Contraction principle for trajectories of random walks and Cramér's theorem for kernel-weighted sums

Vladislav Vysotsky

University of Sussex
Pevensey 2 Building, Falmer Campus,
Brighton BN1 9QH, United Kingdom
and
St. Petersburg Department of Steklov Mathematical Institute
Fontanka 27
St. Petersburg 191011, Russia.
E-mail address: v.vysotskiy@sussex.ac.uk
URL: https://profiles.sussex.ac.uk/p406081-vladislav-vysotskiy

Abstract. In 2013 A.A. Borovkov and A.A. Mogulskii proved a weaker-than-standard "metric" large deviations principle (LDP) for trajectories of random walks in $\mathbb{R}^d$ whose increments have the Laplace transform finite in a neighbourhood of zero. We prove that general metric LDPs are preserved under uniformly continuous mappings. This allows us to transform the result of Borovkov and Mogulskii into standard LDPs. We also give an explicit integral representation of the rate function they found. As an application, we extend the classical Cramér theorem by proving an LPD for kernel-weighted sums of i.i.d. random vectors in $\mathbb{R}^d$.

1. Introduction

The study of large deviations of trajectories of random walks was initiated by A.A. Borovkov in the 1960's. Mogulskii (1976) proved a large deviations result for the trajectories of a multidimensional random walk under the assumption that the Laplace transform of its increments is finite. In this work A.A. Mogulskii also studied the large deviations under the weaker *Cramér moment assumption*, i.e. when the Laplace transform is finite in a neighbourhood of zero, but these results appear to be of a significantly limited use.

The further progress was due to the concept of *metric* large deviations principles (LDPs, in short) on general metric spaces introduced by Borovkov and Mogulskii (2010). The upper bound in such

---

metric LDPs is worse than the conventional one – the infimum of the rate function is taken over shrinking ε-neighbourhoods of a set rather than over its closure as in standard LDPs; compare definitions (2.1) and (2.2) below. These bounds may differ when the rate function is *not tight*, i.e. its sub-level sets are non-compact.

For the scaled trajectories of random walks under the Cramér moment assumption, Borovkov and Mogulskii (2013, 2014) obtained a metric LDP in the space $D[0, 1]$ of càdlàg functions equipped with a Skorokhod-type metric generating a version of the topology M_1; see Theorem A. Their rate function has totally bounded sub-level sets but in general, it is not tight because the metric space is not complete. The authors offer no discussion on whether the usual LDP for the trajectories fails.

Our paper was motivated by the question whether this metric LDP can be converted, in any reasonable sense, to a standard LDP. Our progress in this direction is as follows.

Our first result is a *contraction principle* for metric LDPs on general metric spaces (Theorem 3.1). In particular, it shows that a uniformly continuous mapping to a complete metric space transfers a metric LDP into a standard one if the sub-level sets of the rate function are totally bounded (Corollary 3.2). This allows us to transform the metric LDP of Borovkov and Mogulskii (2013, 2014) into standard LDPs with tight rate functions (Theorem 3.3). For example, this yielded a standard LDP (Proposition 5.2) for scaled trajectories of random walks in $\mathbb{R}^d$ considered as random elements of the space $BV[0, 1]$ of functions of bounded variation equipped with the metric ρ_* of joint convergence of functions in L^1 and of their values at 1. It metrizes weak-* convergence on sets of functions of uniformly bounded variation. It is in a certain sense shorter than the Skorokhod metric generating the topology M_1 (Lemma 4.3).

We used ideas from calculus of variations, which offers well-developed methods for working with integral (action) functionals on the space $BV[0, 1]$ equipped with the weak-* topology. In particular, this allowed us to find an explicit *integral representation* (Theorem 5.1) for the rate function of Borovkov and Mogulskii (2013, 2014), who gave such representation only in dimension one. We also found a wide class of sets where the upper bound in the metric LDP of Borovkov and Mogulskii (2013, 2014) coincides with the standard one (see Propositions 3.4 and 5.4 and Remark 5.5).

We have the following applications of our results on the trajectories of random walks.

First, from our contraction principle for the trajectories (i.e. Theorem 3.3) we obtained the LDPs for the *perimeter* and the *area* of the *convex hull* of a planar random walk, presented in Akopyan and Vysotsky (2021+, Proposition 4.1). This application motivated our initial interest in the questions considered here.

Second, from our LDP for the trajectories in $(BV[0, 1], \rho_*)$ (i.e. Proposition 5.2), we obtained the LDP for *kernel-weighted sums* of i.i.d. random vectors in $\mathbb{R}^d$ (Theorem 6.1). This extends the classical Cramér theorem, without any additional assumptions. We give an explicit formula for the rate function, which is especially accessible in dimension one. To the best of our knowledge, the strongest available results in this direction are by Kiesel and Stadtmüller (2000) and Gantert et al. (2014), both works concerning dimension one only. The former paper considers the "light-tailed" case where the i.i.d. terms have finite Laplace transform. The latter paper considers the "heavy-tailed" case, where the common distribution of the terms has "stretched" exponential tails and thus does not satisfy the Cramér moment assumption. Our Theorem 6.1 complements these one-dimensional results, showing that there is a natural transition of the rate function from the "light-tailed" to the "heavy-tailed" cases; see also Remark 6.2.c.

When the current paper was ready for submission, we became aware of the work of Najim (2002). His main result, Theorem 3.1 on the LDP for weighted sums of i.i.d. random vectors, is stronger than the LDP in our Theorem 6.1. Our new contributions are the simplified explicit expressions for the rate function (formulas (6.2) and (6.3)) and for the corresponding minimizing trajectories we found in certain cases (Remark 6.2.c). As a corollary to his main result, J. Najim obtained an LDP in the weak-* topology on $BV[0, 1]$ (Najim, 2002, Theorems 4.1 and 4.3) for trajectories of

random walks under the Cramér moment assumption. This result is very similar (but not equivalent, cf. Section 4.1) to our Proposition 5.2.

Notably, we obtained our extension of Cramér's theorem as a corollary and only at the last step of our work, as opposed to the argument of Najim (2002) going in the reverse direction. Therefore, despite of a significant intersection with this work, we believe that the corresponding part of our paper is of independent interest since it has a different motivation and uses a different approach – we used Proposition 5.2, which itself follows from the metric LDP in Borovkov and Mogulskii (2013, 2014) using the metrization result of Theorem B, the contraction principle in Theorem 3.1, and geometric Lemma 4.3, which is of its own interest.

This paper is organized as follows. Section 2 presents the metric LDP for trajectories of random walks by Borovkov and Mogulskii (2013, 2014) and also introduces metric LDPs in general metric spaces. In Section 3 we prove a contraction principle for uniformly continuous mappings of metric LDPs, and present its application for random walk trajectories. In Section 4 we define the weak-* and related topologies on the space of functions of bounded variation, and compare them with and the Skorokhod topologies M_1 and M_2 and their modifications. The main result of the section is geometric Lemma 4.3, which compares the metric ρ_* with Skorokhod metrics. The setup and results of this section are used in Section 5, where we give an integral representation for the rate function, prove an LDP for the trajectories in the space $(BV[0,1], \rho_*)$, and present a few types of sets where the upper bound in the metric LDP for the trajectories equals the standard one. The last section concerns our main application, Cramér's theorem for kernel-weighted sums of i.i.d. random vectors in $\mathbb{R}^d$.

2. Metric LDPs for trajectories of random walks

In this section we give necessary definitions and provide a brief summary of the results of Borovkov and Mogulskii (2013, 2014).

2.1. *General metric LDPs.* We start with general definitions. Let $\mathcal{X}$ be a Hausdorff topological space equipped with the Borel σ-algebra. Let $\mathcal{I} : \mathcal{X} \to [0, +\infty]$ be a lower semi-continuous function such that $\mathcal{I} \not\equiv +\infty$. By definition, this means that the *sub-level sets* $\{x \in \mathcal{X} : \mathcal{I}(x) \le \alpha\}_{\alpha \in [0,\infty)}$ of $\mathcal{I}$ are closed. If $\mathcal{X}$ is a metric space, this is equivalent to requiring that for every $x \in \mathcal{X}$, we have $\mathcal{I}(x) \le \liminf_{n \to \infty} \mathcal{I}(x_n)$ for any sequence $(x_n)_{n \ge 1}$ converging to x.

We say that a sequence $(Z_n)_{n \ge 1}$ of random elements of $\mathcal{X}$ satisfies the *large deviations principle (LDP)* in $\mathcal{X}$ with a *rate function* $\mathcal{I}$ and a *speed* $(a_n)_{n \ge 1} \subset (0, \infty)$ if for every Borel set $B \subset \mathcal{X}$,

$$- \inf_{x \in \operatorname{int} B} \mathcal{I}(x) \le \liminf_{n \to \infty} \frac{1}{a_n} \log \mathbb{P}(Z_n \in B) \le \limsup_{n \to \infty} \frac{1}{a_n} \log \mathbb{P}(Z_n \in B) \le - \inf_{x \in \operatorname{cl} B} \mathcal{I}(x), \qquad (2.1)$$

where $\inf_\varnothing := +\infty$ by the usual convention. If $\mathcal{X}$ is equipped with a metric ℓ, we say that $(Z_n)_{n \ge 1}$ satisfies the *metric LDP* in $(\mathcal{X}, \ell)$ with a rate function $\mathcal{I}$ and a speed $(a_n)_{n \ge 1}$ if

$$- \inf_{x \in \operatorname{int} B} \mathcal{I}(x) \le \liminf_{n \to \infty} \frac{1}{a_n} \log \mathbb{P}(Z_n \in B) \le \limsup_{n \to \infty} \frac{1}{a_n} \log \mathbb{P}(Z_n \in B) \le - \lim_{\varepsilon \to 0+} \inf_{x \in B_\ell^\varepsilon} \mathcal{I}(x), \quad (2.2)$$

where B_ℓ^ε denotes the open ε-neighbourhood of B in ℓ (with the convention $\varnothing_\ell^\varepsilon := \varnothing$). For brevity, we will not refer to the speed when $a_n = n$, which essentially is the only case of our interest in this paper. We stress that the concept of metric LDPs is not a topological one, in the sense that (2.2) may cease to hold when ℓ is replaced by another metric generating the same topology.

We say that the rate function $\mathcal{I}$ is *tight* if its sub-level sets are compact. A natural way to relax this condition in the case when $\mathcal{X}$ is a metric space is to assume total boundedness of the sub-level sets of $\mathcal{I}$. Recall that a subset of a metric space is *totally bounded* if it has a finite ε-net for every $\varepsilon > 0$; such subset is relatively compact, i.e. its closure is compact, if the metric space is *complete*.

Note that if $\mathcal{I}$ is tight, then (2.1) and (2.2) coincide because

$$\inf_{x \in \mathrm{cl}\, B} \mathcal{I}(x) = \lim_{\varepsilon \to 0+} \inf_{x \in B_\ell^\varepsilon} \mathcal{I}(x); \tag{2.3}$$

see Borovkov and Mogulskii (2012, Lemma 1.1) or Dembo and Zeitouni (2010, Lemma 4.1.6.b). It is also easy to see that this equality holds true for any $\mathcal{I}$ when B is relatively compact.

Equality (2.3) may not hold in general (see Remark 2.2.d), and thus a metric LDP is in general strictly weaker than the corresponding standard one. For an example when a metric LDP holds true but the standard one does not, see Bazhba et al. (2020, Theorem 2.1 and Section 3.3). On the other hand, a metric LDP is stronger than the corresponding *weak* LDP, defined as the standard one but with the upper bound required only for relatively compact sets. For this reason, we find it misleading to refer to metric LDPs as *extended* ones, which is the original terminology of Borovkov and Mogulskii (2010). Moreover, neither standard nor weak LDPs require a metric structure.

2.2. *Skorokhod topologies.* We will write $x = (x^{(1)}, \ldots, x^{(d)})$ for the Cartesian coordinates of $x \in \mathbb{R}^d$ with $d \in \mathbb{N}$, $|x|$ for the Euclidean norm, and '·' for the scalar product on $\mathbb{R}^d$.

Denote by $D[0,1] = D([0,1]; \mathbb{R}^d)$ the set of *càdlàg* functions on $[0,1]$, that is right-continuous $\mathbb{R}^d$-valued functions without discontinuities of the second kind[1]. We will consider several Skorokhod-type metrics and topologies on $D[0,1]$. We refer to the book by Whitt (2002), where Chapter 12 gives a comprehensive treatise of the matter.

The *completed graph* Γh of a function $h \in D[0,1]$ is a subset of $[0,1] \times \mathbb{R}^d$ defined by

$$\Gamma h := \big\{(t,x) : 0 \le t \le 1, x \in [h(t-), h(t+)]\big\},$$

where $h(0-) := h(0)$, $h(1+) := h(1)$, and $[u_1, u_2]$ denotes the line segment with the endpoints $u_1, u_2 \in \mathbb{R}^d$. We deliberately wrote $h(t+)$ instead of the equal quantity $h(t)$, to stress that Γh is defined even when h has no discontinuities of the second kind but is not right-continuous. We equip the completed graphs with the topology induced from $[0,1] \times \mathbb{R}^d$.

Consider a set of parametrizations of the completed graph:

$$\Pi(h) := \big\{\gamma \big| \gamma : [0,1] \to \Gamma h \text{ is bijective, continuous, and satisfying } \gamma(0) = (0, h(0))\big\}$$

(this set is non-empty by Whitt 2002, Remark 12.3.3). We can thus regard the completed graphs as images of continuous curves in $\mathbb{R}^d$. The metric ρ_1 on $D[0,1]$ is the least uniform distance between parametrizations of completed graphs:

$$\rho_1(h_1, h_2) := \inf_{\gamma_1 \in \Pi(h_1), \gamma_2 \in \Pi(h_2)} \sup_{t \in [0,1]} |\gamma_1(t) - \gamma_2(t)|. \tag{2.4}$$

The topology generated by ρ_1 is called the *Skorokhod topology M_1*; see Whitt (2002, Remark 12.3.4). The metric ρ_2 on $D[0,1]$ is the Hausdorff distance d_H between the completed graphs, i.e. $\rho_2(h_1, h_2) := d_H(\Gamma h_1, \Gamma h_2)$. In other words,

$$\rho_2(h_1, h_2) = \max_{i=1,2} \max_{(s,x) \in \Gamma h_i} \min_{(t,y) \in \Gamma h_{3-i}} |(s,x) - (t,y)|. \tag{2.5}$$

This is a genuine distance because each Γh is compact, being a continuous image of $[0,1]$. The topology generated by ρ_2 is called the *Skorokhod topology M_2*; see Whitt (2002, Theorem 12.10.1).

