - x| \leq a) \\ &\leq \mathbb{P}^x(|X_t - x| > a) + \mathbb{P}^x(\tau \leq t, |X_t - X_\tau| \geq 2a). \end{aligned}$$

In the last line we use the triangle inequality $|X_t - X_\tau| \geq |X_\tau - x| - |X_t - x|$ and the fact that $|X_\tau - x| \geq 3a$. We will now estimate the second term appearing on the right hand side. Using the strong Markov property we get

$$\begin{aligned} \mathbb{P}^x(\tau \leq t, |X_t - X_\tau| \geq 2a) &= \int_{\tau \leq t} \mathbb{P}^{X_\tau(\omega)}(|X_{t-\tau(\omega)} - X_0| \geq 2a) \mathbb{P}^x(d\omega) \\ &\leq \sup_{s \leq t} \sup_{y \in B_{3a}^c(x)} \mathbb{P}^y(|X_s - y| \geq 2a) \mathbb{P}^x(\tau \leq t) \\ &\leq \sup_{s \leq t} \sup_{y \in B_{3a}^c(x)} \mathbb{P}^y(|X_s - y| \geq 2a). \end{aligned} \quad (7.3)$$

In the penultimate estimate we use that $X_\tau \in B_{3a}^c(x)$. Both estimates together prove

$$\begin{aligned} \mathbb{P}^x \left(\sup_{s \leq t} |X_s - x| > 3a \right) &\leq \mathbb{P}^x(|X_t - x| > a) + \mathbb{P}^x(\tau \leq t, |X_t - X_\tau| \geq 2a) \\ &\leq \mathbb{P}^x(|X_t - x| > a) + \sup_{s \leq t} \sup_{y \in B_{3a}^c(x)} \mathbb{P}^y(|X_s - y| \geq 2a) \\ &\leq \mathbb{P}^x(|X_t - x| > a) + \sup_{s \leq t} \sup_{y \in B_{3a}^c(x)} \mathbb{P}^y(|X_s - y| > a). \quad \square \end{aligned}$$

If we assume uniform stochastic continuity, then we get the following Kolmogorov- or Ottaviani-Skorokhod-type inequality. Our proof follows – and simplifies – Itô's presentation [65, Ch. 2.9]. Alternatively, one could use the argument sketched in Remark 5.12 and the penultimate estimate in (7.3), without crudely estimating $\mathbb{P}^x(\tau \leq 1)$ with 1 to get an equivalent version of the next lemma,

see also Kühn [94, Lemma 5.4]. Further proofs of similar results are in Blumenthal [13, Lem. 2.1, 2.2] and Gikhman & Skorokhod [47, p. 420, Lem. 2].

Theorem 7.3 (Ottaviani–Skorokhod-type maximal inequality). *Assume that $(X_t)_{t \geq 0}$ is a uniformly stochastically continuous strong Markov process. For any $\epsilon > 0$ there is some $\delta = \delta(\epsilon)$ such that for all $t, s \geq 0$ satisfying $0 < t - s < \delta$*

$$\mathbb{P}^x \left(\sup_{s \leq u, v \leq t} |X_u - X_v| > 4\epsilon \right) \leq 4\mathbb{P}^x (|X_s - X_t| > \epsilon). \tag{7.4}$$

Proof. Pick any $\epsilon > 0$. Because of uniform stochastic continuity, there is some $\delta = \delta(\epsilon) > 0$ such that

$$\sup_{x \in \mathbb{R}^d, h < \delta} \mathbb{P}^x (|X_h - x| \geq \epsilon) \leq \frac{1}{2},$$

or, equivalently,

$$\inf_{x \in \mathbb{R}^d, h < \delta} \mathbb{P}^x (|X_h - x| < \epsilon) \geq \frac{1}{2}. \tag{7.5}$$

Using the triangle inequality, (7.4) follows from

$$\mathbb{P}^x \left(\sup_{s \leq u \leq t} |X_u - X_s| > 2\epsilon \right) \leq 2\mathbb{P}^x (|X_s - X_t| > \epsilon). \tag{7.6}$$

Denote by $\tau = \tau_{B_{2\epsilon}(z)}$ the first exit time from the open ball $B_{2\epsilon}(z)$ of radius 2ϵ and centre $z \in \mathbb{R}^d$. Combining (7.5) and the strong Markov property yields for $t - s < \delta$

$$\begin{aligned} & \mathbb{P}^z (\tau \leq t - s) \\ &= \mathbb{P}^z (|X_\tau - z| \geq 2\epsilon, \tau \leq t - s) \\ &\leq \int \underbrace{2\mathbb{P}^{X_\tau(\omega)} (|X_{t-s-\tau(\omega)} - X_0| < \epsilon)}_{\geq 1, \text{ by (7.5)}} \mathbf{1}_{\{|X_\tau - z| \geq 2\epsilon\}}(\omega) \mathbf{1}_{\{\tau \leq t-s\}}(\omega) \mathbb{P}^z(d\omega) \\ &= 2\mathbb{P}^z (|X_\tau - z| \geq 2\epsilon, \tau \leq t - s, |X_{t-s} - X_\tau| < \epsilon) \\ &\leq 2\mathbb{P}^z (|X_{t-s} - z| > \epsilon, \tau \leq t - s) \\ &\leq 2\mathbb{P}^z (|X_{t-s} - z| > \epsilon). \end{aligned}$$

Again by the Markov property,

$$\begin{aligned} \mathbb{P}^x \left(\sup_{s \leq u \leq t} |X_u - X_s| > 2\epsilon \right) &= \mathbb{E}^x \left[\mathbb{P}^{X_s} \left(\sup_{u \leq t-s} |X_u - X_0| > 2\epsilon \right) \right] \\ &\leq 2\mathbb{E}^x \left[\mathbb{P}^{X_s} (|X_{t-s} - X_0| > \epsilon) \right] \\ &= 2\mathbb{P}^x (|X_t - X_s| > \epsilon), \end{aligned}$$

and this proves (7.6). □

Maximal estimates of this type can be used to prove sample path regularity of strong Markov processes. For example, we have *quasi-left continuity*, i.e. if σ_n and σ are stopping times, then

$$\sigma_n \uparrow \sigma < \infty \text{ a.s.} \implies \mathbb{P}^x \left(\lim_{n \rightarrow \infty} X_{\sigma_n} = X_\sigma \right) = 1,$$

see Itô [65, Ch. 2.9, Thm. 2], or we can prove the following

Theorem 7.4 (Dynkin–Kinney criterium). *Assume that $(X_t)_{t \geq 0}$ is a strong Markov process satisfying the condition*

$$\forall \epsilon > 0 : \lim_{t \rightarrow 0} \frac{1}{t} \sup_{x \in \mathbb{R}^d} \mathbb{P}^x (|X_t - x| > \epsilon) = 0, \quad (7.7)$$

then $(X_t)_{t \geq 0}$ has a modification with continuous sample paths.

Obviously, the condition (7.7) implies uniform stochastic continuity. Hence, we can use Theorem 7.3 to get a rather simple proof of the Dynkin–Kinney criterium, see Itô [65, Ch. 2.9, Thm. 3]. For a more standard proof, we refer to Wentzell [177, Ch. 9.1] or Dynkin [36, Ch. 6, §5].

If we assume a bit more about the process $(X_t)_{t \geq 0}$, e.g. that it is Feller and that the test functions $C_c^\infty(\mathbb{R}^d)$ are in the domain of the infinitesimal generator, then (7.7) is essentially equivalent to the continuity of the sample paths or to the fact that the infinitesimal generator is a second-order differential operator, see the discussion in Schilling [151, Thm. 7.38, Thm. 7.39, Thm. 23.3].

A further application of such maximal inequalities are estimates on the p -variation of stochastic processes, see Manstavičius [117]; typically, the results of Chapter 6 (e.g. Theorem 6.4) in combination with Theorem 7.3 establish the conditions needed in Manstavičius [117], see Böttcher *et al.* [14, Ch. 5.4].

Remark 7.5 (on sublinear processes). It is also possible (and interesting) to study maximal inequalities for sublinear expectations. A *sublinear* Markov semigroup $(T_t)_{t \geq 0}$ satisfies the usual semigroup property $T_{t+s} = T_t T_s$, but each operator T_t is only subadditive and positively homogeneous, i.e.

$$T_t(f + g) \leq T_t(f) + T_t(g) \quad \text{and} \quad T_t(\lambda f) = \lambda T_t(f),$$

for all $\lambda \geq 0$ and all bounded measurable functions f, g . Sublinear Markov semigroups appear in the study of stochastic processes on sublinear expectation spaces, which can be interpreted as stochastic processes under uncertainty, cf. Hollender [60]. Using techniques similar to those in Chapter 6, it is possible to derive a maximal inequality for this class of processes, cf. Kühn [91, Proposition 5.1]. Applications include an analogue of the Courrège-von Waldenfels theorem, cf. Kühn [93], and the study of sublinear Hamilton–Jacobi–Bellman equations, see Denk *et al.* [31] and Hollender [60] for further details and references.