Furthermore, consider the modified completed graphs $\Gamma' h := \Gamma h \cup [0, h(0)]$ and define ρ_1' and ρ_2' exactly as above using Γ' (parametrized by functions in Π') instead of Γ. Then ρ_1', ρ_2' are metrics too. Equivalently, for $h_1, h_2 \in D[0,1]$ we can write

$$\rho_i'(h_1, h_2) = \rho_i(h_1 \mathbb{1}_{(0,1]}, h_2 \mathbb{1}_{(0,1]}) \tag{2.6}$$

---

[1]A function on $[0,1]$ has no discontinuities of the second kind if it has right and left limits at every point.

if we extend definitions (2.4) and (2.5) to the space of functions on $[0,1]$ that are càdlàg on $(0,1]$ and have right limit at 0, where the ρ_i's still remain metrics because completed graphs uniquely define such functions.

From the definitions above, we readily obtain the following inequalities on $D[0,1]$:

$$\rho_2' \leq \rho_1' \leq \rho_1 \quad \text{and} \quad \rho_2 \leq \rho_1. \tag{2.7}$$

The inequality $\rho_1' \leq \rho_1$ follows because $\rho_1(h_1, h_2)$ equals the infimum in the definition of ρ_1' taken over the pairs of parametrizations $\gamma_1 \in \Pi'(h_1)$, $\gamma_2 \in \Pi'(h_2)$ satisfying $\gamma_1([0,t]) = [0, h_1(0)]$ and $\gamma_2([0,t]) = [0, h_2(0)]$ for some $t \in [0,1]$. Moreover, we have

$$\begin{aligned}
\rho_2'(h_1, h_2) &= d_H\big(\Gamma h_1 \cup [0, h_1(0)], \Gamma h_2 \cup [0, h_2(0)]\big) \\
&\leq \max\Big(d_H(\Gamma h_1, \Gamma h_2), d_H\big([0, h_1(0)], [0, h_2(0)]\big)\Big) \\
&\leq \max\big(\rho_2(h_1, h_2), |h_1(0) - h_2(0)|\big), \tag{2.8}
\end{aligned}$$

where we used the fact that $d_H(A_1 \cup B_1, A_2 \cup B_2) \leq \max(d_H(A_1, A_2), d_H(B_1, B_2))$ for any non-empty $A_1, A_2, B_1, B_2 \subset \mathbb{R}^{d+1}$ (which follows easily from the definition of d_H).

We will work with the topologies M_1' and M_2' on $D[0,1]$ generated respectively by ρ_1' and ρ_2'. The topology M_1' was recently used by Bazhba et al. (2020); the versions of M_1' and M_2' on $D[0,\infty)$ briefly appeared in Whitt (2002, Section 13.6.2). Inequalities (2.7) and (2.8) imply that

$$M_2' \subset M_1' \subset M_1 \quad \text{and} \quad M_2' \subset M_2 \subset M_1, \tag{2.9}$$

where the third inclusion follows from (2.8) using that every M_2'-closed set is also closed in M_2 because convergence of functions in M_2 implies convergence of their values at 0.

The following result, presented without a proof because we will not use it in this paper, describes convergence in M_i' in terms of more standard convergence in M_i. It shows that naturally, the difference between the convergences is only in the behaviour around time 0. In particular, the value at 0 is not an M_i'-continuous functional.

Proposition 2.1. *Let $h, h_1, h_2, \ldots \in D[0,1]$ and $i \in \{1, 2\}$. Then $\lim_{n \to \infty} \rho_i'(h_n, h) = 0$ if and only if there exits a sequence $t_1, t_2, \ldots \in [0,1]$ such that $\lim_{n \to \infty} t_n = 0$,*

$$\lim_{n \to \infty} \rho_i(h_n(\cdot \vee t_n), h) = 0,$$

and

$$\begin{cases}
\lim_{n \to \infty} \sup_{0 \leq t \leq t_n} \sup_{0 \leq s \leq t} \big(|h_n(s)| \cdot |h(0)| - h_n(t) \cdot h(0)\big) = 0, & \text{if } i = 1, \\
\lim_{n \to \infty} \sup_{0 \leq t \leq t_n} \min_{0 \leq s \leq 1} |sh(0) - h_n(t)| = 0, & \text{if } i = 2.
\end{cases}$$

The last condition means that whenever $h(0) \neq 0$, for $i = 1$ the values of h_n on $[0, t_n]$ are nearly proportional to $h(0)$ and their orthogonal projections on $h(0)$ are nearly non-decreasing, and for $i = 2$ they nearly belong to $[0, h(0)]$.

Lastly, we note that each of the metrics ρ_i, ρ_i' is separable and incomplete. Separability follows from Lemma 1 in Section 14 in Billingsley (1968). Observing that $\{\mathbb{1}_{[1/2, 1/2+1/n)}\}_{n \geq 2}$ is a Cauchy sequence in each of the metrics which does not converge shows their incompleteness.

2.3. *Metric LDPs for trajectories of random walks.* Let $(S_n)_{n \geq 1}$, where $S_n = X_1 + \ldots + X_n$, be a random walk with independent identically distributed increments $X_1, X_2, \ldots$ in $\mathbb{R}^d$, where $d \geq 1$. For any $n \in \mathbb{N}$, let $S_n(\cdot)$ be the piece-wise linear function on $[0,1]$ defined by linear interpolation between its values at the points k/n, where $0 \leq k \leq n$, $k \in \mathbb{Z}$, that are given by $S_n(k/n) := S_k$, where $S_0 := 0$. These are time-rescaled trajectories of the random walk $(S_n)_{n \geq 1}$. We will regard them as random elements of the space $D[0,1]$ (or its subsets) equipped with the Borel σ-algebras generated by ρ_1' or ρ_2'.

Let $\mathcal{L}(u) := \mathbb{E}e^{u \cdot X_1}$, where $u \in \mathbb{R}^d$, be the Laplace transform of the random vector X_1 in $\mathbb{R}^d$. Denote by $\mathcal{D}_\mathcal{L} := \{u \in \mathbb{R}^d : \mathcal{L}(u) < \infty\}$ the *effective domain* of $\mathcal{L}$. We say that X_1 satisfies the *Cramér moment assumption* if $\mathcal{L}$ is finite in an open neighbourhood of 0, that is $0 \in \operatorname{int}\mathcal{D}_\mathcal{L}$ in short. The function $K := \log\mathcal{L}$, called the *cumulant moment generating function* of X_1, is always convex. Denote by I the *Legendre–Fenchel transform* of K, i.e.

$$I(v) := \sup_{u \in \mathbb{R}^d} \big(u \cdot v - K(u)\big), \quad v \in \mathbb{R}^d. \tag{2.10}$$

This is a convex lower semi-continuous function with values in $[0, \infty]$. The classical Cramér theorem states that under the assumption $0 \in \operatorname{int}\mathcal{D}_\mathcal{L}$, the sequence $(S_n/n)_{n \geq 1}$ satisfies the LDP in the Euclidean space $\mathbb{R}^d$ with the tight rate function I. This justifies saying that I is the *rate function* of X_1.

The *total variation* $\operatorname{Var}(h)$ of a function $h \in D[0,1]$ is defined by

$$\operatorname{Var}(h) := \sup_{\mathbf{t} \subset (0,1] : \#\mathbf{t} < \infty} \int_0^1 |(h^{\mathbf{t}})'(s)| ds, \tag{2.11}$$

where $h^{\mathbf{t}}$ denotes the continuous function on $[0,1]$ defined by linear interpolation between its values $\mathbf{t} \cup \{0, 1\}$ that are given by $h^{\mathbf{t}}(s) := h(s)$ for $s \in \mathbf{t} \cup \{1\}$ and $h^{\mathbf{t}}(0) := 0$. We can regard $\operatorname{Var}(h)$ as the length of the curve in $\mathbb{R}^d$ obtained by taking the spatial coordinate of a curve in $\Pi'(h)$. Denote by $BV[0,1] := \{h \in D[0,1] : \operatorname{Var}(h) < \infty\}$ the set of càdlàg $\mathbb{R}^d$-valued functions of *bounded variation*, and by $AC_0[0,1]$ its subset of coordinate-wise *absolutely continuous* functions such that $h(0) = 0$.

Lastly, define a non-negative functional I_D on $h \in D[0,1]$:

$$I_D(h) := \sup_{\mathbf{t} \subset (0,1] : \#\mathbf{t} < \infty} \int_0^1 I((h^{\mathbf{t}})'(s)) ds. \tag{2.12}$$

It is worth noting that $I_D(h) = \int_0^1 I(h'(s)) ds$ when $h \in AC_0[0,1]$ (Borovkov and Mogulskii 2013, Theorem 5.3).

We now present the metric LDP for trajectories of random walks.

Theorem A (Borovkov and Mogulskii 2013, 2014). *Assume that X_1 is a random vector in $\mathbb{R}^d$ such that $0 \in \operatorname{int}\mathcal{D}_\mathcal{L}$. Then the random sequences $(S_n(\cdot)/n)_{n \geq 1}$ and $(S_{[n\cdot]}/n)_{n \geq 1}$ satisfy the metric LDPs* (2.2) *in each of the four metric spaces $(D[0,1], \rho_i')$ and $(BV[0,1], \rho_i')$ for $i \in \{1, 2\}$, with the rate function I_D whose sub-level sets are totally bounded (in each space).*

Moreover, I_D is convex and it satisfies, for some constants $c_1, c_2 > 0$,

$$I_D(h) \geq c_1 \operatorname{Var}(h) - c_2, \qquad h \in D[0,1]. \tag{2.13}$$

Remark 2.2. Let us make a number of comments.

a) Note that $I_D(h) = +\infty$ for $h \notin BV[0,1]$ by (2.13).

b) The metric LDPs for $S_n(\cdot)$ and $S_{[n\cdot]}$ are equivalent by $\rho_i(S_n(\cdot)/n, S_{[n\cdot]}/n) \leq 2/n$. The metric LDPs in ρ_1' are stronger than the ones in ρ_2' since $\rho_2' \leq \rho_1'$. We stated these weaker results to match the presentation of Borovkov and Mogulskii (2013, 2014), which puts emphasis on the metric ρ_2'. The only advantage of ρ_2' is in its relative simplicity.

c) Although the rate function I_D is not tight, equality (2.3) still holds true with $\mathcal{I} = I_D$ and $\ell = \rho_1'$ for certain types of non-relatively compact sets B, described in Remark 5.5.

d) In general, (2.3) does not hold for $\ell = \rho_2'$. For example, assume that $d = 1$, $\mathbb{E}X_1 = 0$, $\mathcal{D}_\mathcal{L}$ is bounded, and consider $B := \{h_n\}_{n \geq 4}$ with $h_n := \mathbb{1}_{[1/2 - 2/n, 1/2 - 1/n) \cup [1/2 + 1/n, 1/2 + 2/n)}$. No subsequence $\{h_{n_k}\}$ converges in ρ_2' to an element of $D[0,1]$, hence B is closed w.r.t. ρ_2'. On the other hand, we have $\rho_2'(g_n, B) = 1/n$ for $g_n := \mathbb{1}_{[1/2 - 1/n, 1/2 + 1/n)}$. Then $I_D(g_n) = \frac{1}{2}I_D(h_n) = const > 0$ by equality (5.7) below, therefore (2.3) cannot hold.

It is plausible that (2.3) does not hold for $\ell = \rho_1'$ too but we have no examples. The papers Borovkov and Mogulskii (2010, 2012, 2013, 2014) offer no discussion on this question.

e) The reason why the metric LDPs do not immediately imply the corresponding LDPs is incompleteness of the metric spaces considered. There exists an explicit complete metric that generates the topology M_1 (Whitt 2002, Section 12.8), and it appears that its minor modification should give a complete metric $\tilde{\rho}'_1$ generating M'_1. However, such complete metrics are longer than the initial ones, therefore it seems impossible to have the upper bound in (2.2) with $\ell = \tilde{\rho}'_1$ instead of $\ell = \rho'_1$.

f) Bound (2.13) readily follows from the inequality $I(v) \geq c_1|v| - c_2$ for $v \in \mathbb{R}^d$, which holds true because I is convex and grows at least linearly at infinity; see (5.4) and (5.5) below.

g) In view of (2.6) and given that $S_n(0) = S_0 = 0$, one may argue that it would be more natural to employ the non-standard space $D'[0,1] := \{h\mathbb{1}_{(0,1]} : h \in D[0,1]\}$ equipped with the usual Skorokhod metrics ρ_i. This is essentially done in Borovkov and Mogulskii (2013, 2014). The spaces $(D'[0,1], \rho_i)$ and $(D[0,1], \rho'_i)$ are isometric, so the metric LDPs transfer easily. On the other hand, it is natural to work with the space $BV[0,1]$ of càdlàg modifications of functions of bounded variation because these are distribution functions of vector-valued finite measures. This explains our choice of the standard space $D[0,1]$.

h) The ultimate reason why the space $BV[0,1]$ arises is that the rate function I is not super-linear at infinity. The only exception is when the Laplace transform of the increments is finite (see (5.4) and (5.5)), in which case it suffices to work with the space $(AC_0[0,1], \|\cdot\|_\infty)$ and the rate function I_D is tight due to the super-linearity of I. Such effects of the behaviour of integrand at infinity are well-known in the calculus of variations, which studies minimization of integral functionals.

Theorem A is a combination and adaptation of several results scattered through Borovkov and Mogulskii (2012, 2013, 2014), therefore we shall explain in detail how we obtained it. These authors considered a wider space $\mathbb{D}$ of functions $h : [0,1] \to \mathbb{R}^d$ without discontinuities of the second kind satisfying $h(t) \in [h(t-), h(t+)]$ for $t \in [0,1]$. The functions ρ_i, defined on $\mathbb{D}$ as above in (2.4) and (2.5), are now pseudometrics. Put $\mathbb{D}_0 := \{h \in \mathbb{D} : h(0) = 0\}$ and define the functionals Var and I_D on $\mathbb{D}_0$ as above in (2.11) and (2.12). The functional $I_{\mathbb{D}}$ on $\mathbb{D}$ is defined by $I_{\mathbb{D}} := I_D$ on $\mathbb{D}_0$ and $I_{\mathbb{D}} := +\infty$ on $\mathbb{D}_0^c$, see Borovkov and Mogulskii (2012, Definition 2.1). Lastly, putting $h^+(t) := h(t+)$ for $t \in [0,1]$ defines an isometry from $(\mathbb{D}_0, \rho_i)$ onto $(D[0,1], \rho'_i)$.

Then $(S_{[n\cdot]}/n)_{n\geq 1}$ satisfies the metric LDP in $(\mathbb{D}, \rho_2)$ with convex rate function $I_{\mathbb{D}}$ by Borovkov and Mogulskii (2013, Theorem 5.5), where measurability refers to the Borel σ-algebra (see Borovkov and Mogulskii 2012, Definition 1.4) and $I_{\mathbb{D}}$ is convex and lower semi-continuous by Borovkov and Mogulskii (2013, Theorem 5.2.i and ii) and its sub-level sets are totally bounded by Borovkov and Mogulskii (2013, Lemma 5.3) combined with inequality (2.13). By Lemma 3.5.a, the metric LDP remains valid on $(\mathbb{D}_0, \rho_2)$ because $\mathbb{P}(S_{[n\cdot]} \in \mathbb{D}_0) = 1$ for every n and $I_{\mathbb{D}} = +\infty$ on $\mathbb{D}_0^c$. Finally, by our Theorem 3.1, the isometry $h \mapsto h^+$ transforms this metric LDP into the one on $(D[0,1], \rho'_2)$ with the rate function I_D because $S_{[n\cdot]}^+ = S_{[n\cdot]}$ and $I_{\mathbb{D}}(h) = I_D(h^+)$ for every $h \in \mathbb{D}_0$ by Borovkov and Mogulskii (2013, Theorem 5.1).

Denote by $\mathbb{V}_0$ the subset of $\mathbb{D}_0$ of functions of finite variation. By Borovkov and Mogulskii (2014, Theorem 6.2), $(S_{[n\cdot]}/n)_{n\geq 1}$ satisfies the metric LDP in $(\mathbb{V}_0, \rho_1)$ with the rate function $I_{\mathbb{D}}$, which is lower semi-continuous (in ρ_1) because it is so in the shorter pseudometric ρ_2, and its sub-level sets are totally bounded by Borovkov and Mogulskii (2014, Lemma 6.2) combined with (2.13). This metric LDP remains valid in $(BV[0,1], \rho'_1)$ by the same argument as above using Remark 2.2.a. It in turn implies the weaker metric LDP in $(BV[0,1], \rho'_2)$ and also implies the one in $(D[0,1], \rho'_1)$ by Lemma 3.5.b using that I_D is lower semi-continuous in ρ'_1 because it is so in ρ'_2.

3. Contraction principle for metric LDPs

The following general result is analogous to the usual contraction principle for standard LDPs; cf. Dembo and Zeitouni (2010, Theorem 4.2.1 and Remark (c)). To state it, we first give two definitions.

For a function $J : \mathcal{X} \to [0, +\infty]$ defined on a topological space $\mathcal{X}$, denote by $\operatorname{cl} J$ its *closure* (or the *lower semi-continuous regularization*), i.e. the function whose epigraph is the closure (in the product topology on $\mathcal{X} \times [0, +\infty]$) of the epigraph of J. Recall that $\mathcal{D}_J = \{x \in \mathcal{X} : J(x) < \infty\}$ denotes the effective domain of J. We say that a mapping F between metric spaces is uniformly continuous *on a subset A* of the domain if $F|_A$ is uniformly continuous.

Theorem 3.1. *Let $(\mathcal{X}, \ell_1)$ and $(\mathcal{Y}, \ell_2)$ be metric spaces and $(Z_n)_{n \geq 1}$ be a sequence of random elements that satisfies a metric LDP in $(\mathcal{X}, \ell_1)$ with some speed and a rate function $\mathcal{I}$. Let $F : \mathcal{X} \to \mathcal{Y}$ be a measurable mapping that is continuous at every $x \in \mathcal{D}_\mathcal{I}$ and uniformly continuous on every sub-level set of $\mathcal{I}$. Then the sequence $(F(Z_n))_{n \geq 1}$ satisfies the metric LDP in $(\mathcal{Y}, \ell_2)$ with the same speed and the rate function $\operatorname{cl} \tilde{\mathcal{J}}$, where $\tilde{\mathcal{J}}(y) := \inf_{x \in F^{-1}(y)} \mathcal{I}(x)$ for $y \in \mathcal{Y}$.*

Moreover, if the sub-level sets of $\mathcal{I}$ are totally bounded, then the same is true for $\operatorname{cl} \tilde{\mathcal{J}}$.

The interest in this result is in its corollary, which allows one to bring the metric LDPs (2.2) into the standard form (2.1).

Corollary 3.2. *If the metric space $(\mathcal{Y}, \ell_2)$ is complete and the sub-level sets of $\mathcal{I}$ are totally bounded, then the sequence $(F(Z_n))_{n \geq 1}$ satisfies the (standard) LDP with the tight rate function $\operatorname{cl} \tilde{\mathcal{J}}$.*

This follows from Theorem 3.1 by equality (2.3) and the fact that closed totally bounded subsets of complete metric spaces are compact.

Our main application of Corollary 3.2 is in the context of random walks trajectories:

Theorem 3.3. *Assume that X_1 is a random vector in $\mathbb{R}^d$ such that $0 \in \operatorname{int} \mathcal{D}_\mathcal{L}$. Let $\mathcal{Y}$ be a complete metric space and $F : BV[0,1] \to \mathcal{Y}$ be a mapping that is continuous in ρ_1' and uniformly continuous in ρ_1' on $\{h : \operatorname{Var}(h) \leq R\}$ for every $R > 0$. Then both sequences of random elements $(F(S_n(\cdot)/n))_{n \geq 1}$ and $(F(S_{[n \cdot]}/n))_{n \geq 1}$ satisfy the (standard) LDP in $\mathcal{Y}$ with the tight rate function $\operatorname{cl} \tilde{\mathcal{J}}$, where $\tilde{\mathcal{J}}(y) := \inf_{h \in F^{-1}(y)} I_D(h)$ for $y \in \mathcal{Y}$.*

This follows from Corollary 3.2 combined with Theorem A using the lower bound (2.13) for $I_D(h)$ in terms of $\operatorname{Var}(h)$.

Proof of Theorem 3.1. For any $A \subset \mathcal{Y}$, we have

$$\{x : \mathcal{I}(x) < \infty\} \cap F^{-1}(\operatorname{int} A) = \{x : \mathcal{I}(x) < \infty\} \cap \operatorname{int}(F^{-1}(\operatorname{int} A))$$

because for every x in the set in the l.h.s., $\{x\}_{\ell_1}^\delta \subset F^{-1}(\operatorname{int} A)$ holds for some $\delta > 0$ by continuity of F at x, and thus $x \in \operatorname{int}(F^{-1}(\operatorname{int} A))$. Hence

$$\inf_{x \in \operatorname{int}(F^{-1}(\operatorname{int} A))} \mathcal{I}(x) = \inf_{x \in F^{-1}(\operatorname{int} A)} \mathcal{I}(x) = \inf_{y \in \operatorname{int} A} \tilde{\mathcal{J}}(y), \tag{3.1}$$

where the second equality holds true by the definition of $\tilde{\mathcal{J}}$. Furthermore, we claim that

$$\lim_{\delta \to 0+} \inf_{x \in (F^{-1}(A))_{\ell_1}^\delta} \mathcal{I}(x) \geq \lim_{\varepsilon \to 0+} \inf_{x \in F^{-1}(A_{\ell_2}^\varepsilon)} \mathcal{I}(x) = \lim_{\varepsilon \to 0+} \inf_{y \in A_{\ell_2}^\varepsilon} \tilde{\mathcal{J}}(y). \tag{3.2}$$

The inequality is trivial when its l.h.s. is infinite, otherwise denote the l.h.s. by R. By uniform continuity of F on sub-level sets of $\mathcal{I}$, for any $\varepsilon > 0$ there exists a $\delta_0 > 0$ such that $\ell_2(F(x), F(x')) < \varepsilon$ whenever $\ell_1(x, x') < \delta_0$ and $\max(\mathcal{I}(x), \mathcal{I}(x')) \leq R$. Hence

$$\left(F^{-1}(A) \cap \{x : \mathcal{I}(x) \leq R\}\right)_{\ell_1}^{\delta_0} \subset F^{-1}(A_{\ell_2}^\varepsilon),$$

and for any $\delta \in (0, \delta_0)$ we have

$$R \geq \inf_{x \in (F^{-1}(A))_{\ell_1}^{\delta_0}} \mathcal{I}(x) = \inf_{x \in (F^{-1}(A))_{\ell_1}^{\delta_0} \cap \{\mathcal{I} \leq R\}_{\ell_1}^{\delta_0}} \mathcal{I}(x) = \inf_{x \in (F^{-1}(A) \cap \{\mathcal{I} \leq R\})_{\ell_1}^{\delta_0}} \mathcal{I}(x) \geq \inf_{x \in F^{-1}(A_{\ell_2}^\varepsilon)} \mathcal{I}(x),$$

which implies (3.2) by first taking $\delta \to 0+$ and then $\varepsilon \to 0+$.

Denoting by $(a_n)_{n \geq 1}$ the speed in the metric LDP for $(Z_n)_{n \geq 1}$, for any Borel set $A \subset \mathcal{Y}$,

$$\limsup_{n \to \infty} \frac{1}{a_n} \log \mathbb{P}(F(Z_n) \in A) \leq -\lim_{\delta \to 0+} \inf_{x \in (F^{-1}(A))^{\delta}_{\ell_1}} \mathcal{I}(x) \leq -\lim_{\varepsilon \to 0+} \inf_{y \in A^{\varepsilon}_{\ell_2}} \tilde{\mathcal{J}}(y),$$

where the second inequality follows from (3.2). We also have

$$\liminf_{n \to \infty} \frac{1}{a_n} \log \mathbb{P}(F(Z_n) \in A) \geq - \inf_{x \in \text{int}(F^{-1}(\text{int } A))} \mathcal{I}(x) = - \inf_{y \in \text{int } A} \tilde{\mathcal{J}}(y),$$

where the equality follows from (3.1).

Finally, we can replace $\tilde{\mathcal{J}}$ above by the function $\mathcal{J}$ given by $\mathcal{J} := \text{cl}\,\tilde{\mathcal{J}}$, which is lower semi-continuous by definition, non-negative, and not identically $+\infty$. Indeed, we have

$$\inf_{y \in \text{int } A} \mathcal{J}(y) = \inf_{y \in \text{int } A} \tilde{\mathcal{J}}(y), \qquad A \subset \mathcal{Y}, \tag{3.3}$$

which follows easily from the representation (e.g., see Rassoul-Agha and Seppäläinen, 2015, Lemma 2.8)

$$\mathcal{J}(y) = \sup \left\{ \inf_{z \in U} \tilde{\mathcal{J}}(z) : y \in U, U \subset \mathcal{Y}, U \text{ is open} \right\}, \qquad y \in \mathcal{Y}. \tag{3.4}$$

Thus, the sequence $(F(Z_n))_{n \geq 1}$ of random elements of $\mathcal{Y}$ satisfies the metric LDP in $(\mathcal{Y}, \ell_2)$ with the rate function $\mathcal{J}$ and speed $(a_n)_{n \geq 1}$, as stated.

Note that representation (3.4) also implies that $\text{cl}\,\tilde{\mathcal{J}} \leq \tilde{\mathcal{J}}$ and

$$\{y : \mathcal{J}(y) < \alpha\} \subset \text{cl}\{y : \tilde{\mathcal{J}}(y) < \alpha\}, \qquad \alpha > 0. \tag{3.5}$$

Indeed, if there is a $y \in \mathcal{Y}$ such that $\mathcal{J}(y) < \alpha$ but $y \notin \text{cl}\{y : \tilde{\mathcal{J}}(y) < \alpha\}$, then since $\mathcal{Y}$ is a metric space, there is an open ball U centred at y that does not intersect with $\text{cl}\{y : \tilde{\mathcal{J}}(y) < \alpha\}$. Thus, $\tilde{\mathcal{J}} \geq \alpha$ on U, hence $\mathcal{J}(y) \geq \alpha$ by (3.4), which is a contradiction.

If T is a totally bounded subset of $\mathcal{X}$ and F is uniformly continuous on T, then $F(T)$ is totally bounded in $\mathcal{Y}$. Therefore, if the sub-level sets of $\mathcal{I}$ are totally bounded in $\mathcal{X}$, by

$$\{y : \tilde{\mathcal{J}}(y) < \alpha\} = \left\{ y : \inf_{x \in F^{-1}(y)} \mathcal{I}(x) < \alpha \right\} \subset F(\{x : \mathcal{I}(x) < \alpha\}),$$

the set on the l.h.s. is totally bounded in $\mathcal{Y}$, and so is its closure. Then the sub-level sets of $\mathcal{J}$ are totally bounded by (3.5), as claimed.

$\square$

The proof presented actually reveals a wide class of sets where equality (2.3) holds true and thus the metric LDP bound (2.2) can be strengthened to the standard one (2.1). Let us state this as a separate assertion.

Proposition 3.4. *Let $(\mathcal{X}, \ell)$, $(\mathcal{Y}, \ell_2)$ be metric spaces, and $F : \mathcal{X} \to \mathcal{Y}$, $\mathcal{I} : \mathcal{X} \to [0, \infty]$ be mappings such that F is uniformly continuous on the sub-level sets of $\mathcal{I}$. Assume that the function $\tilde{\mathcal{J}}(y) := \inf_{x \in F^{-1}(y)} \mathcal{I}(x)$ on $\mathcal{Y}$ is tight. Then equality (2.3) holds true for every non-empty set B such that $B = F^{-1}(A)$ for some $A \subset \mathcal{Y}$ satisfying $\text{cl}(F^{-1}(A)) = F^{-1}(\text{cl } A)$, with the infimum on the l.h.s. of (2.3) attained at some $x \in \text{cl } B$.*

For a direct application of this result to random walks trajectories, see Proposition 5.4.