8. Gaussian processes

The methods used in the study of Gaussian processes are quite different from those introduced in the previous sections; quite often, they come from discrete

mathematics or functional analysis or from a combination of both. Moreover, the full strength of these methods unfolds, if one considers Gaussian processes which are indexed by elements of a (compact) metric space. As before, we restrict ourselves to processes indexed by $[0, \infty)$ and taking values in \mathbb{R} – this is enough to illustrate some of the methods used in the theory. Our standard references for Gaussian processes are the books by Lifshits [109] and [110]. Gaussian random variables and Gaussian processes with values in Banach spaces are discussed in Ledoux & Talagrand [102].

Definition 8.1. A random process $(X_t)_{t \geq 0}$ taking values in \mathbb{R} is said to be Gaussian, if every vector $(X_{t_1}, \dots, X_{t_n})$, $0 \leq t_1 < \dots < t_n$ is an n -dimensional Gaussian random variable.

By definition, a Gaussian random process is uniquely determined through the mean $m(t) = \mathbb{E}[X_t]$ and the covariance $K(t, s) = \mathbb{E}[X_t X_s] - m(t)m(s)$ functions.

Example 8.2. Typical examples of Gaussian processes (indexed by $[0, \infty)$) are:

1. *Brownian motion:* $K(t, s) = t \wedge s$ and $m(t) \equiv 0$.
2. *Fractional Brownian motion:* $K(t, s) = \frac{1}{2}(t^\alpha + s^\alpha - |s - t|^\alpha)$ with the Hurst index $H = \alpha/2 \in (0, 1]$ and $m(t) \equiv 0$.
3. *Bi-fractional Brownian motion:* $K(t, s) = \frac{1}{2}((t^\alpha + s^\alpha)^k - |s - t|^{\alpha k})$ with $\alpha, k \in (0, 2]$, $\alpha k \leq 2$ and $m(t) \equiv 0$.
4. *Gaussian Markov processes:* Let $f, g: [0, \infty) \rightarrow [0, \infty)$ be functions such that g is strictly increasing. The process $X_t := f(t)W_{g(t)}$ where $(W_t)_{t \geq 0}$ is a standard Brownian motion, is a mean-zero Gaussian Markov process. Conversely, every mean-zero Gaussian Markov process is of this form, see Marcus & Rosen [120, Chap. 5.1.1].

Examples in this class are Brownian motion, the Brownian bridge ($K(s, t) = s \wedge t - st$, $X_t = (1 + t)W_{t(1+t)^{-1}}$) or the Ornstein–Uhlenbeck process ($K(s, t) = \exp(-|s - t|)$, $X_t = e^{-t}W_{e^{2t}}$ – this is the only stationary Gaussian Markov process).

8.1. Concentration of measure methods

A key tool in the analysis of Gaussian processes are *tail estimates* (sometimes called *Borell estimates*). We will sketch a method how to obtain such estimates from concentration inequalities. As a preparation we need the following logarithmic Sobolev inequality in Gaussian spaces which was discovered by Gross [50]. We do not provide a proof, but refer to Ledoux [103, Thm. 5.1] (for a semigroup proof) or Shigekawa [160, Thm. 2.12] (for a probabilistic proof). Recall that the *entropy of a random variable* $Z \geq 0$ is defined as $\text{Ent}(Z) = \mathbb{E}(Z \log Z) - (\mathbb{E}Z) \log(\mathbb{E}Z)$.

Theorem 8.3 (logarithmic Sobolev inequality). *Let $G = (G_1, \dots, G_n) \in \mathbb{R}^n$ be a vector whose entries are iid standard normal random variables and $f: \mathbb{R}^n \rightarrow \mathbb{R}$*

a C^1 -function. Then the entropy of $f^2(G)$ satisfies

$$\mathbb{E} [f^2(G) \log f^2(G)] - \mathbb{E} [f^2(G)] \log \mathbb{E} [f^2(G)] \leq 2\mathbb{E} [|\nabla f(G)|^2]. \quad (8.1)$$

Theorem 8.4. Let $G = (G_1, \dots, G_n) \in \mathbb{R}^n$ be a vector whose entries are iid standard normal random variables and $f: \mathbb{R}^n \rightarrow \mathbb{R}$ a Lipschitz function with global Lipschitz constant L , i.e. $|f(x) - f(y)| \leq L|x - y|$ holds for all $x, y \in \mathbb{R}^n$. Then

$$\mathbb{E} \left[e^{\lambda(f(G) - \mathbb{E}f(G))} \right] \leq e^{\frac{1}{2}\lambda^2 L^2}, \quad \lambda > 0. \quad (8.2)$$

Proof. It is enough to show (8.2) for C^1 -functions whose derivative is bounded by L . Indeed, if f is Lipschitz with Lipschitz constant L , we can approximate f pointwise by a sequence of C^1 -functions f_n such that $|\nabla f_n(x)| \leq L$, and (8.2) remains stable under the limit if we use Fatou's lemma. So, without loss of generality, $f \in C^1$ and $|\nabla f(x)| \leq L$. Moreover, we can always assume that $\mathbb{E}f(G) = 0$, otherwise we use $f \rightsquigarrow f - \mathbb{E}f(G)$.

We apply Theorem 8.3 to the function $\lambda \mapsto e^{\lambda f(G)/2}$ to get

$$\begin{aligned} \mathbb{E} \left[\lambda f(G) e^{\lambda f(G)} \right] - \mathbb{E} \left[e^{\lambda f(G)} \right] \cdot \log \mathbb{E} \left[e^{\lambda f(G)} \right] &\leq 2\mathbb{E} \left[\left| \nabla e^{\frac{1}{2}\lambda f(G)} \right|^2 \right] \\ &= \frac{\lambda^2}{2} \mathbb{E} \left[e^{\lambda f(G)} |\nabla f(G)|^2 \right] \\ &\leq \frac{\lambda^2}{2} L^2 \mathbb{E} \left[e^{\lambda f(G)} \right]. \end{aligned}$$

If we set $\phi(\lambda) := \mathbb{E} [e^{\lambda f(G)}]$, this inequality becomes a differential inequality

$$\lambda \phi'(\lambda) - \phi(\lambda) \log \phi(\lambda) \leq \frac{\lambda^2}{2} L^2 \phi(\lambda),$$

which we can rewrite as

$$\frac{d}{d\lambda} \left(\frac{\log \phi(\lambda)}{\lambda} \right) = \frac{\phi'(\lambda)}{\lambda \phi(\lambda)} - \frac{\log \phi(\lambda)}{\lambda^2} \leq \frac{1}{2} L^2.$$

From this we see that $\phi(\lambda) \leq e^{\lambda^2 L^2/2}$, and the claim follows. \square

Corollary 8.5 (Gaussian concentration). Let $G = (G_1, \dots, G_n) \in \mathbb{R}^n$ be a vector whose entries are iid standard normal random variables and $f: \mathbb{R}^n \rightarrow \mathbb{R}$ a Lipschitz function with global Lipschitz constant L . Then

$$\mathbb{P} (\pm(f(G) - \mathbb{E}f(G)) > \lambda) \leq e^{-\lambda^2/2L^2}, \quad \lambda > 0. \quad (8.3)$$

Proof. For $x, \lambda > 0$ we find using the Markov inequality and Theorem 8.4

$$\begin{aligned} \mathbb{P} (\pm(f(G) - \mathbb{E}f(G)) > \lambda) &= \mathbb{P} (\pm x(f(G) - \mathbb{E}f(G)) > x\lambda) \\ &\leq e^{-x\lambda} \mathbb{E} \left(e^{\pm x(f(G) - \mathbb{E}f(G))} \right) \leq e^{-x\lambda + x^2 L^2/2}, \end{aligned}$$

and the claim follows, since the right hand side is minimal for $x = \lambda/L^2$. \square

An important feature in the estimates (8.2) and (8.3) is that the right hand side does not depend on the dimension n of the Gaussian vector (G_1, \dots, G_n) . This allows us to derive

Corollary 8.6 (Borell's inequality). *Let $(X_t)_{t \geq 0}$ be a separable modification¹⁸ of a centred Gaussian process and $M_t := \sup_{s \leq t} X_s$ its running maximum. If $\sigma_t^2 := \sup_{s \leq t} \mathbb{E}[X_s^2] < \infty$ for all $t \geq 0$, then $\mathbb{V}[M_t] < \infty$ and*

$$\mathbb{P}(\pm(M_t - \mathbb{E}M_t) > \lambda) \leq e^{-\lambda^2/2\sigma_t^2}, \quad \lambda > 0. \quad (8.4)$$

Proof. Fix $t > 0$ and let $0 \leq t_1 < \dots < t_n \leq t$. Since $\Xi := (X_{t_1}, \dots, X_{t_n})^\top$ is a Gaussian random vector, we can write it as $\Xi = \sqrt{C}\Gamma$ where $\sqrt{C} = (\gamma_{ij})_{i,j}$ is the (unique, positive semidefinite) square root of the covariance matrix $C = (K(t_i, t_j))_{i,j}$ and $\Gamma := (G_1, \dots, G_n)^\top$ is a Gaussian vector with iid standard normal entries. In particular,

$$\max_{1 \leq i \leq n} X_{t_i} \sim \max_{1 \leq i \leq n} (\sqrt{C}\Gamma)_i =: f(G_1, \dots, G_n).$$

For $x, y \in \mathbb{R}^n$ we have for every $i = 1, 2, \dots, n$

$$\left| (\sqrt{C}x)_i - (\sqrt{C}y)_i \right| = \left| \sum_{j=1}^n \gamma_{ij}(x_j - y_j) \right| \leq \left(\sum_{j=1}^n \gamma_{ij}^2 \right)^{1/2} |x - y|.$$

Since $\sum_{j=1}^n \gamma_{ij}^2 = \mathbb{V}(X_{t_i})$, we see that

$$|f(x) - f(y)| \leq \sigma_t |x - y|.$$

Pick the t_i in such a way that $\{t_1, \dots, t_n\} \uparrow \mathbb{Q} \cap [0, t]$. Take $y = 0$ and use the separability of the process to get $\mathbb{V}(M_t) = \sup_n \mathbb{V}f(G_1, \dots, G_n) \leq \sigma_t$. Finally, we can apply Corollary 8.5 and see that

$$\begin{aligned} \mathbb{P}(\pm(M_t - \mathbb{E}M_t) > \lambda) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\pm\left[\max_{1 \leq i \leq n} X_{t_i} - \mathbb{E}\left[\max_{1 \leq i \leq n} X_{t_i}\right]\right] > \lambda\right) \\ &\leq e^{-\lambda^2/2\sigma_t^2}. \quad \square \end{aligned}$$

Corollary 8.7 (Fernique's theorem). *Let $(X_t)_{t \geq 0}$ be a (separable modification of a) centred Gaussian process and $M_t := \sup_{s \leq t} X_s$ its running maximum. If $\sigma_t^2 := \sup_{s \leq t} \mathbb{E}[X_s^2] < \infty$ for all $t \geq 0$, then*

$$\mathbb{E}\left[e^{pM_t^2}\right] < \infty, \quad 0 \leq p < \frac{1}{2\sigma_t^2}. \quad (8.5)$$

¹⁸A necessary and sufficient condition for the existence of a separable modification is that $(s, t) \mapsto \|X_s - X_t\|_{L^2}$ is separable, e.g. if $K(s, t)$ is continuous; see Lifshits [109, p. 26] and the discussion at the beginning of Chapter 3.

Proof. From Borell's inequality we see that

$$\mathbb{P}(|M_t - \mathbb{E}M_t| > \lambda) \leq \mathbb{P}(M_t - \mathbb{E}M_t > \lambda) + \mathbb{P}(\mathbb{E}M_t - M_t > \lambda) \leq 2e^{-\lambda^2/2\sigma_t^2},$$

and with the layer-cake formula we get

$$\begin{aligned} \mathbb{E} \left[e^{p(M_t - \mathbb{E}M_t)^2} \right] &= \int_0^\infty 2\lambda p e^{p\lambda^2} \mathbb{P}(|M_t - \mathbb{E}M_t| > \lambda) d\lambda \\ &\leq \int_0^\infty 4\lambda p e^{p\lambda^2} e^{-\lambda^2/2\sigma_t^2} d\lambda = \frac{4p\sigma_t^2}{2p\sigma_t^2 - 1}. \quad \square \end{aligned}$$

The setting and proofs of Corollaries 8.6 and 8.7 show that we are in a truly exceptional situation: From $\mathbb{P}(|M_t - \mathbb{E}M_t| > \lambda) \leq 2e^{-\lambda^2/2\sigma_t^2}$ we conclude

$$\mathbb{P}(M_t < \infty) = 1 \iff \mathbb{E}M_t < \infty.$$

Together with (8.5) we get, that for a centred Gaussian process with $\sigma_t^2 := \sup_{s \leq t} \mathbb{V}X_s < \infty$, we even have

$$M_t < \infty \text{ a.s.} \iff \mathbb{E}M_t < \infty \iff \mathbb{E}e^{pM_t^2} < \infty, \quad p < (2\sigma_t)^{-1}.$$

8.2. Metric entropy methods

Let us briefly discuss a further technique in the study of maximal inequalities for Gaussian processes. It is connected with the way in which we can cover the index set optimally. These methods came up at the end of the 1960s, mainly due to Dudley [34], and led to a breakthrough in the study of Gaussian processes.

Definition 8.8. Let (\mathbb{T}, d) be a compact metric space. One defines for every $\epsilon > 0$ the

$$\begin{aligned} \text{covering number} \quad N(\epsilon) &:= \min \{n \mid \mathbb{T} \subseteq \bigcup_{k=1}^n A_k, \text{diam}(A_i) \leq \epsilon\}; \\ \text{metric entropy} \quad H(\epsilon) &:= \log N(\epsilon); \\ \text{packing number} \quad M(\epsilon) &:= \max \left\{ n \mid t_1, \dots, t_n \in \mathbb{T}, \forall i \neq j : d(t_i, t_j) > \epsilon \right\}; \\ \text{metric capacity} \quad C(\epsilon) &:= \log M(\epsilon). \end{aligned}$$

Note that $N(2\epsilon) \leq M(\epsilon) \leq N(\epsilon)$, see Lifshits [110, Prop. 10.1]. We write $N_{\mathbb{T}}, H_{\mathbb{T}}$ etc. if we need to specify the index set \mathbb{T} .

Entropy numbers play also an important role in the theory of function spaces and in the approximation of operators. Among the standard references in this direction are Carl & Stephani [20] or Edmunds & Triebel [38]. The notions introduced in Definition 8.8 have all their roots in the concept of metric entropy of a set which Kolmogorov introduced in the 1930s and 1950s, see Kolmogorov [79, 80, 81] and Kolmogorov & Tikhomirov [82].

We mention, without proof, a key result based on the notions introduced in Definition 8.8.

Theorem 8.9 (Dudley–Sudakov bounds). *Assume that $(X_t)_{t \geq 0}$ is (the separable modification of) a centred Gaussian process and $\sigma_t^2 = \sup_{s \leq t} \mathbb{E}(X_s^2) < \infty$ for every $t \geq 0$. Then the following estimates hold*

$$\frac{\epsilon}{2\sqrt{2}} \sqrt{C_{[0,t]}(\epsilon)} \leq \mathbb{E}[X_t^*] \leq 2\sqrt{2} \int_0^{\sigma_t} \sqrt{H_{[0,t]}(2\epsilon)} d\epsilon, \quad \epsilon > 0. \quad (8.6)$$

For a proof and further information we refer to the discussion in Lifshits [109, Thm. 1, p. 179, Thm. 5, p. 193] and [110, Thm. 10.1, Thm. 10.5], see also Lemma 3.3. Note that the finiteness of the *Dudley integral* appearing in the upper bound of (8.6), is *necessary and sufficient for the continuity and boundedness of the process $(X_t)_{t \in [0,T]}$ for any fixed $T > 0$* , see Lifshits [109, p. 224]. The idea behind the (upper part of the) Dudley–Sudakov bounds can be vastly generalized. We would like to mention here the pioneering work by Talagrand [170], who uses systematically majorizing measures and the generic chaining method to derive bounds for X^* , pushing the theory well beyond Gaussian processes. For example, one can get upper and lower bounds for infinitely divisible processes. The (second edition of the) research monograph Talagrand [171] contains the latest developments in this direction.

A further important topic in Gaussian processes are the so-called *small ball probabilities* or *small deviations*. The typical problem is to study the behaviour of

$$\mathbb{P} \left(\sup_{s \leq t} |X_s| \leq \epsilon \right) \quad \text{as } \epsilon \rightarrow 0, \quad (8.7)$$

or of any other interesting functional $\Phi(X_s, s \leq t)$ replacing the supremum $\sup_{s \leq t} |X_s|$. With the Wiener process in mind, one would expect asymptotics like $\sim \kappa \epsilon^a \exp(-c\epsilon^{-b})$, where b is the *small deviation rate* and c is the *small deviation constant*. For a Wiener process we get $\kappa = 4/\pi$, $a = 0$, $c = \pi^2/8$ and $b = 2$; details can be found in Lifshits [110, Chap. 11], see also Schilling [151, Lem. 12.8]. Such estimates are interesting if one wants to determine the exact short-time behaviour of the sample paths of $t \mapsto X_t$.