Proof: By (3.2) and (2.3) applied to $\mathcal{I}$ replaced by $\tilde{\mathcal{J}}$, which is tight by assumption (and hence lower semi-continuous), we have

$$\inf_{x \in \text{cl}(F^{-1}(A))} \mathcal{I}(x) \geq \lim_{\delta \to 0+} \inf_{x \in (F^{-1}(A))^{\delta}_{\ell}} \mathcal{I}(x) \geq \lim_{\varepsilon \to 0+} \inf_{y \in A^{\varepsilon}_{\ell_2}} \tilde{\mathcal{J}}(y) = \inf_{y \in \text{cl } A} \tilde{\mathcal{J}}(y) = \inf_{x \in F^{-1}(\text{cl } A)} \mathcal{I}(x).$$

By $\mathrm{cl}(F^{-1}(A)) = F^{-1}(\mathrm{cl}\,A)$, the inequalities above are equalities, thus establishing (2.3). The penultimate infimum is attained at some $y \in \mathrm{cl}\,A$ by tightness of $\tilde{\mathcal{J}}$, hence the last infimum is attained at some $x \in \mathrm{cl}\,B$, as needed. $\qquad\square$

Our last simple claim, analogous to Lemma 4.1.5 in Dembo and Zeitouni (2010), describes the behaviour of metric LDPs under inclusions.

Lemma 3.5. *Let $(Z_n)_{n\geq 1}$ be a sequence of random elements of a metric space $(\mathcal{X}, \ell)$ and $\mathcal{Y} \subset \mathcal{X}$ be a Borel set such that $\mathbb{P}(Z_n \in \mathcal{Y}) = 1$ for each $n \in \mathbb{N}$.*

a) If $(Z_n)_{n\geq 1}$ satisfies a metric LDP in $(\mathcal{X}, \ell)$ with some speed and a rate function $\mathcal{I}$ such that $\mathcal{I} = +\infty$ on $\mathcal{Y}^c$, then $(Z_n)_{n\geq 1}$ satisfies the metric LDP in $(\mathcal{Y}, \ell)$ with the same speed and the rate function $\mathcal{I}|_{\mathcal{Y}}$.

b) Conversely, if $(Z_n)_{n\geq 1}$ satisfies a metric LDP in $(\mathcal{Y}, \ell)$ with some speed and a rate function $\mathcal{I}$, then $(Z_n)_{n\geq 1}$ satisfies the metric LDP in $(\mathcal{X}, \ell)$ with the same speed and the rate function $\mathrm{cl}\,\mathcal{I}$, where $\mathcal{I}$ is extended to $\mathcal{X}$ by putting $\mathcal{I} := +\infty$ on $\mathcal{Y}^c$.

Proof: a) Clearly, $\mathcal{I}|_{\mathcal{Y}}$ is non-negative, lower semi-continuous, and not identically $+\infty$ (otherwise $\mathcal{I} = +\infty$ on $\mathcal{X}$). By monotonicity of probability, the lower bound in a metric LDP follows if we establish it for open sets. Since the topology of $(\mathcal{Y}, \ell)$ is the subspace topology induced from $(\mathcal{X}, \ell)$, for every set $B \subset \mathcal{Y}$ that is open in $(\mathcal{Y}, \ell)$ we have $B = \tilde{B} \cap \mathcal{Y}$ for some $\tilde{B} \subset \mathcal{X}$ open in $(\mathcal{X}, \ell)$. Then, denoting by $(a_n)_{n\geq 1}$ the speed in the metric LDP,

$$- \inf_{y \in B} \mathcal{I}|_{\mathcal{Y}}(y) = - \inf_{x \in \tilde{B}} \mathcal{I}(x) \leq \liminf_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}(Z_n \in \tilde{B}) = \liminf_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}(Z_n \in B), \qquad (3.6)$$

where in the first equality we used that $\mathcal{I} = +\infty$ on $\mathcal{Y}^c$ and in the last one we used that $\mathbb{P}(Z_n \in \mathcal{Y}) = 1$ for each n. This proves the lower bound required. The upper bound required follows from the fact that $\inf_{x \in B_{\mathcal{X}}^\varepsilon} \mathcal{I}(x) = \inf_{y \in B_{\mathcal{Y}}^\varepsilon} \mathcal{I}|_{\mathcal{Y}}(y)$ for every $\varepsilon > 0$.

b) For every $\tilde{B} \subset \mathcal{X}$ open in $(\mathcal{X}, \ell)$, the set $B := \tilde{B} \cap \mathcal{Y}$ is open in $(\mathcal{Y}, \ell)$. Then (3.6) holds true (switch the sides in both equalities), and the lower bound required follows from the equality $\inf_{x \in \tilde{B}} \mathcal{I}(x) = \inf_{x \in \tilde{B}} \mathrm{cl}\,\mathcal{I}(x)$; see (3.3). And the upper bound follows from the fact that $\inf_{y \in B_{\mathcal{Y}}^\varepsilon} \mathcal{I}(y) \geq \inf_{x \in B_{\mathcal{X}}^\varepsilon} \mathrm{cl}\,\mathcal{I}(x)$ for every $\varepsilon > 0$, which holds true by $\mathcal{I} \geq \mathrm{cl}\,\mathcal{I}$. $\qquad\square$

4. The weak-* and related topologies on $BV[0,1]$

In this section we introduce the weak-* topology W_* on the space of functions of bounded variation, then present a convenient metric topology $\widetilde{W_*}$ that coincides with W_* on strongly bounded sets.

4.1. *The weak-* topology and a related metric.* Every $h \in BV[0,1]$ is the distribution function of the $\mathbb{R}^d$-valued finite Borel measure on $[0,1]$, which we denote by dh, that satisfies $dh([0,x]) = h(x)$ for $x \in [0,1]$. As in the case $d = 1$, this correspondence is bijective (Folland 1999, Theorem 3.29). Note that this book considers only complex-valued measures but all the cited results from there are actually valid for any $d \geq 1$ since the consideration of $\mathbb{R}^d$-valued finite measures is coordinate-wise. For example, the integral of a measurable function $f : [0,1] \to \mathbb{R}^d$ w.r.t. dh, is given by

$$\int_0^1 f \cdot dh := \sum_{k=1}^d \int_0^1 f^{(k)} dh^{(k)}, \qquad h \in BV[0,1], \qquad (4.1)$$

with the agreement that the notation above always means integration over $[0,1]$.

Recall that $\mathrm{Var}(h)$ denotes the total variation of an $h \in BV[0,1]$; see (2.11). This is a norm on $BV[0,1]$, and it generates a topology. Both will be referred to as *strong*.

Denote by $C[0,1] = C([0,1]; \mathbb{R}^d)$ the set of continuous functions on $[0,1]$, and equip it with the supremum norm $\|\cdot\|_\infty$. By the Riesz theorem (Folland 1999, Theorem 7.17), the dual of $(C[0,1], \|\cdot\|_\infty)$ is isometrically isomorphic to $(BV[0,1], \mathrm{Var}(\cdot))$ since we regard $BV[0,1]$ as the space of finite $\mathbb{R}^d$-valued Borel measures on $[0,1]$. In particular, we have

$$\mathrm{Var}(h) = \sup_{f \in C[0,1]: \|f\|_\infty \le 1} \int_0^1 f \cdot dh, \qquad h \in BV[0,1], \qquad (4.2)$$

i.e. the strong (total variation) norm is the operator norm. The *weak-* topology* on $BV[0,1]$, denoted by W_*, is the coarsest topology such that all the linear functionals on $BV[0,1]$ of the form $h \mapsto \int_0^1 f \cdot dh$ for $f \in C[0,1]$ are continuous. The convergence defined by W_* is called the weak-* convergence; it is traditionally referred to as weak convergence in probabilistic literature.

We would want to apply our contraction principle (Theorem 3.1) to the natural embedding $(BV[0,1], \rho_1') \to (BV[0,1], W_*)$, but we should seek for a substitute of the weak-* topology W_*, which is known to be non-metrizable. However, it is metrizable on strongly bounded subsets of $BV[0,1]$ (cf. a general metrization result Dunford and Schwartz 1958, Theorem V.5.1) This can be done using an explicit metric ρ_* defined as follows.

Consider the norm

$$\|h\|_* := \int_0^1 |h(s)| ds + |h(1)|, \quad h \in BV[0,1].$$

on $BV[0,1]$, which is simply the L^1-norm of h w.r.t. the sum of the Lebesgue measure on $[0,1]$ and the δ-measure at 1. The metric $\rho_*(g,h) := \|g - h\|_*$ generates a topology, which we denote by $\widetilde{W_*}$. We have the following.

Theorem B (Högnäs 1977). *Suppose that $\{g_\alpha\}_{\alpha \in A} \subset BV[0,1]$ is a strongly bounded net, i.e. $\sup_{\alpha \in A} Var(g_\alpha) < \infty$. Then the following are equivalent:*
1) $\lim_{\alpha \in A} \|g_\alpha\|_* = 0$;
2) $\lim_{\alpha \in A} \int_0^1 f \cdot dg_\alpha = 0$ for any $f \in C[0,1]$, i.e. $\{g_\alpha\}_{\alpha \in A}$ converges weakly-* to zero on $[0,1]$.

By (4.1), this result fully reduces to $d = 1$, the only case considered in Högnäs (1977).

Remark 4.1. If the net $\{g_\alpha\}_{\alpha \in A}$ is a sequence, i.e. $A = \mathbb{N}$, then by the uniform boundedness principle and (4.2), $\sup_{\alpha \in A} \mathrm{Var}(g_\alpha) < \infty$ if and only if $\sup_{\alpha \in A} \left| \int_0^1 f \cdot dg_\alpha \right| < \infty$ for any $f \in C[0,1]$. Hence, a weakly-* convergent sequence also converges in the metric ρ_* (but not vice versa).

Corollary 4.2. *A strongly bounded subset of $BV[0,1]$ that is closed or compact in one of the topologies $\widetilde{W_*}$ and W_*, is closed and compact in each of them.*

This result is the main reason why we have chosen to employ the weak-* topology.

Proof: It follows from the Banach–Alaoglu theorem that strongly bounded sets that are closed in W_* are compact in W_*. Conversely, every set compact in the metric topology $\widetilde{W_*}$ is closed in $\widetilde{W_*}$. The claim then follows from Theorem B because in a topological space, a set is closed if and only if together with any converging net it contains all its limits (Engelking 1989, Corollary 1.6.4), and it is compact if and only if any decreasing sequence of its closed non-empty subsets has non-empty intersection. $\qquad \square$

Let us clarify the relationship between the the topologies $\widetilde{W_*}$ and W_*. Denote by $\mathrm{seq}(W_*)$ the topology on $BV[0,1]$ where a set is closed if and only if it is *sequentially* closed in W_*; we will use this topology in Section 5. We have $W_* \subset \mathrm{seq}(W_*)$ because in any topology, a closed set is sequentially closed. We also have $\widetilde{W_*} \subset \mathrm{seq}(W_*)$ since by Remark 4.1, sequential convergence in W_* implies convergence in $\widetilde{W_*}$. However, this argument *does not* imply that $\widetilde{W_*}$ is weaker than W_* because it is known that $W_* \ne \mathrm{seq}(W_*)$. And indeed, the topologies $\widetilde{W_*}$ and W_* are

incomparable, with $\widetilde{W}_* \not\subset W_*$ following from the observation that for any $f_1, \ldots, f_n \in C[0,1]$, the set $\{h \in AC_0[0,1] : \int_0^1 f \cdot dh = 0, 1 \le i \le n\}$ is unbounded in ρ_*.

4.2. *Comparison with the Skorokhod topologies.* First compare ρ_* with the metrics ρ_2 and ρ_2' defined in Section 2.2.

Lemma 4.3. *For any $h \in BV([0,1]; \mathbb{R}^d)$ and $g \in D([0,1]; \mathbb{R}^d)$, we have*

$$\int_0^1 |g(s) - h(s)| ds \le 2d(\mathrm{Var}(h) - |h(0)| + 1)\rho_2(g, h) + \pi d \rho_2^2(g, h)$$

and

$$\int_0^1 |g(s) - h(s)| ds \le 2d(\mathrm{Var}(h) + 1)\rho_2'(g, h) + \pi d (\rho_2')^2(g, h).$$

Proof: We start by proving the first inequality for $d = 1$. Consider the set

$$U := \big\{(s, x) \in \mathbb{R}^2 : 0 \le s \le 1, g(s) \wedge h(s) \le x \le g(s) \vee h(s)\big\}.$$

It is Borel because g and h are càdlàg on $[0,1]$. We claim that $U \subset \mathrm{cl}\big((\Gamma h)^{\rho_2(g,h)}\big)$, where $(\Gamma h)^r$ denotes the Euclidean open r-neighbourhood of Γh, the completed graph of h. Then by Fubini's theorem,

$$\int_0^1 |g(s) - h(s)| ds = \lambda(U) \le \lambda\big(\mathrm{cl}\big((\Gamma h)^{\rho_2(g,h)}\big)\big). \tag{4.3}$$

where λ denotes the Lebesgue measure on the plane.

In order to prove the claim, pick an $s \in [0,1]$. There is a point $(t, y) \in \Gamma h$ such that $|(t, y) - (s, g(s))| \le \rho_2(g, h)$. Hence a) $(s, x) \in \mathrm{cl}\big((\Gamma h)^{\rho_2(g,h)}\big)$ for any $x \in [g(s) \wedge y, g(s) \vee y]$, and b) since $\Gamma h = \gamma([0,1])$ for a continuous planar curve $\gamma \in \Pi h$ (see Section 2.2), by the intermediate value theorem applied to the spatial coordinate of γ, for any $x \in [h(s) \wedge y, h(s) \vee y]$ there is a $u \in [s \wedge t, s \vee t]$ such that $(u, x) \in \Gamma h$, and by $|s - t| \le \rho_2(g, h)$ this implies $(s, x) \in \mathrm{cl}\big((\Gamma h)^{\rho_2(g,h)}\big)$. Put together, a) and b) imply that $U \subset \mathrm{cl}\big((\Gamma h)^{\rho_2(g,h)}\big)$, as claimed, by

$$[g(s) \wedge h(s), g(s) \vee h(s)] \subset [g(s) \wedge y, g(s) \vee y] \cup [h(s) \wedge y, h(s) \vee y].$$

Furthermore, denote by $\ell(\gamma)$ the length of γ. It is easy to check, using the definition of the total variation of h, that

$$\ell(\gamma) \le \mathrm{Var}(h) - |h(0)| + 1. \tag{4.4}$$

Thus, γ is a rectifiable curve (i.e. offinite length), therefore (Federer 1969, Theorem 3.2.39)

$$\ell(\gamma) = \lim_{r \to 0+} \lambda(\mathrm{cl}((\Gamma h)^r))/(2r), \tag{4.5}$$

where the limit is known as the one-dimensional *Minkowski content* of the set Γh.