The following example is a classic application of small-ball probabilities; at the same time, it nicely illustrates the scope and strength of small-ball methods. The result is due to Lévy [108, p. 96, pp. 107–109], and it is the archetype of a so-called Stroock-and-Varadhan *support theorem* for diffusions (see e.g. Stroock & Varadhan [168], Ikeda & Watanabe [63, Ch. VI.8]) or for jump processes (e.g. Simon [162], Kulik [97]).

Example 8.10 (Lévy). *Let $(B_t)_{t \geq 0}$ be a one-dimensional Brownian motion. Then*

$$\mathbb{P} \left(\sup_{0 \leq t \leq 1} |B_t - f(t)| < \epsilon \right) > 0, \quad \epsilon > 0,$$

for any continuous function $f: [0, 1] \rightarrow \mathbb{R}$ with $f(0) = 0$.

Proof. Fix $\epsilon > 0$. By the uniform continuity of f on $[0, 1]$, there is some $n \in \mathbb{N}$ such that $|f(s) - f(t)| < \epsilon$ for all $|s - t| \leq 1/n$. Set $t_i := i/n$, then

$$\left\{ \sup_{0 \leq t \leq 1} |B_t - f(t)| < \epsilon \right\} \supseteq \bigcap_{i=0}^{n-1} \left\{ \sup_{t \in [t_i, t_{i+1}]} |B_t - f(t_i)| < \frac{\epsilon}{2(n-i)} \right\} =: \bigcap_{i=0}^{n-1} A_i,$$

and so, by the Markov property,

$$\begin{aligned} \mathbb{P} \left(\sup_{0 \leq t \leq 1} |B_t - f(t)| < \epsilon \right) &\geq \mathbb{P} \left(\bigcap_{i=0}^{n-1} A_i \right) \\ &\geq \mathbb{E} \left(\prod_{i=0}^{n-2} \mathbb{1}_{A_i} \cdot \mathbb{P}^{B_{t_{n-1}}} \left[\sup_{s \leq 1/n} |B_s - f(t_{n-1})| < \frac{\epsilon}{2} \right] \right). \end{aligned}$$

Since $|B_{t_{n-1}} - f(t_{n-1})| < \epsilon/4$ on A_{n-2} , it follows that

$$\mathbb{P} \left(\sup_{0 \leq t \leq 1} |B_t - f(t)| < \epsilon \right) \geq \mathbb{P} \left(\bigcap_{i=0}^{n-2} A_i \right) \mathbb{P} \left(\sup_{s \leq 1/n} |B_s| < \frac{\epsilon}{4} \right).$$

The reflection principle shows that for any $\epsilon > 0$ and $n \in \mathbb{N}$ the probability

$$\mathbb{P} \left(\sup_{s \leq 1/n} |B_s| < \frac{\epsilon}{4} \right)$$

is strictly positive, and the claim follows by iteration. \square

Index of notation

This is intended to aid cross-referencing, so notation that is specific to a single section is generally not listed. Some symbols are used locally, without ambiguity, in senses other than those given below.

Unless otherwise stated, binary operations between functions such as $f \pm g$, $f \cdot g$, $f \wedge g$, $f \vee g$, comparisons $f \leq g$, $f < g$ or limiting relations $f_n \xrightarrow[n \rightarrow \infty]{} f$, $\lim_n f_n$, $\liminf_n f_n$, $\limsup f_n$, $\sup_i f_i$ or $\inf_i f_i$ are always understood pointwise.

Generalities

positive	always ≥ 0
negative	always ≤ 0
increasing	always non-strict sense
decreasing	always non-strict sense
\mathbb{N}	natural numbers: 1, 2, 3, ...
\mathbb{N}_0	positive integers: 0, 1, 2, ...
$\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$	integer, rational, real, complex numbers
$\inf \emptyset, \sup \emptyset$	$\inf \emptyset = +\infty, \sup \emptyset = -\infty$
$a \vee b$	maximum of a and b
$a \wedge b$	minimum of a and b
$[x]$	$\max\{n \in \mathbb{Z} \mid n \leq x\}$

Functions and sets

$\mathbb{1}_A$	$\mathbb{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$
f^+, X^+	positive part
f^-, X^-	negative part
$f \asymp g$	$\forall x : cf(x) \leq g(x) \leq Cf(x)$
$\log^+ x$	$\max\{\log(x), 0\}$
\hat{f}, \check{g}	(inverse) Fourier transform, p. 385
$f^*(x)$	Hardy-Littlewood maximal function, p. 417
$f^\diamond(x)$	square maximal function, p. 417
$C(\epsilon)$	metric capacity, p. 470
$H(\epsilon)$	metric entropy, p. 470
$N(\epsilon)$	covering number, p. 408, 470
$M(\epsilon)$	packing number, p. 470
ψ, ψ_X	characteristic exponent, p. 385, 435
ψ^*, ψ_X^*	maximum function, p. 386
$ \psi ^*, \psi_X ^*$	maximum function, p. 386
(b, Q, ν)	Lévy triplet, p. 385, 435
$(b(x), Q(x), \nu(x, dy))$	infinitesimal

characteristics, p. 450

$\ \cdot\ _{\text{BMO}}$	bounded mean oscillation, p. 412 (3.11), (3.12), p. 419
$B_r(x)$	open ball with radius r and centre x

Random variables and processes

$\mathbb{P}, \mathbb{E}, \mathbb{V}$	probability, expectation, variance
$X \sim Y$	$\text{law}(X) = \text{law}(Y)$
$X \sim \mu$	$\text{law}(X) = \mu$
M_t	$\sup_{s \leq t} X_s$
X_t^*	$\sup_{s \leq t} X_s $
X^*	$\sup_{s \geq 0} X_s $
ΔX_t	$X_t - X_{t-}, (t > 0)$
$c \bullet X_n$	martingale transform, p. 421
$H \bullet X_t$	stochastic integral, p. 441
$[X]_n$	compensator, p. 423
$\langle X \rangle_t, [X]_t$	(previsible) quadratic variation, p. 424, 430
$\alpha_X(\ell)$	Milman concentration function, p. 396
$D(\mu, \ell), G(\mu, \ell), K(\mu, \ell), M(b, \nu, \ell)$	truncated moments, p. 386
$Q_X(\ell)$	Lévy concentration function, p. 392

Abbreviations

a.a., a.e.	almost all/every(where)
BDG	Burkholder–Davis–Gundy
BMO	bounded mean oscillation
càdlàg	right-continuous, finite left-hand limits
iid	independent and identically distributed
(L1), (L2), (L3)	properties of Lévy processes, pp. 433–434

Index

- BMO, 411, 412, 419
- Brownian motion, 427, 433, 436, 467, 471
 - time-change, 426
- censored moment, 386, 388, 390, 439
- chaining, 406
- characteristic exponent, 385, 435, 450
- concentration function
 - Lévy, 392, 440
 - Milman, 396
 - vs. censored moments, 393, 396
 - vs. characteristic function, 393
- conditionally controlled increments, 408
- covering number, 406, 470
- Dirichlet form, 452
- doubling condition, 401
- Dynkin–Kinney criterium, 466
- Feller process, 451
- Fourier transform, 385
- G -Lévy process, 466
- Gaussian process, 467
- generalized moment, 383, 445
- inequality
 - Azuma, 420
 - BDG (continuous case), 424
 - BDG (jump case), 429–431
 - BDG, counterexample, 431
 - Birnbaum–Marshall, 415
 - Borell, 469
 - controlled jumps, 408, 409
 - Doob’s L^p , 414
 - Doob’s maximal, 414
 - Dudley–Sudakov, 471
 - Etemadi, 438, 464
 - Gaussian concentration, 468
 - good- λ , 400, 401
 - John–Nirenberg, 411
 - Khintchine, 421
 - log Sobolev, 467
 - maximal tail, 439, 441, 453, 458, 460, 461
 - McKean, 427
 - mean exit time, 442, 443, 455
 - Ottaviani–Skorokhod, 438, 465
- infinitely divisible, 385
- infinitesimal characteristics, 450
- interpolation, 402, 419
- layer cake formula, 398
- Lévy process, 433, 451
 - characterization, 435
 - strong Markov property, 434
- Lévy triplet, 386, 435, 450
- Lévy–Khintchine formula, 385, 435, 450
- Lévy-driven SDE, 441, 451, 457
- Lévy-type process, 450
- $L \log L$ class, 417
- martingale, 413
- maximal function
 - dyadic, 417
 - Hardy–Littlewood, 417
 - sharp ($\#$), 419
- maximum function ψ^* , $|\psi|^*$, 386, 439
 - estimates, 388, 391
 - vs. censored moments, 388, 389
- metric entropy, 470
- moderate growth, 432
- moment, *see* censored/generalized m.
- negative definite function, 386
 - upper bound, 388
- pseudo Poisson process, 451
- reflection principle, 436, 437
- sector condition, 391
- separable process, 404
 - sufficient condition, 404
- square bracket, 423, 430
- stable Lévy process, 461
- stable-like process, 452, 457
- strong law of large numbers, 416
- sublinear process, 466
- submultiplicative function, 444, 445
- symbol of Lévy-type process, 450, 452
- theorem
 - Fernique, 469
 - John–Nirenberg, 411