On the other hand, for any compact connected planar set F, the function $r \mapsto \lambda(F^r) - \pi r^2$ is known to be concave on $(0, \infty)$ (Fast 1959, Theorem on p. 139 and Sz.-Nagy 1959, Theorem 1). Its right derivative at 0 is $2\ell(\gamma)$ by (4.5). Then, since Γh is connected,

$$\lambda\big(\mathrm{cl}\big((\Gamma h)^r\big)\big) \le 2r\ell(\gamma) + \pi r^2, \qquad r > 0. \tag{4.6}$$

This inequality, which is sometimes referred to as *Steiner's inequality* (cf. Steiner's formula), is actually available in Fast (1959, p. 146); the assumptions imposed in Fast (1959) are satisfied since γ is a rectifiable simple (i.e. injective) curve.

Put together, inequalities (4.3), (4.4), (4.6) imply the first inequality of Lemma 4.3 for $d = 1$. This in turn proves the inequality in any dimension using that $|x| \le |x^{(1)}| + \ldots + |x^{(d)}|$ for $x \in \mathbb{R}^d$, and $\mathrm{Var}(h^{(k)}) - |h^{(k)}(0)| \le \mathrm{Var}(h) - |h(0)|$ and $\rho_2(g^{(k)}, h^{(k)}) \le \rho_2(g, h)$ for $k = 1, \ldots, d$. The last inequality can be obtained from definition (2.5) of the metric ρ_2 as follows: first estimate $|(s, x) - (t, y)| \ge |(s, x^{(k)}) - (t, y^{(k)})|$ and then, since the r.h.s. of this inequality does not depend

on the remaining coordinates, eliminate them from the constraints under the maximum and the minimum in (2.5).

The second inequality of Lemma 4.3 follows by the same argument using that by (2.6), the modified completed graphs $\Gamma'g$ and $\Gamma'h$ can be regarded as the usual completed graphs of the modified functions $g_0 := g\mathbb{1}_{(0,1]}$ and $h_0 := h\mathbb{1}_{(0,1]}$, which are càdlàg on $(0,1]$ and have right limits at 0. Hence $\rho_2'(g,h) = \rho_2(g_0,h_0)$, and if $\gamma \in \Pi'(h)$ is a parametrization of $\Gamma'h$, then $\ell(\gamma) \leq \mathrm{Var}(h)+1$. It remains to use that $\int_0^1 |g(s) - h(s)|ds = \int_0^1 |g_0(s) - h_0(s)|ds$. $\qquad\square$

We now use Lemma 4.3 to clarify the relationship between the topologies introduced. With no risk of confusion, in the rest of the paper we use the original notation M_i, M_i' for the induced topologies on $BV[0,1]$. We have $\widetilde{W}_* \subset M_1'$ since $\rho_2' \leq \rho_1'$ and convergence of càdlàg functions in the metric ρ_1' implies convergence of their values at the endpoint 1.

Neither $\widetilde{W}_*$ nor W_* is comparable with M_2. For example, for $g_n := \mathbb{1}_{[1-1/n,1)}$ and $g := \mathbb{1}_{\{1\}}$, we have $\rho_2(g_n,g) \to 0$ but $\rho_*(g_n,g) \not\to 0$ as $n \to \infty$. However, by Lemma 4.3, convergence in either ρ_2 or ρ_2' implies convergence in ρ_* if the limit function is continuous at 1. Moreover, W_* is incomparable with M_1. For example, for $g_n := \{n\cdot\}/\sqrt{n}$, where $\{\cdot\}$ denotes the fractional part, we have $\rho_1(g_n,0) \to 0$ but g_n does not converge weakly-* since its total variation explodes. Likewise, W_* is incomparable with the topology of uniform convergence, which is weaker than the strong topology on $BV[0,1]$ by inequality (5.2) below.

5. Application of the weak-*-related topologies to the study of I_D

In this section we use the topologies $\widetilde{W}_*$ and $\mathrm{seq}(W_*)$ on $BV[0,1]$, introduced in Section 4, to study properties of the rate function I_D using the results of variational calculus. Namely, we prove *sequential weak-* lower semi-continuity* of I_D and use this property to obtain an explicit *integral representation* for I_D, which is written using the *directional decomposition* of the total variation of functions in $BV([0,1];\mathbb{R}^d)$. Moreover, we standardize the upper bound in the metric LDP of Borovkov and Mogulskii (2013, 2014) for a few types of sets, and prove a standard LDP for trajectories of random walks in the space $(BV[0,1], \rho_*)$.

5.1. *Directional decomposition of total variation.* Recall that $AC_0[0,1]$ denotes the set of coordinate-wise absolutely continuous functions from $[0,1]$ to $\mathbb{R}^d$ that equal 0 at 0. These are exactly the distribution functions of $\mathbb{R}^d$-valued finite Borel absolutely continuous measures on $[0,1]$. For any $h \in BV[0,1]$, put $h_a(t) := \int_0^t h'(s)ds$, where h' exists a.e. and is integrable by Folland (1999, Proposition 3.30), which also ensures that the measure $dh_s := dh - dh_a$ is singular. We say that $h = h_a + h_s$ is the *Lebesgue decomposition* of the vector-valued function h; let us stress that $h_a \in AC_0[0,1]$.

Denote by V^h the total variation *function* of an $h \in BV[0,1]$, defined by $V^h(t) := \mathrm{Var}(h(\cdot \wedge t))$ for $t \in [0,1]$; cf. (2.11). It is non-decreasing, càdlàg, and satisfies $\mathrm{Var}(h) = V^h(1)$, hence $V^h \in BV[0,1]$. By Folland (1999, Theorem 3.29 and Exercise 21 in Section 3.3), dV^h is the total variation *measure* of the vector-valued measure dh, that is the equality

$$dV^h(B) = \sup\left\{\sum_{i=1}^{\infty} |dh(B_i)| : B_1, B_2, \ldots \text{ are disjoint Borel sets}, \bigcup_{i=1}^{\infty} B_i = B\right\} \qquad (5.1)$$

holds true for every Borel set $B \subset [0,1]$. This implies that $V^{h_1+h_2} = V^{h_1} + V^{h_2}$ whenever dh_1 and dh_2 are singular. In particular, we have $V^h = V^{h_a} + V^{h_s}$.

It holds $dh \ll dV^h$ and the Radon–Nykodim density $\dot{h} : [0,1] \to \mathbb{R}^d$, defined by $dh = \dot{h}\, dV^h$, satisfies $|\dot{h}| = 1$ dV^h-a.e. (Folland 1999, Proposition 3.13.b). In particular, this implies that

$$\|h\|_\infty \leq \mathrm{Var}(h), \qquad h \in BV[0,1]. \qquad (5.2)$$

We say that the push-forward measure $d\sigma^h := dV^h \circ (\dot{h})^{-1}$ on the unit sphere $\mathbb{S}^{d-1}$ is the *directional decomposition* of the total variation of h. For example, if $d = 1$, the Hahn–Jordan decomposition gives the unique representation $h = h^+ - h^-$, where $h^\pm \in BV[0,1]$ are non-decreasing functions, and so $d\sigma^h = h^+(1)\delta_1 + h^-(1)\delta_{-1}$. Then $\text{Var}(h) = \sigma^h(\mathbb{S}^{d-1})$ and

$$\text{Var}(h) = \int_0^1 |h'(t)|dt + \sigma^{h_s}(\mathbb{S}^{d-1}). \tag{5.3}$$

To prove (5.3), note that it follows from (5.1) that the measure dV^{h_a} is absolutely continuous (because so is dh_a). Then from the equalities $h'dt = dh_a = \dot{h}_a \, dV^{h_a}$ and $|\dot{h}_a| = 1 \, dV^{h_a}$-a.e., we see by equating densities that $dV^{h_a} = |h'|dt$ (because the unit vector $\dot{h}_a$ multiplied by the scalar density of dV^{h_a} equals h'). This implies (5.3) by $dV^h = dV^{h_a} + dV^{h_s}$.

5.2. *Lower semi-continuity of I_D w.r.t. ρ_* and related results.* Recall that $\mathcal{D}_\mathcal{L}$ is the subset of $\mathbb{R}^d$ where the Laplace transform of X_1 is finite. This set is convex. Denote by

$$I_\infty(v) := \sup\{u \cdot v : u \in \mathcal{D}_\mathcal{L}\}, \qquad v \in \mathbb{R}^d, \tag{5.4}$$

its *support function*. This name reflects that I_∞ equals the so-called *recession function* of I (Rockafellar 1970, Theorem 13.3), which is convex, lower semi-continuous and positively homogeneous on $\mathbb{R}^d$, and has the property (Rockafellar 1970, Theorem 8.5)

$$I_\infty(v) = \lim_{t \to \infty} I(u + vt)/t = \sup_{t > 0} \left[(I(u + vt) - I(u))/t\right], \qquad u \in \mathcal{D}_\mathcal{L}, v \in \mathbb{R}^d. \tag{5.5}$$

Note that in dimension $d = 1$, we have $\text{int}\,\mathcal{D}_\mathcal{L} = (-I_\infty(-1), I_\infty(1))$.

We can now state the main result of the section.

Theorem 5.1. *Assume that X_1 is a random vector in $\mathbb{R}^d$ such that $0 \in \text{int}\,\mathcal{D}_\mathcal{L}$. Then the functional I_D on $BV[0,1]$, defined in (2.12), is tight w.r.t. ρ_*. Moreover, we have*

$$I_D(h) = \int_0^1 I(h'(t))dt + \int_{\mathbb{S}^{d-1}} I_\infty(\ell)\,\sigma^{h_s}(d\ell), \qquad h \in BV[0,1], \tag{5.6}$$

which is dimension $d = 1$ reads as

$$I_D(h) = \int_0^1 I(h'(t))dt + h_s^+(1)I_\infty(1) + h_s^-(1)I_\infty(-1). \tag{5.7}$$

Formula (5.7) is available in Borovkov and Mogulskii (2013, Theorem 3.3). If $\mathcal{D}_\mathcal{L} = \mathbb{R}^d$, that is the Laplace transform of X_1 is finite on $\mathbb{R}^d$, then $I_\infty(v) = +\infty$ for $v \neq 0$, hence $I_D(h) = +\infty$ for $h \notin AC_0[0,1]$.

The advantage of integral representation (5.6) is in its explicitness. It becomes more transparent when compared with equality (5.3), where the total variation of the singular component of a function is expressed using its directional decomposition. We can get (5.3) by formally substituting the Euclidean norm $|\cdot|$ for I in (5.6).

The next two statement are corollaries to Theorem 5.1.

Proposition 5.2. *Assume that X_1 is a random vector in $\mathbb{R}^d$ such that $0 \in \text{int}\,\mathcal{D}_\mathcal{L}$. Then both random sequences $(S_n(\cdot)/n)_{n \geq 1}$ and $(S_{[n\cdot]}/n)_{n \geq 1}$ satisfy the (standard) LDP in the separable metric space $(BV[0,1], \rho_*)$ with the tight convex rate function I_D.*

The main application of this result is our LDP for kernel-weighted sums of i.i.d. random vectors in $\mathbb{R}^d$, presented in Section 6. It is worth to compare Proposition 5.2 with the results by Gantert (1998, Theorems 1 and 2), who proved LDPs in L^1 (i.e., "almost" in the same topology as in our case but on a different space) for one-dimensional random walks with so-called semi-exponential increments.

Proof of Proposition 5.2. This follows from Theorem A combined with Theorem 3.1 applied to the natural embedding $F : (BV[0,1], \rho_1') \to (BV[0,1], \rho_*)$, where $\tilde{\mathcal{J}} = \mathcal{I} = I_D$, and I_D is a tight rate function w.r.t. ρ_* by Theorem 5.1. The assumptions of Theorem 3.1 are satisfied because by Lemma 4.3 and the inequality $\rho_2' \leq \rho_1'$, F is continuous and it is uniformly continuous on strongly bounded subsets of $BV[0,1]$, and also on the sub-level sets of $\mathcal{I}$ by (2.13). Lastly, by Corollary 4.2, the metric space $(\{h : \mathrm{Var}(h) \leq n\}, \rho_*)$ is compact, hence totally bounded, hence separable for every $n \in \mathbb{N}$, therefore $(BV[0,1], \rho_*)$ is separable. $\qquad\square$

Remark 5.3. Note that we cannot apply Theorem 3.3 instead of Theorem 3.1 in the proof presented because the metric space $(BV[0,1], \rho_*)$ is not complete. A way around is to consider the natural embedding of $(BV[0,1], \rho_1)$ into the complete metric space $(L^1[0,1], \rho_*)$.

Namely, for any $g \in L^1[0,1]$, put $I_D(g) := I_D(h)$ if there exists an $h \in BV[0,1]$ such that $\rho_*(g,h) = 0$ and $I_D(g) := +\infty$ otherwise. Let us check that this extended version of I_D remains lower semi-continuous. Since I_D has this property on $(BV[0,1], \rho_*)$ by Theorem 5.1, it suffices to prove that for any $h_1, h_2, \ldots \in BV[0,1]$ and $g \in L^1[0,1] \setminus BV[0,1]$ such that $\lim_{n\to\infty} \rho_*(h_n, g) = 0$ and $\rho_*(g,h) > 0$ for every $h \in BV[0,1]$, we have $\liminf_{n\to\infty} I_D(h_n) = +\infty$. This follows from (2.13) because $\liminf_{n\to\infty} \mathrm{Var}(h_n) = +\infty$ (no subsequence of $(h_n)_n$ is strongly bounded since otherwise by Corollary 4.2 we can choose a further subsubsequence converging in ρ_* to some $h \in BV[0,1]$, hence we arrive at the contradictory $\rho_*(g,h) = 0$).

Theorem 3.3 then implies that $(S_n(\cdot)/n)_{n\geq 1}$ and $(S_{[n\cdot]}/n)_{n\geq 1}$ satisfy the LDP in $(L^1[0,1], \rho_*)$ with the tight rate function I_D; it reduces to the LDP in $(BV[0,1], \rho_*)$ by Dembo and Zeitouni (2010, Lemma 4.1.5.b).

Let us use subscripts to indicate in which metric (or topology) we take closures.

Proposition 5.4. *Assume that X_1 is a random vector in $\mathbb{R}^d$ such that $0 \in \mathrm{int}\,\mathcal{D}_{\mathcal{L}}$. Let $B \subset BV[0,1]$ be such that $\mathrm{cl}_{\rho_1'}(B) = \mathrm{cl}_{\rho_*}(B)$ (where ρ_1' is restricted to $BV[0,1]$). Then*

$$\inf_{h \in \mathrm{cl}_{\rho_1'}(B)} I_D(h) = \lim_{\varepsilon \to 0+} \inf_{h \in B_{\rho_1'}^\varepsilon} I_D(h),$$

with the infimum on the l.h.s. attained at some $h \in \mathrm{cl}_{\rho_1'}(B)$ unless B is empty.

We prove this applying Proposition 3.4 to the natural embedding as in the proof of Proposition 5.2. We thus see that for the sets B that are closed both in ρ_1' and ρ_*, the upper bound in the metric LDP of Theorem A matches the standard LDP one.

Remark 5.5. A set B is closed in ρ_* and in ρ_1' when it is a sub-level set of a functional on $BV[0,1]$ that is lower semi-continuous in both metrics. Examples of such functionals include:

a) The *action* functional J_D defined by the r.h.s. of (5.6) with I replaced by any *convex* lower semi-continuous function $J : \mathbb{R}^d \to [0, +\infty]$ that satisfies $J(u) \geq c_1|u| - c_2$ for some $c_1, c_2 > 0$ and every $u \in \mathbb{R}^d$; for example, from (5.3) we see that $J_{|\cdot|} = \mathrm{Var}$. The action functional J_D is sequentially weak-* lower semi-continuous by Buttazzo (1989, Corollary 3.4.2) applied exactly as in the proof of Theorem 5.1 below. By Theorem B and Remark 4.1, J_D is also lower semi-continuous in ρ_* (and hence in the stronger metric ρ_1') because its sub-level sets are strongly bounded. Indeed, from (5.3) it follows that $J_D(h) \geq c_1 \mathrm{Var}(h) - c_2$ for $h \in BV[0,1]$ since $J_\infty(u) \geq c_1$ for $u \in \mathbb{R}^d$ by (5.5).

b) The *maximum* functional $h \mapsto \sup_{0 \leq t \leq 1}(h(t) \cdot \ell)$, where $\ell \in \mathbb{S}^{d-1}$ is a fixed direction. To check its lower semi-continuity in ρ_* (which suffices since $\widetilde{W}_* \subset M_1'$), note that the value of the functional on an $h \in BV[0,1]$ is either $h(t_0-) \cdot \ell$ or $h(t_0) \cdot \ell$ for some $t_0 \in [0,1]$. This fact, combined with the *càdlàg* property of h and the fact that every ρ_*-convergent sequence contains a subsequence that converges pointwise on a dense subset of $[0,1]$ that contains 1, yields the required property of the functional.

Note in passing that the maximum functional is continuous in ρ_i but not in ρ_i'. However, the *positive maximum* $h \mapsto \sup_{0 \leq t \leq 1}(h(t) \cdot \ell)_+$ is continuous in ρ_i'.

c) For $d = 2$, the *perimeter of the convex hull* (and the *mean width* in higher dimensions) of $h([0,1])$, the image of a planar curve h. Large deviations of the perimeter of the convex hull of the first n steps of a planar random walk were studied by Akopyan and Vysotsky (2021+). For the perimeter functional, lower semi-continuity in ρ_* follows from the combination of Cauchy's formula for perimeter of a planar convex set, the result of Example b), and Fatou's lemma.

The perimeter functional is continuous in ρ_i but not in ρ_i'. However, the perimeter of the convex hull of the set $h([0,1]) \cup \{0\}$ is continuous in ρ_i'.

One more type of sets satisfying the assumption of Proposition 5.4 is as follows.

d) $B_f := \{h \in BV[0,1] : h \leq f\}$ for some $f \in C[0,1]$. This set is closed in ρ_* (and ρ_1') by the same argument as we used in Example b). If $B_f \cap \mathcal{D}_{I_D} \neq \varnothing$, the minimizers of I_D over B exist, and we call them the *taut strings*. In a probabilistic setup, taut strings were considered by Lifshits and Setterqvist (2015), who were interested in those corresponding to the sets of the form $\{h \in BV[0,1] : f_1 \leq h \leq f_2\} = B_{f_1} \cap (-B_{-f_2})$ for $f_1, f_2 \in C[0,1]$.

5.3. *Proof of Theorem 5.1.* We will use results and methods of the calculus of variations, referring to the book by Buttazzo (1989). The action (integral) functionals on the spaces of finite Borel vector-valued measures are considered in Chapter 3 of this book, where the notation $C_0([0,1];\mathbb{R}^d)$ corresponds to our $C[0,1]$; see Buttazzo (1989, Section 3.1). Consider an integral functional I_C defined by $I_C(h) := \int_0^1 I(h')dt$ for $h \in AC_0[0,1]$, and extend it formally to $BV[0,1]$ by putting $I_C(h) := +\infty$ for $h \notin AC_0[0,1]$. The main idea, which applies in a more general setup as described in Section 1.3 of Buttazzo (1989), is that a natural extension is actually given by $\mathrm{cl}_{\mathrm{seq}(W_*)}(I_C)$, referred to as *relaxed functional*. The book offers results which will allow us to find this extension explicitly, and we will show that it equals I_D.

We have $I_C = I_D$ on $AC_0[0,1]$ by Borovkov and Mogulskii (2013, Theorem 5.3), hence

$$I_D(h) \leq I_C(h), \qquad h \in BV[0,1]. \tag{5.8}$$

In the new notation, the definition (2.12) of I_D reads as

$$I_D(h) = \sup_{\mathbf{t} \subset (0,1]: \#\mathbf{t} < \infty} I_C(h^{\mathbf{t}}), \qquad h \in BV[0,1]. \tag{5.9}$$

Then for any dense sequence $(t_n)_{n \geq 1}$ in $(0,1]$, for $\mathbf{t}_n := \{t_1, \ldots, t_n\}$ we have

$$I_D(h) = \lim_{n \to \infty} I_C(h^{\mathbf{t}_n}), \qquad h \in BV[0,1]. \tag{5.10}$$

because I_D is lower semi-continuous w.r.t. ρ_2' by Theorem A and $\rho_2'(h^{\mathbf{t}_n}, h) \to 0$, which follows from the càdlàg property of h. Moreover, $h^{\mathbf{t}_n} \to h$ weakly-*. Indeed, Lemma 4.3 and the equality $h^{\mathbf{t}_n}(1) = h(1)$ imply that $\rho_*(h^{\mathbf{t}_n}, h) \to 0$, and the weak-* convergence then follows from Theorem B, which applies since $\mathrm{Var}(h^{\mathbf{t}_n}) \leq \mathrm{Var}(h)$ by (2.11).

Let us prove lower semi-continuity of I_D in the metric ρ_*. Use that lower semi-continuity in metric spaces is a sequential property. Assume that there are $g, g_1, g_2, \ldots \in BV[0,1]$ such that $\rho_*(g_n, g) \to 0$ but $I_D(g) > \liminf_n I_D(g_n)$ as $n \to \infty$. Since $\rho_*(g_n, g) \to 0$ means convergence in L_1 and $g_n(1) \to g(1)$, by considering a subsequence, we can assume w.l.o.g. that the convergence is point-wise on a subset of $(0,1]$ of full Lebesgue measure. Pick a sequence $(s_n)_{n \geq 1}$ of distinct elements of this set that is dense in $(0,1]$ and satisfies $s_1 = 1$.

For any integer $k \geq 1$, put $\mathbf{s}_k := \{s_1, \ldots, s_k\}$, and let σ_k be the permutation of length k such that $s_{\sigma_k(1)} < \ldots < s_{\sigma_k(k)}$. For any $i, k, n \in \mathbb{N}$ satisfying $1 \leq i \leq k$, we have $g_n^{\mathbf{s}_k}(s_i) = g_n(s_i)$ and also $g_n(s_{\sigma_k(i)}) \to g(s_{\sigma_k(i)})$ as $n \to \infty$. Hence, by lower semi-continuity and non-negativity of I, for

any fixed integer $k \geq 1$ we have

$$I_C(g^{\mathbf{s}_k}) = s_{\sigma_k(1)} I\Big(\frac{g(s_{\sigma_k(1)})}{s_{\sigma_k(1)}}\Big) + \sum_{i=1}^{k-1} (s_{\sigma_k(i+1)} - s_{\sigma_k(i)}) I\Big(\frac{g(s_{\sigma_k(i+1)}) - g(s_{\sigma_k(i)})}{s_{\sigma_k(i+1)} - s_{\sigma_k(i)}}\Big)$$

$$\leq \liminf_{n\to\infty} \Big[s_{\sigma_k(1)} I\Big(\frac{g_n(s_{\sigma_k(1)})}{s_{\sigma_k(1)}}\Big) + \sum_{i=1}^{k-1} (s_{\sigma_k(i+1)} - s_{\sigma_k(i)}) I\Big(\frac{g_n(s_{\sigma_k(i+1)}) - g_n(s_{\sigma_k(i)})}{s_{\sigma_k(i+1)} - s_{\sigma_k(i)}}\Big) \Big]$$

$$= \liminf_{n\to\infty} I_C(g_n^{\mathbf{s}_k}).$$

From (5.9) we see that $I_C(g_n^{\mathbf{s}_k}) \leq I_D(g_n)$, hence $I_C(g^{\mathbf{s}_k}) \leq \liminf_{n\to\infty} I_D(g_n)$. It remains to take $k \to \infty$ and use (5.10) to arrive at $I_D(g) \leq \liminf_{n\to\infty} I_D(g_n)$, which contradicts our assumption that the lower semi-continuity does not hold.

Furthermore, the sub-level sets of I_D are strongly bounded by (2.13). They are closed in the metric ρ_* since I_D is lower semi-continuous in ρ_*. Therefore, they are compact by Corollary 4.2, and thus I_D is tight, as claimed.

It remains to prove integral representation (5.6). Denote by $\mathcal{I}(h)$ its r.h.s. By Buttazzo (1989, Corollar 3.4.2) (which applies because the change of measure formula in the definition of directional total variation σ^h brings to $\mathcal{I}$ into the form Buttazzo 1989, Eq. (3.4.1) and Condition (i) in Buttazzo (1989, Lemma 2.2.3) is satisfied for $z_0 = \mathbb{E}X_1$), the functional $\mathcal{I}$ is sequentially weakly-* lower semi-continuous on $BV[0,1]$. Equivalently, $\mathcal{I}$ is lower semi-continuous in $\mathrm{seq}(W_*)$, the topology where a set is closed if and only if it is sequentially weakly-* closed (Buttazzo 1989, Proposition 1.1.5.ii).

We claim that $\mathcal{I} = \mathrm{cl}_{\mathrm{seq}(W_*)}(I_C)$. It holds $\mathcal{I} \leq \mathrm{cl}_{\mathrm{seq}(W_*)}(I_C)$ because the on r.h.s. we have the maximal functional that is lower semi-continuous in $\mathrm{seq}(W_*)$ and dominated by I_C (Buttazzo 1989, Propositions 1.1.2.ii). Therefore, since $\mathcal{I} = I_C$ on $AC_0[0,1]$, we have $\mathrm{cl}_{\mathrm{seq}(W_*)}(I_C) = I_C$ on $AC_0[0,1]$. On the other hand, by Buttazzo (1989, Theorem 3.3.1), $\mathrm{cl}_{\mathrm{seq}(W_*)}(I_C)$ on $AC_0[0,1]$ is of the form $h \mapsto \int_0^1 J(t, h'(t)) dt$ for some measurable function $J : [0,1] \times \mathbb{R}^d \to [0, +\infty]$ such that $J(t, \cdot)$ is convex and lower semi-continuous for a.e. t. This is possible only when $J(t, \cdot) = I$ for a.e. t by Buttazzo (1989, Proposition 2.1.3) (applied with $\psi = 0$), hence $\mathcal{I} = \mathrm{cl}_{\mathrm{seq}(W_*)}(I_C)$, as claimed.

Furthermore, I_D is lower semi-continuous in $\mathrm{seq}(W_*)$ because I_D is lower semi-continuous in the metric ρ_* and the topology $\widetilde{W}_*$ generated by ρ_* is coarser than $\mathrm{seq}(W_*)$ (see Section 4.1). On the other hand, by lower semi-continuity of $\mathcal{I}$ in $\mathrm{seq}(W_*)$ and equality (5.10), where $I_C(h^{\mathbf{t}_n}) = \mathcal{I}(h^{\mathbf{t}_n})$ and $h^{\mathbf{t}_n} \to h$ weakly-* as $n \to \infty$ (this is explained right after (5.10)), we have $\mathcal{I} \leq I_D$. We also have $I_D \leq I_C$ by (5.8). Therefore, since $\mathcal{I} = \mathrm{cl}_{\mathrm{seq}(W_*)}(I_C)$ and $\mathrm{cl}_{\mathrm{seq}(W_*)}(I_C)$ is the maximal functional that is lower semi-continuous in $\mathrm{seq}(W_*)$ and dominated by I_C, we have $\mathcal{I} = I_D$, as required.

6. Cramér's theorem for kernel-weighted sums

In this section we present an application of Proposition 5.2, yielding an LDP for *kernel-weighted sums* of i.i.d. random vectors in $\mathbb{R}^d$. This extends the classical Cramér theorem.

It appears that large deviations of weighted i.i.d. random variables were first studied by Book (1972). The next results were due to Kiesel and Stadtmüller (2000), who considered only the "light-tailed" case where the i.i.d. terms have finite Laplace transform, i.e. $\mathcal{D}_{\mathcal{L}} = \mathbb{R}$. They proved an LDP and found the rate function (available in Kiesel and Stadtmüller, 2000, Theorem on p. 933 with the coefficients a_ν for the kernel weights given in Kiesel and Stadtmüller 1996, p. 976); below we will present a more explicit expression (6.3) for the rate function. A recent result Gantert et al. (2014, Theorem 3) is an LDP (with polynomial speed and an explicit rate function) for kernel-weighted sums of i.i.d. random variables with stretched (super-) exponential tails. In this "heavy-tailed" case where the random variables have no exponential moments, the rate function is defined by the supremum of the (non-negative) kernel. This corresponds to the first two terms in our formula (6.3) (cf. Remark 6.2.c). Since these terms vanish for "light-tailed" increments, in the case $0 \in \mathrm{int}\, \mathcal{D}_{\mathcal{L}}$

with $\mathcal{D}_{\mathcal{L}} \neq \mathbb{R}$ there is a natural transition of the rate function from the "light-tailed" case to the "heavy-tailed" one. One more related reference, added in proof, is Gantert et al. (2016).