Kolmogorov–Chentsov, 405
 Pisier, 406
 truncated moment, *see* censored m.
 uniform stochastic continuity, 463
 usual conditions, 385

Acknowledgments

We thank the editor-in-chief of *Probability Surveys* for the invitation to contribute to this journal and for the careful handling of the submission. The efforts and advice of two anonymous reviewers are much appreciated. We are deeply indebted to our colleagues and students, Dr. David Berger, Dr. Wojciech Cychan, Mr. Mustafa Hamadi and Ms. Cailing Li, for proofreading and helpful comments. A substantial part of this paper was written while the first-named author was research associate at the Mathematics Department of TU Dresden. She would like to thank the department for the good working conditions.

References

- [1] AMIR, D. and MILMAN, V.D. (1980). Unconditional and symmetric sets in n -dimensional normed spaces. *Israel J. Math.* **37** 3–20. [MR0599298](#)
- [2] AZUMA, K. (1967). Weighted sums of certain dependent random variables. *Tohoku Math. J.* **19** 357–367. [MR0221571](#)
- [3] BARLOW, M.T. and YOR, M. (1982). Semimartingale inequalities via the Garsia–Rodemich–Rumsey lemma, and applications to local times. *J. Functional Analysis* **49** 198–229. [MR0680660](#)
- [4] BASS, R.F. (1988) Uniqueness in law for pure jump Markov processes. *Probab. Theory Related Fields* **79** 271–287. [MR0958291](#)
- [5] BASS, R.F. (1995). *Probabilistic Techniques in Analysis*. Springer, New York. [MR1329542](#)
- [6] BASS, R.F. and LEVIN, D. (2002) Harnack inequalities for jump processes. *Potential Anal.* **17** 375–388. [MR1918242](#)
- [7] BENNETT, C. and SHARPLEY, R. (1988). *Interpolation of Operators*. Academic Press, Boston, MA. [MR0928802](#)
- [8] BERG, C. and FORST, G. (1973). Non-symmetric translation invariant Dirichlet forms. *Invent. Math.* **21** 199–212. [MR0427662](#)
- [9] BERG, C. and FORST, G. (1975). *Potential Theory on Locally Compact Abelian Groups*. Springer, Berlin. [MR0481057](#)
- [10] BERGER, D., KÜHN, F. and SCHILLING, R.L. (2022). Lévy processes, generalized moments and uniform integrability. *Probab. Math. Statist.* **42** 109–131. [MR4490672](#)
- [11] BILLINGSLEY, P. (1995). *Probability and Measure* 3rd ed. Wiley, New York. [MR1324786](#)
- [12] BIRNBAUM, Z.W. and MARSHALL, A.W. (1961). Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *Ann. Math. Statist.* **32** 687–703. [MR0148106](#)
- [13] BLUMENTHAL, R.M. (1957). An extended Markov property. *Trans. Amer. Math. Soc.* **85** 52–72. [MR0088102](#)

- [14] BÖTTCHER, B., SCHILLING, R.L. and WANG, J. (2014). *Lévy Matters III. Lévy-type processes: construction, approximation and sample path properties*. Springer Lecture Notes in Mathematics, **2099**, Springer, New York. [MR3156646](#)
- [15] BOUCHERON, S., LUGOS, G. and MASSART, P. (2013). *Concentration Inequalities. A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford. [MR3185193](#)
- [16] BURKHOLDER, D.L. (1973). Distribution function inequalities for martingales. *Ann. Probability* **1** 19–42. [MR0365692](#)
- [17] BURKHOLDER, D.L. and GUNDY, R.F. (1972). Extrapolation and interpolation of quasi-linear operators on martingales. *Acta Math.* **124** 249–304. [MR0440695](#)
- [18] BURKHOLDER, D.L. and GUNDY, R.F. (1972). Distribution function inequalities for the area integral. *Studia Math.* **44** 527–544. [MR0340557](#)
- [19] BURKHOLDER, D.L., DAVIS, B.J. and GUNDY, R.F. (1972). Integral inequalities for convex functions of operators on martingales. In: LECAM, L. et al. eds. (1972). *Proceedings 6th Berkeley Symposium on Mathematical Statistics and Probability* **2** 223–240. University of California Press, Berkeley, CA. [MR0400380](#)
- [20] CARL, B. and STEPHANI, I. (1990). *Entropy, compactness and the approximation of operators*. Cambridge University Press, Cambridge. [MR1098497](#)
- [21] CHERSI, F. (1970). Martingales et intégrabilité de $X \log^+ X$ d’après R.F. Gundy. *Séminaire de Probabilités IV (1968/69)*. Lecture Notes in Math. **124** 37–45. Springer, Berlin.
- [22] CHOW, Y.S. (1960). A martingale inequality and the law of large numbers. *Proc. Amer. Math. Soc.* **11** 107–111. [MR0112190](#)
- [23] COHEN, S.N. and ELLIOTT, R.J. (2015). *Stochastic Calculus and Applications* (2nd ed). Birkhäuser/Springer, New York. [MR3443368](#)
- [24] CSÖRGÖ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York. [MR0666546](#)
- [25] DELLACHERIE, C. (1979). Inégalités de convexité pour les processus croissants et les sousmartingales. *Lecture Notes in Math.* **721**, *Séminaire de Probabilités XIII* 371–384, Springer, Berlin. [MR0544807](#)
- [26] DELLACHERIE, C. and MEYER, P.A. (1975). *Probabilités et potentiel. Chapitres I à IV*. Hermann, Paris. [MR0488194](#)
- [27] DELLACHERIE, C. and MEYER, P.A. (1980). *Probabilités et potentiel. Chapitres V à VIII*. Hermann, Paris. [MR0566768](#)
- [28] DELLACHERIE, C., MAISONNEUVE, B. and MEYER, P.A. (1992). *Probabilités et potentiel. Chapitres XVII à XXIV*. Hermann, Paris.
- [29] DEMBO, A. and ZEITOUNI, O. (1992). *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston. [MR1202429](#)
- [30] DENG, C.-S. and SCHILLING, R.L. On shift Harnack inequalities for subordinate semigroups and moment estimates for Lévy processes. *Stoch. Proc. Appl.* **125** 3851–3878. [MR3373306](#)
- [31] DENK, R., KUPPER, M. and NENDEL, M. (2020). A semigroup ap-

- proach to nonlinear Lévy processes. *Stoch. Proc. Appl.* **130** 1616–1642. [MR4058284](#)
- [32] DIRKSEN, S. (2014). Itô isomorphisms for L_p -valued Poisson stochastic integrals. *Ann. Probab.* **42** 2595–2643. [MR3265175](#)
- [33] DOOB, J.L. (1953). *Stochastic Processes*. Wiley, New York. [MR0058896](#)
- [34] DUDLEY, R.M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis* **1** 290–330. [MR0220340](#)
- [35] DUBINS, L.E. (1968). On a theorem of Skorohod. *Ann. Math. Statist.* **39** 2094–2097. [MR0234520](#)
- [36] DYNKIN, E.B. (1961). *Die Grundlagen der Theorie der Markoffschen Prozesse*. Springer, Berlin. [MR0131899](#)
English translation (2006): *Theory of Markov Processes*. Dover, Mineola (NY). [MR2305744](#)
- [37] DZHAPARIDZE, K. and VALKEILA, E. (1990). On the Hellinger type distances for filtered experiments. *Probab. Theory Rel. Fields* **85** 105–117. [MR1044303](#)
- [38] EDMUNDS, D. and TRIEBEL, H. (1996). *Function Spaces, Entropy Numbers and Differential Operators*. Cambridge University Press, Cambridge. [MR1410258](#).
- [39] ESSEEN, C.G. (1968). On the concentration function of a sum of independent random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **9** 290–308. [MR0231419](#)
- [40] ETEMADI, N. (1985). On some classical results in probability. *Sankhya, Ser. A* **47** 215–221. [MR0844022](#)
- [41] ETHIER, S.N. and KURTZ, T.G. (1985). *Markov Processes: Characterization and Convergence*. Wiley, New York. [MR838085](#)
- [42] FEFFERMAN, C. and STEIN, E.M. (1972). H^p spaces of several variables. *Acta Math.* **129** 137–193. [MR0447953](#)
- [43] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications. Volume II* (2nd ed). Wiley, New York. [MR0270403](#)
- [44] GEISS, S. (1997). BMO_ψ -spaces and applications to extrapolation theory. *Studia Math.* **122** 235–274. [MR1434474](#)
- [45] GEISS, S. (2005). Weighted BMO and discrete time hedging within the Black-Scholes model. *Probab. Theory Related Fields* **132** 13–38. [MR2136865](#)
- [46] GETTOOR, R.K. and SHARPE, M.J. (1972) Conformal martingales. *Invent. Math.* **16** 271–308. [MR0305473](#)
- [47] GIHMANN, I.I. and SKOROHOD, A.V. (1974). *The Theory of Stochastic Processes I*. Springer, New York. [MR0346882](#)
- [48] GINÉ, E. and MARCUS, M.B. (1983). The central limit theorem for stochastic integrals with respect to Lévy processes. *Ann. Probab.* **11** 58–77. [MR0682801](#)
- [49] GRAFAKOS, L. (2009). *Modern Fourier Analysis* (2nd ed). Springer, New York. [MR2463316](#)
- [50] GROSS, L. (1975). Logarithmic Sobolev inequalities. *Amer. J. Math.* **97**