We introduce more notation to state our result. Let us agree to write $\sup g$, $\max g$, etc. for the supremum, maximum, etc. of a real-valued function g over its (effective) domain. For any real-valued increasing function g on an interval (a, b), where $-\infty \leq a < b \leq +\infty$, denote by $\bar{g}$ its extension to $\mathbb{R}$ given by $\bar{g}(x) := g(b-)$ for $x \geq b$ and $\bar{g}(x) := g(a+)$ for $x \leq a$. For a real x, put $x_+ := \max(x, 0)$ and $x_- := (-x)_+$, and use the same notation for functions. By convention, put $\frac{1}{0} := +\infty$ and $\frac{C}{0} := \mathbb{R}^d$ for any set $C \subset \mathbb{R}^d$ satisfying $0 \in \mathrm{int}\, C$. Recall that $K(u) = \log \mathcal{L}$ is the cumulant generating function of X_1, finite on its effective domain $\mathcal{D}_{\mathcal{L}}$.

Theorem 6.1. *Assume that X_1 is a random vector in $\mathbb{R}^d$ such that $0 \in \mathrm{int}\, \mathcal{D}_{\mathcal{L}}$ and $f : [0, 1] \to \mathbb{R}$ is a non-zero Lipschitz function. Then the sequence of random vectors $\left(\frac{1}{n}\sum_{k=1}^n f(\frac{k}{n})X_k\right)_{n \geq 1}$ satisfies the LDP in $\mathbb{R}^d$ with the tight rate function I_f that is the Legendre–Fenchel transform of the convex function*

$$E_f(\lambda) := \int_0^1 K(\lambda f(t))dt, \quad \lambda \in \mathbb{R}^d. \tag{6.1}$$

Moreover, if $\dim(\mathrm{supp}(X_1)) = d$, the rate function satisfies

$$I_f(x) = \int_0^1 I(\nabla K((\nabla E_f)^{-1}(x)f(t)))dt, \qquad x \in \nabla E_f(\mathrm{int}\, D_f), \tag{6.2}$$

where $D_f := \frac{\mathcal{D}_{\mathcal{L}}}{\max f_+} \cap \frac{-\mathcal{D}_{\mathcal{L}}}{\max f_-}$ and ∇E_f is an injective function (on its domain $\mathrm{int}\, D_f$). For $d = 1$, equality (6.2) extends to

$$I_f(x) = M_+(x - \sup E_f')_+ + M_-(x - \inf E_f')_- + \int_0^1 I\left(K'\left(\overline{(E_f')^{-1}}(x)f(t)\right)\right)dt, \quad x \in \mathbb{R}, \tag{6.3}$$

where $M_\pm := \min\left(\frac{I_\infty(1)}{\max f_\pm}, \frac{I_\infty(-1)}{\max f_\mp}\right)$ and $K'(\pm I_\infty(\pm 1)) := K'(\pm I_\infty(\pm 1)\mp)$, with the symbol $\mp$ standing for the left/right limit.

Remark 6.2. Let us make some comments.

a) Equality (6.2) remains valid when $\dim(\mathrm{supp}(X_1)) < d$, in which case the integrand on the r.h.s. still is a well-defined function, calculated by taking any element of the set $(\nabla E_f)^{-1}(x)$. We will prove this together with the main case where $\mathrm{supp}(X_1)$ has full dimension.

b) If $\mathcal{D}_{\mathcal{L}} = \mathbb{R}^d$ and $\dim(\mathrm{supp}(X_1)) = d$, then $\mathrm{int}(\mathcal{D}_{I_f}) = \nabla E_f(\mathrm{int}\, D_f)$, hence in this case equality (6.2) completely defines I_f by lower semi-continuity; and if we additionally assume that $\mathrm{conv}(\mathrm{supp}(X_1)) = \mathbb{R}^d$, then $\nabla E_f(\mathrm{int}\, D_f) = \mathbb{R}^d$ and if $d = 1$, the first two terms in (6.3) vanish. We will prove the claims of Items b) and c) below after proving Theorem 6.1.

c) Assume that $\dim(\mathrm{supp}(X_1)) = d$. Then for any $x \in \nabla E_f(\mathrm{int}\, D_f)$, the function $h^{(x)}(t) := \int_0^t \nabla K((\nabla E_f)^{-1}(x)f(s))ds$ is the unique minimizer of I_D on $\{h \in BV[0,1] : \int_0^1 f dh = x\}$. If we additionally assume that $d = 1$, $M_+ < \infty$, $f \geq 0$ for simplicity, and $\mathrm{argmax}_{t \in [0,1]} f(t)$ has Lebesgue measure zero to fully distinguish from the usual case $f \equiv 1$ of equal weights, then for (say) any $x \geq \sup E_f'$, all the minimizers of I_D are of the form

$$h(t) = \frac{x - \sup E_f'}{\max f_+}g(t) + \int_0^t K'(M_+ f(s))ds, \qquad t \in [0, 1],$$

where $g \in BV[0, 1]$ is a non-decreasing function such that $g(0) = 0$, $g(1) = 1$, and dg is supported on $\mathrm{argmax}_t f(t)$. For example, if $\mathrm{argmax}_t f(t)$ has a unique element t_0, the only possible g is $\mathbb{1}_{[t_0, 1]}$. The singular part of such h provides the first term on the r.h.s. of (6.3).

Also note that in the case when $\mathcal{D}_{\mathcal{L}} = \mathbb{R}^d$, the function $h^{(x)}$ is the unique solution to the *Euler–Lagrange equation* for the Lagrangian $L(t, p) := I(p) + f(t)\lambda \cdot p$, where $t \in [0, 1]$ and $p \in \mathbb{R}^d$, but

the corresponding results of classical variational calculus (see, e.g., Cesari 1983, Sections 2.2 and 2.7) require the additional assumptions $f \in C^1[0,1]$ and $\mathcal{D}_I = \mathbb{R}^d$; the latter one is equivalent to $\operatorname{conv}(\operatorname{supp}(X_1)) = \mathbb{R}^d$.

 d) Let us give a concrete example. Assume that X_1 is a non-generate Gaussian vector in $\mathbb{R}^d$. We have $K(v) = \frac{1}{2}v^\top \Sigma v + \mu \cdot v$ and $I(v) = \frac{1}{2}(v - \mu)^\top \Sigma^{-1}(v - \mu)$ for $v \in \mathbb{R}^d$, where $\mu := \mathbb{E}X_1$ and $\Sigma := \mathbb{E}(X_1 X_1^\top) - \mu\mu^\top$. Then $\nabla K(v) = \Sigma v + \mu$ and $\nabla E_f(\lambda) = m_1\mu + m_2\Sigma\lambda$ for $\lambda \in \mathbb{R}^d$, where $m_i := \int_0^1 f^i(s)ds$ for $i \in \{1, 2\}$, hence $h^{(x)}(t) = t\mu + m_2^{-1}(x - m_1\mu)F(t)$ for $x \in \mathbb{R}^d$, where $F(t) := \int_0^t f(s)ds$. This gives $I_f(x) = \frac{1}{2}m_2^{-1}(x - m_1\mu)^\top\Sigma^{-1}(x - m_1\mu)$.

Proof of Theorem 6.1: We use the representation

$$\frac{1}{n}\sum_{k=1}^n f\left(\frac{k}{n}\right)X_k = \int_0^1 fd\left(\frac{1}{n}S_{[n\cdot]}\right)$$

for the kernel-weighted sums. Any Lipschitz function is absolutely continuous, and the integration by parts formula (Folland 1999, Theorem 3.36) yields $\int_0^1 fdh = f(1)h(1) - \int_0^1 f'hdt$ for $h \in BV[0,1]$. Hence the functional $h \mapsto \int_0^1 fdh$ on $BV[0,1]$ is continuous in the metric ρ_* since f' is bounded and $L^\infty[0,1]$ is dual to $L^1[0,1]$. Therefore, by the usual contraction principle (Dembo and Zeitouni 2010, Theorem 4.2.1) it follows from Proposition 5.2 that the sequence of kernel-weighted sums satisfies the LDP in $\mathbb{R}^d$ with the tight rate function

$$I_f(x) = \inf_{h\in BV([0,1];\mathbb{R}^d):\int_0^1 fdh=x} I_D(h), \qquad x \in \mathbb{R}^d, \tag{6.4}$$

where the infimum is always attained at some h. Let us *compute* this function.

 Denote by χ_C the *convex-analytic* characteristic function of a set $C \subset \mathbb{R}^d$, defined to be 0 on the set and $+\infty$ on its complement. By definition (2.10) of the Legendre–Fenchel transform, which we denote by *, for any $\lambda \in \mathbb{R}^d$,

$$I_f^*(\lambda) = \sup_{x\in\mathbb{R}^d}\left(\lambda \cdot x - I_f(x)\right) = \sup_{x\in\mathbb{R}^d}\left(\lambda \cdot x - \inf_{h\in BV}\left[I_D(h) + \chi_{\{x\}}\left(\int_0^1 fdh\right)\right]\right)$$

$$= \sup_{x\in\mathbb{R}^d}\sup_{h\in BV}\left(\lambda \cdot x - I_D(h) - \chi_{\{x\}}\left(\int_0^1 fdh\right)\right)$$

$$= \sup_{h\in BV}\left(\int_0^1 \lambda f \cdot dh - I_D(h)\right),$$

where the last equality follows after interchanging the suprema. Thus, $I_f^*(\lambda) = I_D^*(\lambda f)$, where I_D^* is the Legendre–Fenchel transform of the function I_D on BV, defined by the standard duality (given by the respective integral) between the spaces $BV[0,1]$ and $C[0,1]$.

 By representation (5.6), we have

$$I_f^*(\lambda) = \sup_{h\in AC_0}\left(\int_0^1 \lambda f \cdot h'dt - \int_0^1 I(h')dt\right) + \sup_{h\in BV: h_a=0}\left(\int_0^1 \lambda f \cdot dh - \int_{\mathbb{S}^{d-1}} I_\infty(\ell)\sigma^h(d\ell)\right).$$

To find the first supremum, we use Proposition IX.2.1 in the book by Ekeland and Témam (1999), which computes the Legendre–Fenchel transform of the functional $g \mapsto \int_0^1 I(g(t))dt$ on $L^1([0,1];\mathbb{R}^d)$, defined by the standard duality (given by the respective integral) between the spaces $L^1[0,1]$ and $L^\infty[0,1]$. This result applies, in the notation and terminology of Ekeland and Témam (1999), with $\alpha = 1$, $u_0 \equiv \mathbb{E}X_1$, and the integrand $f = I$, which is non-negative and normal (i.e., lower semi-continuous) on $B = \mathbb{R}^d$. This gives, by $I^* = K$,

$$I_f^*(\lambda) = \int_0^1 K(\lambda f(t))dt + \sup_{h\in BV: h_a=0}\left(\int_0^1 \lambda f \cdot dh - \int_{\mathbb{S}^{d-1}} I_\infty(\ell)\sigma^h(d\ell)\right), \tag{6.5}$$

where the first term is $E_f(\lambda)$.

Consider the second term in (6.5). For $h \in BV[0,1]$, put $h^{\pm}(t) := dh([0,t] \cap \{\pm f > 0\})$ for $t \in [0,1]$ and $h^= := h - h^+ - h^-$. We have $h^{\pm}, h^= \in BV[0,1]$. It follows from (5.1) that $V^h = V^{h^+} + V^{h^-} + V^{h^=}$, hence $\sigma^h = \sigma^{h^+} + \sigma^{h^-} + \sigma^{h^=}$ by the definition of directional total variation. Therefore,

$$
\begin{aligned}
\int_0^1 \lambda f \cdot dh - \int_{\mathbb{S}^{d-1}} I_\infty(\ell)\sigma^h(d\ell) &= \sum_{\varsigma \in \{+,-,=\}} \left[\int_0^1 \lambda f \cdot \dot{h}^\varsigma \, dV^{h^\varsigma} - \int_{\mathbb{S}^{d-1}} I_\infty(\ell)\sigma^{h^\varsigma}(d\ell) \right] \\
&\leq \sum_{\varsigma \in \{+,-\}} \left[\int_0^1 (\varsigma \max f_\varsigma \cdot \lambda \cdot \dot{h}^\varsigma)_+ \, dV^{h^\varsigma} - \int_{\mathbb{S}^{d-1}} I_\infty(\ell)\sigma^{h^\varsigma}(d\ell) \right] \\
&= \sum_{\varsigma \in \{+,-\}} \left[\int_{\mathbb{S}^{d-1}} \big((\varsigma \max f_\varsigma \cdot \lambda \cdot \ell)_+ - I_\infty(\ell)\big)\sigma^{h^\varsigma}(d\ell) \right], \qquad (6.6)
\end{aligned}
$$

where the inequality follows from $0 \leq \pm f \leq \max f_{\pm}$ $(dV^{h^{\pm}})$-a.e. We estimate the integrands using that $I_\infty(\ell) = \sup_{u \in \mathcal{D}_\mathcal{L}} u \cdot \ell = \sup_{u \in \mathrm{cl}\,\mathcal{D}_\mathcal{L}} u \cdot \ell$, and recall that $D_f = \frac{\mathcal{D}_\mathcal{L}}{\max f_+} \cap \frac{-\mathcal{D}_\mathcal{L}}{\max f_-}$. Then

$$
\int_0^1 \lambda f \cdot dh - \int_{\mathbb{S}^{d-1}} I_\infty(\ell)\sigma^h(d\ell) \leq \chi_{\mathrm{cl}\,\mathcal{D}_\mathcal{L}}\big(\max f_+ \lambda\big) + \chi_{\mathrm{cl}\,\mathcal{D}_\mathcal{L}}\big(-\max f_- \lambda\big) = \chi_{\mathrm{cl}\,D_f}(\lambda). \qquad (6.7)
$$

By (6.5) and (6.7), we have $I_f^*(\lambda) = E_f(\lambda)$ for $\lambda \in \mathrm{cl}\,D_f$. On the other hand, $I_f^*(\lambda) \geq E_f(\lambda) = +\infty$ for $\lambda \notin \mathrm{cl}\,D_f$ because for such λ, $K(\lambda f(t)) = +\infty$ for t in a non-empty interval since f is continuous on $[0,1]$ and $\mathcal{D}_K = \mathcal{D}_\mathcal{L}$. All together, we get $I_f^* = E_f$.

The function E_f is convex on $\mathbb{R}^d$ as a mixture of convex functions $\lambda \mapsto K(\lambda f(t))$. Since K is lower semi-continuous by Fatou's lemma, so is E_f, again by Fatou's lemma. Therefore, $I_f^* = E_f$ yields the required identity $I_f = E_f^*$ by Rockafellar (1970, Theorem 12.2 and Corollary 12.1.1).