- 1061–1083. [MR0420249](#)
- [51] GRZYWNY, T. (2014) On Harnack inequality and Hölder regularity for isotropic unimodal Lévy processes. *Potential Anal.* **41** 1–29. [MR3225805](#)
- [52] GRZYWNY, T. and SZCZYPKOWSKI, K. (2020) Lévy processes: concentration function and heat kernel bounds. *Bernoulli* **26** 3191–3223. [MR4140542](#)
- [53] GUNDY, R.F. (1969). On the class $L \log L$, martingales, and singular integrals. *Studia Math.* **33** 107–118. [MR0243588](#)
- [54] HAAGERUP, U. (1982). The best constants in the Khintchine inequality. *Studia Math.* **70** 231–283. [MR654838](#)
- [55] HENGARTNER, W. and THEODORESCU, R. (1973). *Concentration Functions*. Academic Press, New York. [MR0331448](#)
- [56] HITCZENKO, P. (1990). Upper bounds for the L_p -norms of martingales. *Probab. Theory Related Fields* **86** 225–238. [MR1065280](#)
- [57] HÖEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. [MR0144363](#)
- [58] HOH, W. (1998). *Pseudo-Differential Operators Generating Markov Processes*. Habilitationsschrift, University of Bielefeld.
- [59] HOH, W. (2000). Pseudo differential operators with negative definite symbols of variable order. *Rev. Mat. Iberoamericana* **16** 219–241. [MR1809340](#)
- [60] HOLLENDER, J. (2016). *Lévy-Type Processes under Uncertainty and Related Nonlocal Equations*. Ph.D. Thesis, TU Dresden. CreateSpace Independent Publishing Platform. ISBN: 978-1-5355-5384-1. <https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa-211795>
- [61] HULANICKI, A. (1980). A class of convolution semi-groups of measures on a Lie group. In: WERON, A. (ed.): *Probability Theory on Vector Spaces II. Lecture Notes in Math.* **828** 82–101, Springer, Berlin. [MR0611711](#)
- [62] HYTÖNEN, T., VAN NEERVEN, J., VERAAR, M. and WEIS, L. (2016). *Analysis in Banach Spaces. Volume 1: Martingales and Littlewood–Paley Theory*. Springer, Cham. [MR3617205](#)
- [63] IKEDA, N. and WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes* 2nd ed. North-Holland Publishing Company, Amsterdam and Kodansha, Tokyo. [MR1011252](#)
- [64] ITÔ, K. (1961, 1984). *Lectures on Stochastic Processes*. Tata Institute of Fundamental Research, Bombay. Reprinted (1984), Springer, Berlin. [MR0759892](#)
- [65] ITÔ, K. (1969, 2004). *Stochastic Processes. Lectures given at Aarhus University*. Lecture Notes series no. **16**, Matematisk Institut, Aarhus Universitet, Aarhus. Reprinted (2004), Springer, Berlin. [MR2053326](#)
- [66] JACOB, N. (2001–2005). *Pseudo-Differential Operators and Markov Processes I–III*. Imperial College Press/World Scientific, London. [MR1873235](#), [MR1917230](#), [MR2158336](#)
- [67] JACOB, N. and LEOPOLD, H.-G. (1993). Pseudo-differential operators with variable order of differentiation generating Feller semigroups. *Integral Equations Operator Theory* **17** 544–553. [MR1243995](#)
- [68] JACOD, J. and SHIRYAEV, A.N. (2003). *Limit Theorems for Stochastic*

- Processes* (second edition). Springer, Berlin. [MR1943877](#)
- [69] JOULIN, A. (2007). On maximal inequalities for stable stochastic integrals. *Potential Anal.* **26** 57–78. [MR2276525](#)
- [70] KALLENBERG, O. and SZTENCCEL, R. (1991). Some dimension-free features of vector-valued martingales. *Probab. Theory Related Fields* **88** 215–247. [MR1096481](#)
- [71] KAMIŃSKI, Ł. and OSEKOWSKI, A. (2023). On the best constant in the L^p estimate for the sharp maximal function. *J. Math. Anal. Appl.* **518** 126733, 18 pp.
- [72] KHINTCHINE, A. (1923). Über dyadische Brüche. *Math. Z.* **18** 109–116. [MR1544623](#)
- [73] KNOPOVA, V. (2014). Compound kernel estimates for the transition probability density of a Lévy process. *Theory Probab. Math. Statist.* **89** 57–70. [MR3235175](#)
- [74] KNOPOVA, V. and KULIK, A. (2013). Intrinsic small time estimates for distribution densities of Lévy-type processes. *Random Oper. Stoch. Equ.* **21** 321–344. [MR3139314](#)
- [75] KNOPOVA, V. and KULIK, A. (2017). Intrinsic compound kernel estimates for the transition probability density of Lévy-type processes and their applications. *Probab. Math. Statist.* **37** 53–100. [MR3652202](#)
- [76] KNOPOVA, V., KULIK, A. and SCHILLING, R.L. (2021). Construction and heat kernel estimates of general stable-like Markov processes. *Dissertationes Math.* **569** 86 pp. [MR4361582](#)
- [77] KNOPOVA, V. and SCHILLING, R.L. (2014). On the small-time behaviour of Lévy-type processes. *Stochastic Process. Appl.* **124** 2249–2265. [MR3188355](#)
- [78] KNOPOVA, V., SCHILLING, R.L. and WANG, J. (2015). Lower bounds of the Hausdorff dimension for the images of Feller processes. *Statist. Probab. Lett.* **97** 222–228. [MR3299774](#)
- [79] KOLMOGOROV, A.N. (1936) Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse *Ann. of Math.* **37** 107–110. [MR1503273](#)
English Translation in: TIKHOMIROV, V.M. (ed.) (1991). *Selected Works of A.N. Kolmogorov. Volume 1: Mathematics and Mechanics*. Kluwer, Dordrecht, pp. 202–205. [MR1175399](#)
- [80] KOLMOGOROV, A.N. (1955) Estimates of the minimal number of elements of ϵ -nets in various function classes and their application to the problem of representing functions of several variables by superpositions of a smaller number of variables (Russian). *Uspekhi Mat. Nauk* **10.1** 192–194.
- [81] KOLMOGOROV, A.N. (1956) On the representation of continuous functions of several variables as superpositions of continuous functions of a smaller number of variables (Russian). *Dokl. Acad. Nauk SSSR* **108** 179–182. [MR0080129](#)
English Translation in: TIKHOMIROV, V.M. (ed.) (1991). *Selected Works of A.N. Kolmogorov. Volume 1: Mathematics and Mechanics*. Kluwer, Dordrecht, pp. 378–382. [MR1175399](#)
- [82] KOLMOGOROV, A.N. and TIKHOMIROV, (1959) ϵ -entropy and ϵ -capacity

- of sets in function spaces (Russian). *Uspekhi Mat. Nauk* **14.2** 8–86. [MR0112032](#)
 English Translation in: *Amer. Math. Soc. Transl. (2)* (1961) **17** 277–364. [MR0124720](#)
 Reprinted in: SHIRYAYEV, A.N. (ed.) (1993). *Selected Works of A.N. Kolmogorov. Volume 3: Information Theory and the Theory of Algorithms*. Kluwer, Dordrecht, pp. 86–170. [MR1228446](#)
- [83] KOLOKOLTSOV, V.N. (2011). Markov processes, semigroups and generators. De Gruyter, Berlin. [MR2780345](#)
- [84] KRUGLOV, V.M. (2017). A note on infinitely divisible distributions. *Theory Probab. Appl.* **15** 319–324. [MR0272018](#)
- [85] KRÜHNER, P. and SCHNURR, A. (2018). Time change equations for Lévy-type processes. *Stoch. Proc. Appl.* **128** 963–978. [MR3758344](#)
- [86] KRYLOV, N.V. (1995). *Introduction to the Theory of Diffusion Processes*. American Mathematical Society, Providence, RI. [MR1311478](#)
- [87] KÜHN, F. (2017). *Lévy Matters VI. Lévy-Type Processes: Moments, Construction and Heat Kernel Estimates*. Springer Lecture Notes in Mathematics, **2187**, Springer, New York. [MR3701414](#)
- [88] KÜHN, F. (2017). Existence and estimates of moments for Lévy-type processes. *Stoch. Proc. Appl.* **127** 1018–1041. [MR3605719](#)
- [89] KÜHN, F. (2018). Solutions of Lévy-driven SDEs with unbounded coefficients as Feller processes. *Proc. Amer. Math. Soc.* **146** 3591–3604. [MR3803683](#)
- [90] KÜHN, F. (2019). Perpetual integrals via random time changes. *Bernoulli* **25** 1755–1769. [MR3961229](#)
- [91] KÜHN, F. (2019). Viscosity solutions to Hamilton-Jacobi-Bellman equations associated with sublinear Lévy(-type) processes. *ALEA* **16** 531–559. [MR3941868](#)
- [92] KÜHN, F. (2020). Existence of (Markovian) solutions to martingale problems associated with Lévy-type operators. *Electron. J. Probab.* **25** 1–26. [MR4073677](#)
- [93] KÜHN, F. (2021). On infinitesimal generators of sublinear Markov semigroups. *Osaka J. Math.* **58** 487–508. [MR4350042](#)
- [94] KÜHN, F. (2022). Upper functions for sample paths of Lévy(-type) processes. *Bernoulli* **28** 2874–2908.
- [95] KÜHN, F. and KUNZE, M. (2021). Feller generators with measurable lower order terms. *Positivity* **26**, 85–41 pp. [MR4474566](#)
- [96] KÜHN, F. and SCHILLING, R.L. (2019). On the domain of fractional Laplacians and related generators of Feller processes. *J. Funct. Anal.* **276** 2397–2439. [MR3926121](#)
- [97] KULIK, A. (2023). Support theorem for Lévy driven SDEs. To appear in *J. Theor. Probab.* Preprint, arXiv: arxiv.org/abs/2112.11106
- [98] KUNITA, H. (2004) Stochastic Differential Equations Based on Lévy Processes and Stochastic Flows of Diffeomorphisms. In: RAO, M.M. (2004). *Real and Stochastic Analysis*. Trends in Mathematics. Birkhäuser, Boston, pp. 305–373. [MR2090755](#)

- [99] KWAPIEŃ, S. and WOYCZYŃSKI, W. (1992) *Random Series and Stochastic Integrals: Single and Multiple*. Birkhäuser, Boston. [MR1167198](#)
- [100] LE CAM, L. (1965). On the distribution of sums of independent random variables. In: NEYMAN, J. and LE CAM, L. eds. (1965). *Bernoulli 1713, Bayes 1763, Laplace 1813. Anniversary Volume* pp. 179–202. Springer, Berlin. [MR0199871](#)
- [101] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. [MR0856411](#)
- [102] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces. Isoperimetry and Processes*. Springer, Berlin. [MR1102015](#)
- [103] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon*. American Mathematical Society, Providence, RI. [MR1849347](#)
- [104] LENGART, E., LEPINGLE, D. and PRATELLI, M. (1980). Presentation unifiée de certaines inégalités de la théorie des martingales. *Lecture Notes in Math.* **784**, *Séminaire de Probabilités XIV* 26–52, Springer, Berlin. [MR0580107](#)
- [105] LÉVY, P. (1922). *Leçons d'Analyse Fonctionnelle*. Gauthier–Villars, Paris.
- [106] LÉVY, P. (1937, 1954). *Théorie de l'Addition des Variables Aléatoires*, 1st ed. 1937, 2nd revised ed. 1954. Gauthier–Villars, Paris.
- [107] LÉVY, P. (1951). *Problèmes Concrets d'Analyse Fonctionnelle* (2nd ed. of [105]). Gauthier–Villars, Paris. [MR0041346](#)
- [108] LÉVY, P. (1953). La mesure de Hausdorff de la courbe du mouvement brownien. *Giorn. Ist. Ital. Attuari* **16** 1–37. [MR0064344](#)
- [109] LIFSHITS, M.A. (1995). *Gaussian Random Functions*. Kluwer Academic Publishers, Dordrecht. [MR1472736](#)
- [110] LIFSHITS, M.A. (2012). *Lectures on Gaussian Processes*. Springer, Heidelberg. [MR3024389](#)
- [111] LIN, Z. and BAI, Z. (2010). *Probability Inequalities*. Science Press, Beijing. [MR2789096](#)
- [112] LINDE, W. (1986). *Probability in Banach Spaces – Stable and Infinitely Divisible Distributions*. Wiley, Chichester. [MR0874529](#)
- [113] LIPTSER, R.SH. and SHIRYAYEV, A.N. (1989). *Theory of Martingales*. Kluwer Academic Publishers, Dordrecht. [MR1022664](#)
- [114] LITTLEWOOD, J.E. (1930). On bounded bilinear forms in an infinite number of variables. *Quarterly J. Math.* **1** 164–174.
- [115] LIU, X. and XI, G. (2020). On a maximal inequality and its application to SDEs with singular drift. *Stochastic Process. Appl.* **130** 4275–4293. [MR4102266](#)
- [116] LOÉVE, M. (1977). *Probability Theory I, II* 4th ed. Springer, New York. [MR0651017](#), [MR0651018](#)
- [117] MANSTAVIČIUS, M. (2004). p -variation of strong Markov processes. *Ann. Probab.* **32** 2053–2066. [MR2073185](#)
- [118] (2017). MANSTAVIČIUS, M. and SCHNURR, A. Criteria for the finiteness of the strong p -variation for Lévy-type processes. *Stoch. Anal. Appl.* **35** 873–899. [MR3686473](#)
- [119] MARCINKIEWICZ, J. and ZYGMUND, A. (1938). Quelques théorèmes sur

- les fonctions indépendantes. *Studia Mathematica* **7** 104–120.
- [120] MARCUS, M.B. and ROSEN, J. (2006). *Markov Processes, Gaussian Processes, and Local Times*. Cambridge University Press, Cambridge. [MR2250510](#)
- [121] MARINELLI, C. and RÖCKNER, M. (2016). On the maximal inequalities of Burkholder, Davis and Gundy. *Expo. Math.* **34** 1–26. [MR3463679](#)
- [122] MARTIN, G. (2015). Stackexchange (last accessed 8/January/2023). <https://math.stackexchange.com/questions/1175189/prove-the-bound-fx-leq-c-fx-for-a-submultiplicative-functionf>
- [123] MCDIARMID, C. (1998) Concentration. In: HABIB, M. et al. eds. (1998). *Probabilistic Methods for Algorithmic Discrete Mathematics* 195–248. Springer, Berlin. [MR1678578](#)
- [124] MEERSCHAERT, M.M., SCHILLING, R.L. and SIKORSKII, A. (2015). Stochastic solutions for fractional wave equations. *Nonlinear Dyn.* **80** 1685–1695. [MR3343424](#)
- [125] MEYER, P.A. (1972). Martingales and stochastic integrals. I. *Lecture Notes in Math.* **284**, Springer, Berlin. [MR0426145](#)
- [126] MEYER, P.A. (1976). Un cours sur les intégrales stochastiques. *Lecture Notes in Math.* **511**, *Séminaire de Probabilités X* 245–400, Springer, Berlin. [MR0501332](#)
- [127] MEYER, P.A. (1981). Flot d’une équation différentielle stochastique (d’après Malliavin, Bismut, Kunita). *Lecture Notes in Math.* **850**, *Séminaire de Probabilités XV* 103–117, Springer, Berlin. [MR0622556](#)
- [128] MIL’MAN, V.D. (1971). New proof of the theorem of A. Dvoretzky on intersections of convex bodies. *Functional Anal. Appl.* **5** 288–295. [MR0293374](#)
- [129] MONROE, I. (1972). On embedding right continuous martingales in Brownian motion. *Ann. Math. Statist.* **43** 1293–1311. [MR0343354](#)
- [130] MONROE, I. (1978). Processes that can be embedded in Brownian motion. *Ann. Probability* **6** 42–56. [MR0455113](#)
- [131] NOVIKOV, A.A. (1971). The moment inequalities for stochastic integrals. *Theory Probab. Appl.* **16** 538–541. [MR0288844](#)
- [132] NOVIKOV, A.A. (1975). Discontinuous martingales. *Theory Probab. Appl.* **20** 11–26. [MR0394861](#)
- [133] NOVIKOV, A.A. and SHIRYAEV, A. (2013). Remarks on moment inequalities and identities for martingales. *Statist. Probab. Lett.* **83** 1260–1261. [MR3041402](#)
- [134] OSEKOWSKI, A. (2012). *Sharp Martingale and Semimartingale Inequalities*. Birkhäuser, Basel. [MR2964297](#)
- [135] PALEY, R.E.A.C. and ZYGMUND, A. (1930). On some series of functions, (1). *Proc. Cambridge Phil. Soc.* **26** 337–357.
- [136] PETROV, V.V. (1975). *Sums of Independent Random Variables*. Springer, Berlin. [MR0388499](#)
- [137] PETROV, V.V. (1995). *Limit Theorems of Probability Theory. Sequences of Independent Random Variables*. Oxford University Press, Oxford. [MR1353441](#)

- [138] PISIER, G. (1979/80). Conditions d'entropie assurant la continuité de certains processus et applications à l'analyse harmonique. *Seminar on Functional Analysis, 1979–1980* Exposé No. **13–14** 43 pp., École Polytechnique, Palaiseau. [MR0604395](#)
- [139] PROTTER, PH. (2004). *Stochastic Integration and Differential Equations*, 2nd ed. Springer, Berlin. [MR2020294](#)
- [140] PRUITT, W. (1981). The growth of random walks and Lévy processes. *Ann. Probab.* **9** 948–956. [MR0632968](#)
- [141] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. Springer, Berlin. [MR1725357](#)
- [142] ROGERS, L.C.G. and WILLIAMS, D. (1994). *Diffusions, Markov Processes, and Martingales. Volume One: Foundations*, 2nd ed. Wiley, Chichester. [MR1331599](#)
- [143] SATO, K. (1999, 2013). *Lévy Processes and Infinitely Divisible Distributions*, revised ed. Cambridge University Press, Cambridge. [MR3185174](#), [MR1739520](#)
- [144] SCHILLING, R.L. (1998). Growth and Hölder conditions for the sample paths of Feller processes. *Probab. Theory Related Fields* **112** 565–611. [MR1664705](#)
- [145] SCHILLING, R.L. (1998). Feller processes generated by pseudo-differential operators: on the Hausdorff dimension of their sample paths. *J. Theoret. Probab.* **11** 303–330. [MR1622573](#)
- [146] SCHILLING, R.L. (1998). Conservativeness and extensions of Feller semigroups. *Positivity* **2** 239–256. [MR1653474](#)
- [147] SCHILLING, R.L. (2000). Sobolev embedding for stochastic processes. *Expo. Math.* **18** 239–242. [MR1763890](#)
- [148] SCHILLING, R.L. (2016) An Introduction to Lévy and Feller Processes. In: KHOSHNEVISAN, D. and SCHILLING, R.L. (2016). *From Lévy-Type Processes to Parabolic SPDEs*. Birkhäuser, Cham. [MR3587832](#)
- [149] SCHILLING, R.L. (2017). *Measures, Integrals and Martingales*, 2nd ed. Cambridge University Press, Cambridge. [MR3644418](#)
- [150] SCHILLING, R.L. (2018). *Martingale und Prozesse*. De Gruyter, Berlin. [MR3793613](#)
- [151] SCHILLING, R.L. (2021). *Brownian Motion. A Guide to Random Processes and Stochastic Calculus*, 3rd ed. De Gruyter, Berlin. [MR3234570](#)
- [152] SCHILLING, R.L. (2021). *Measure, Integral, Probability & Processes*. Independently Published, Dresden. ISBN: 979-8-5991-0488-9.
- [153] SCHILLING, R.L. and SCHNURR, A. (2010). The Symbol Associated with the Solution of a Stochastic Differential Equation. *Electron. J. Probab.* **15** 1369–1393. [MR2721050](#)
- [154] SCHILLING, R.L. and UEMURA, T. (2007). On the Feller property of Dirichlet forms generated by pseudo differential operators. *Tohoku Math. J.* **59** 401–422. [MR2365348](#)
- [155] SCHILLING, R.L. and WANG, J. (2012). Strong Feller continuity of Feller processes and semigroups. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **15** 1250010, 28 pp. [MR2957137](#)

- [156] SCHILLING, R.L. and WANG, J. (2015). Lower bounded semi-Dirichlet forms associated with Lévy type operators. In: CHEN, Z.-Q. ET AL. (eds.) *Festschrift Masatoshi Fukushima*, pp. 507–526. World Scientific, Hackensack, NJ. [MR3379356](#)
- [157] SCHNURR, A. (2009) *The Symbol of a Markov Semimartingale*. Ph.D. thesis, TU Dresden. Shaker-Verlag, Aachen. ISBN: 978-3-8322-8244-8. <https://nbn-resolving.org/urn:nbn:de:bsz:14-ds-1244626491491-70401>
- [158] SCHNURR, A. (2013). Generalization of the Blumenthal–Gettoor index to the class of homogeneous diffusions with jumps and some applications. *Bernoulli* **19** 2010–2032. [MR3129042](#)
- [159] SHARPE, M. (1988). *General Theory of Markov Processes*. Academic Press, Boston. [MR0958914](#)
- [160] SHIGEKAWA, I. (2004). *Stochastic Analysis*. American Mathematical Society, Providence RI. [MR2060917](#)
- [161] SIEBERT, E. (1982) Continuous convolution semigroups integrating a sub-multiplicative function. *manuscripta math.* **37** 383–391. [MR0657526](#)
- [162] SIMON, T. (2000) Support theorem for jump processes. *Stochastic Process. Appl.* **89** 1–30. [MR1775224](#)
- [163] SKOROKHOD, A.V. (1991). *Random Processes with Independent Increments*. Kluwer, Dordrecht. [MR1155400](#)
- [164] SONG, R. and VONDRAČEK, Z. (2004) Harnack inequality for some classes of Markov processes. *Math. Z.* **246** 177–202. [MR2031452](#)
- [165] STEIN, E.M. (1969). Note on the class $L \log L$. *Studia Math.* **32** 305–310. [MR0247534](#)
- [166] STEIN, E.M. (1970). *Singular Integrals and Differentiability Properties of Functions*. Princeton University Press, Princeton, NJ. [MR0290095](#)
- [167] STEIN, E.M. (1970). *Topics in Harmonic Analysis Related to the Littlewood–Paley Theory*. Annals of Mathematics Studies **63**, Princeton University Press, Princeton, NJ. [MR0252961](#)
- [168] STROOCK, D.W. and VARADHAN, S.R.S. (1972). On the support of diffusion processes with applications to the strong maximum principle. In: LECAM, L. et al. eds. (1972). *Proceedings 6th Berkeley Symposium on Mathematical Statistics and Probability* **3** 333–359. University of California Press, Berkeley, CA. [MR0400425](#)
- [169] TALAGRAND, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.* **81** 73–205. [MR1361756](#)
- [170] TALAGRAND, M. (1996). Majorizing measures: the generic chaining. *Ann. Probab.* **24** 1049–1103. [MR1411488](#)
- [171] TALAGRAND, M. (2014/21). *Upper and Lower Bounds for Stochastic Processes. Modern Methods and Classical Problems* (1st ed. 2014, 2nd ed. 2021). Springer, Berlin. [MR3184689](#), [MR4381414](#)
- [172] TAYLOR, S.J. (1967). Sample Path Properties of a Transient Stable Process. *J. Math. Mech.* **16** 1229–1246. [MR0208684](#)
- [173] TORCHINSKY, A. (1986). *Real-Variable Methods in Harmonic Analysis*.

- Academic Press, San Diego, CA. [MR0869816](#)
- [174] TOTOKI, H. (1961/62). A method of construction of measures on function spaces and its applications to stochastic processes. *Mem. Fac. Sci. Kyushu Univ. Ser. A* **15** 178–190. [MR0138127](#)
- [175] VAN CASTEREN, J.A. (2011). *Markov Processes, Feller Semigroups and Evolution Equations*. World Scientific, New Jersey. [MR2779929](#)
- [176] WANG, G. (1991). Sharp square-function inequalities for conditionally symmetric martingales. *Trans. Amer. Math. Soc.* **328** 393–419. [MR1018577](#)
- [177] WENTZELL (VEN'CEL), A.D. (1979). *Theorie zufälliger Prozesse*. Birkhäuser, Basel and Akademie-Verlag, Berlin. [MR0571888](#), [MR0563923](#). English translation (1981): *A Course in The Theory of Stochastic Processes*. McGraw–Hill, New York. [MR0781738](#)