Next we prove formula (6.2) for I_f. By Theorem 26.4 in Rockafellar (1970), which applies because E_f is a continuous convex function differentiable on $\mathrm{int}\,D_f$ (since so is K on $\mathrm{int}\,\mathcal{D}_\mathcal{L}$), we have

$$
I_f(x) = x \cdot (\nabla E_f)^{-1}(x) - E_f((\nabla E_f)^{-1}(x)), \quad x \in \nabla E_f(\mathrm{int}\,D_f).
$$

The result used also states that the r.h.s. is well-defined even if $(\nabla E_f)^{-1}(x)$ contains more than one element, in which case we shall understand the r.h.s. replacing $(\nabla E_f)^{-1}(x)$ by any $\lambda \in (\nabla E_f)^{-1}(x)$, and the resulting value does not depend on the particular choice of λ. This justifies Remark 6.2.a.

For any $x \in \nabla E_f(\mathrm{int}\,D_f)$ and $\lambda \in (\nabla E_f)^{-1}(x)$, we have

$$
x = \nabla E_f(\lambda) = \nabla\left(\int_0^1 K(\lambda f(t))dt\right) = \int_0^1 \nabla K(\lambda f(t))f(t)dt,
$$

because the cumulant generating function K is smooth on $\mathrm{int}\,\mathcal{D}_\mathcal{L}$ and $\lambda f(t) \in \mathrm{int}\,\mathcal{D}_\mathcal{L}$ for every $t \in [0,1]$ by $\lambda \in (\nabla E_f)^{-1}(x) \subset \mathrm{int}\,D_f$. Then

$$
\begin{aligned}
I_f(x) &= x \cdot \lambda - \int_0^1 K(\lambda f(t))dt \\
&= \int_0^1 \big(\nabla K(\lambda f(t)) \cdot \lambda f(t) - K(\lambda f(t))\big)dt = \int_0^1 I(\nabla K(\lambda f(t)))dt,
\end{aligned}
$$

where in the last equality we applied Rockafellar (1970, Theorem 26.4) again. This proves (6.2).

Furthermore, $\dim(\mathrm{supp}(X_1)) < d$ if and only if X_1 is supported on a hyperplane, in which case K, and hence E_f, is constant along the lines orthogonal to the hyperplane. Therefore ∇E_f cannot be injective in this case. On the contrary, if $\dim(\mathrm{supp}(X_1)) = d$, then it follows by a standard application of the Cauchy–Schwartz inequality that the Hessian of K is positive-definite on the interior of its effective domain. Then the same holds for E_f, which is a mixture of functions

$\lambda \mapsto K(\lambda f(t))$ which are positively definite when $f(t) \neq 0$. This implies that ∇E_f is injective, as claimed, since the scalar function $t \mapsto \nabla E_f((1-t)\lambda_1 + t\lambda_2) \cdot (\lambda_2 - \lambda_1)$ for $t \in [0,1]$ has strictly positive derivative whenever $\lambda_1, \lambda_2 \in \mathbb{R}^d$ are distinct.

It remains to prove equality (6.3). Here $d = 1$ and $\operatorname{int} D_f = (-M_-, M_+)$. For $x \in E_f'(\operatorname{int} D_f)$, (6.3) reduces to equality (6.2) and there is nothing to prove. Since K is continuously differentiable and convex on $\operatorname{int} \mathcal{D}_{\mathcal{L}}$, so is E_f on $\operatorname{int} D_f$. Therefore $E_f'(\operatorname{int} D_f) = (\inf E_f', \sup E_f')$. Assume that the complement of this set is non-empty and consider an x from there. W.l.o.g., we can assume that $\sup E_f' < \infty$ and prove (6.3) only on $[\sup E_f', \infty)$. Also, assume that X_1 is not constant, otherwise the claim is trivial.

We can check that equality (6.3) holds true for $x = \sup E_f'$ by taking $x \nearrow \sup E_f'$ in (6.3) and using lower semi-continuity of I combined with the facts that I_f increases on $[m_1\mu, +\infty)$ and $\sup E_f' > E_f'(0) = m_1\mu$, where $m_1 = \int_0^1 f(t)dt$ and $\mu = \mathbb{E}X_1$. Therefore it remains to prove (6.3) for $x > \sup E_f'$. We have $I_f(x) = E_f^*(x) = \sup_{\lambda \in \mathbb{R}^d}(x \cdot \lambda - E_f(\lambda))$. If $M_+ = \infty$, this gives $I_f(x) = \infty$ for $x > \sup E_f'$, matching (6.3). If $M_+ < \infty$, this gives, by taking into account that $E_f(\lambda) = +\infty$ for $\lambda > M_+$, that $I_f(x) = M_+ x - E_f(M_+)$ for $x > \sup E_f'$. By the equality $I(K'(u)) = uK'(u) - K(u)$ for $u \in \operatorname{cl}(-M_-, M_+)$, we get

$$I_f(x) = M_+ x - M_+ \int_0^1 f(t)K'(M_+ f(t))dt + \int_0^1 \left(M_+ f(t) \cdot K'(M_+ f(t)) - K(M_+ f(t))\right)dt$$

$$= M_+(x - E_f'(M_+-)) + \int_0^1 I(K'(M_+ f(t))dt,$$

which coincides with (6.3) for $x > \sup E_f'$ since $\sup E_f' = E_f'(M_+-)$ and $\overline{(E_f')^{-1}}(x) = M_+$. $\qquad\square$

Proof of Remark 6.2: b) If $\dim(\operatorname{supp}(X_1)) = d$, then a standard application of Hölder's inequality implies strict convexity of K on $\mathcal{D}_{\mathcal{L}}$. So is E_f on its effective domain $\operatorname{int} D_f$. If we additionally require that $\mathcal{D}_{\mathcal{L}} = \mathbb{R}^d$, then $D_f = \mathbb{R}^d$ and hence $\operatorname{int}(\mathcal{D}_{I_f}) = \nabla E_f(\operatorname{int} D_f)$ by Theorem 26.5 in Rockafellar (1970). If we further assume that $\operatorname{conv}(\operatorname{supp}(X_1)) = \mathbb{R}^d$, then $\operatorname{conv}(\operatorname{supp}(u \cdot X_1)) = \mathbb{R}$ for any non-zero $u \in \mathbb{R}^d$. This readily implies $\lim_{t \to \infty} K(tu)/t = +\infty$. Therefore Theorem 26.6 from Rockafellar (1970) applies, ensuring that ∇E_f is a homeomorphism from $\mathbb{R}^d$ to $\mathbb{R}^d$.

c) Fix any $x \in \mathbb{R}^d$ and $h \in BV[0,1]$ such that $x = \int_0^1 fdh$. Put $x_a := \int_0^1 fdh_a$ and $x_s := \int_0^1 fdh_s$. For any $\lambda \in \mathbb{R}^d$, by Fenchel's inequality we have

$$I(h'(t)) + K(\lambda f(t)) \geq \lambda f(t) \cdot h'(t), \quad \text{a.e. } t \in [0,1], \tag{6.8}$$

hence $I_D(h_a) \geq \lambda x_a - E_f(\lambda)$. Combining this with inequality (6.7) and optimizing over $\lambda \in \operatorname{cl} D_f$ yields

$$I_D(h) \geq \sup_{\lambda \in \operatorname{cl} D_f} \left(\lambda x - E_f(\lambda)\right) = I_f(x), \tag{6.9}$$

where the equality follows from the facts that $I_f = E_f^*$ and $\mathcal{D}_{E_f} \subset \operatorname{cl} D_f$.

If $x \in \nabla E_f(\operatorname{int} D_f)$, the supremum in (6.9) is attained at $\lambda = (\nabla E_f)^{-1}(x)$; recall that ∇E_f is injective due to the assumption $\dim(\operatorname{supp}(X_1)) = d$. Therefore, $I_D(h) = I_f(x)$ implies that (6.7) is an equality and (6.8) is an a.e. equality for this particular λ. Since $\lambda \in \operatorname{int} D_f$, we have $\lambda f(t) \in \operatorname{int} \mathcal{D}_K$ for every $t \in [0,1]$, hence Fenchel's inequality (6.8) is an equality if and only if $h'(t) = \nabla K(\lambda f(t))$ for a.e. t (Rockafellar 1970, Theorem 23.5). Also, since $\max f_\pm \cdot \lambda \in \operatorname{int} \mathcal{D}_{\mathcal{L}}$, inequality (6.7) is strict when $h_s \neq 0$ because the integrands in the last line of (6.6) are strictly negative. All together, this means that $h(t) = \int_0^t \nabla K((\nabla E_f)^{-1}(x)f(s))ds$ is the unique minimizer of I_D, i.e. $h = h^{(x)}$, as claimed.

If $d = 1$, $\sup E_f' < \infty$, and $M_+ < \infty$, the implication $(I_D(h) = I_f(x)) \Rightarrow (h = h^{(x)})$ extends to $x = \sup E_f'$ (corresponding to $\lambda = M_+$) by continuity as in the corresponding argument in the proof

of Theorem 6.1. For $\lambda = M_+$ and when $\text{argmax}_t f(t)$ is a singular set, Fenchel's inequality (6.8) still is an equality if and only if $h'(t) = \nabla K(M_+ f(t))$ for a.e. t because $M_+ f(t) \in \text{int } \mathcal{D}_K$ for a.e. t. However, for $\lambda = M_+$ and when $f \geq 0$, the last line of (6.6) is zero, and thus (6.6) is an equality if and only if dh_s is a non-negative finite measure supported on $\text{argmax}_t f(t)$. Thus, for $x > \sup E'_f$, the equality $I_D(h) = I_f(x)$ implies that $h_a = h^{(\sup E'_f)}$ and $h_s(1) = \frac{x - \sup E'_f}{\max f_+}$, as claimed. $\qquad\square$

Acknowledgements

References

Akopyan, A. and Vysotsky, V. Large deviations of convex hulls of planar random walks and Brownian motions (2021+). To appear in *Ann. H. Lebesgue*.

Bazhba, M., Blanchet, J., Rhee, C.-H., and Zwart, B. Sample path large deviations for Lévy processes and random walks with Weibull increments. *Ann. Appl. Probab.*, **30** (6), 2695–2739 (2020). MR4187125.

Billingsley, P. *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney (1968). MR0233396.

Book, S. A. Large deviation probabilities for weighted sums. *Ann. Math. Statist.*, **43**, 1221–1234 (1972). MR331486.

Borovkov, A. A. and Mogulskii, A. A. On large deviation principles in metric spaces. *Sib. Math. J.*, **51** (6), 989–1003 (2010). MR2797595.

Borovkov, A. A. and Mogulskii, A. A. On large deviation principles for the trajectories of random walks. I. *Theory Probab. Appl.*, **56** (4), 538–561 (2012). MR3137061.

Borovkov, A. A. and Mogulskii, A. A. Large deviation principles for random walk trajectories. II. *Theory Probab. Appl.*, **57** (1), 1–27 (2013). MR3201636.

Borovkov, A. A. and Mogulskii, A. A. Large deviation principles for random walk trajectories. III. *Theory Probab. Appl.*, **58** (1), 25–37 (2014). MR3267282.

Buttazzo, G. *Semicontinuity, relaxation and integral representation in the calculus of variations*, volume 207 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York (1989). ISBN 0-582-01859-5. MR1020296.

Cesari, L. *Optimization—theory and applications. Problems with ordinary differential equations*, volume 17 of *Applications of Mathematics (New York)*. Springer-Verlag, New York (1983). ISBN 0-387-90676-2. MR688142.

Dembo, A. and Zeitouni, O. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin (2010). ISBN 978-3-642-03310-0. MR2571413.

Dunford, N. and Schwartz, J. T. *Linear Operators. I. General Theory*. Pure and Applied Mathematics, Vol. 7. Interscience Publishers, Inc., New York; Interscience Publishers, Ltd., London (1958). MR0117523.

Ekeland, I. and Témam, R. *Convex analysis and variational problems*, volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, English edition (1999). ISBN 0-89871-450-8. MR1727362.

Engelking, R. *General topology*, volume 6 of *Sigma Series in Pure Mathematics*. Heldermann Verlag, Berlin, second edition (1989). ISBN 3-88538-006-4. MR1039321.

Fast, G. Area of a generalized circle as a function of its radius. I, II. *Fund. Math.*, **46**, 137–163 (1959). MR104204.

Federer, H. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York (1969). MR0257325.

Folland, G. B. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition (1999). ISBN 0-471-31716-0. MR1681462.

Gantert, N. Functional Erdős-Renyi laws for semiexponential random variables. *Ann. Probab.*, **26** (3), 1356–1369 (1998). MR1640348.

Gantert, N., Kim, S. S., and Ramanan, K. Cramér's theorem is atypical. In *Advances in the mathematical sciences*, volume 6 of *Assoc. Women Math. Ser.*, pp. 253–270. Springer (2016). MR3654500.

Gantert, N., Ramanan, K., and Rembart, F. Large deviations for weighted sums of stretched exponential random variables. *Electron. Commun. Probab.*, **19**, no. 41, pp. 14 (2014). MR3233203.

Högnäs, G. Characterization of weak convergence of signed measures on $[0, 1]$. *Math. Scand.*, **41** (1), 175–184 (1977). MR482909.

Kiesel, R. and Stadtmüller, U. Erdős-Rényi-Shepp laws and weighted sums of independent identically distributed random variables. *J. Theoret. Probab.*, **9** (4), 961–982 (1996). MR1419871.

Kiesel, R. and Stadtmüller, U. A large deviation principle for weighted sums of independent identically distributed random variables. *J. Math. Anal. Appl.*, **251** (2), 929–939 (2000). MR1794779.

Lifshits, M. and Setterqvist, E. Energy of taut strings accompanying Wiener process. *Stochastic Process. Appl.*, **125** (2), 401–427 (2015). MR3293288.

Mogulskii, A. A. Large deviations for the trajectories of multidimensional random walks. *Theor. Probab. Appl.*, **21** (2), 300–315 (1976). MR0420798.

Najim, J. A Cramér type theorem for weighted random variables. *Electron. J. Probab.*, **7**, no. 4, pp. 32 (2002). MR1887624.

Rassoul-Agha, F. and Seppäläinen, T. *A course on large deviations with an introduction to Gibbs measures*, volume 162 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI (2015). ISBN 978-0-8218-7578-0. MR3309619.

Rockafellar, R. T. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J. (1970). MR0274683.

Sz.-Nagy, B. Über Parallelmengen nichtkonvexer ebener Bereiche. *Acta Sci. Math. (Szeged)*, **20**, 36–47 (1959). MR107213.

Whitt, W. *Stochastic-process limits. An introduction to stochastic-process limits and their application to queues*. Springer Series in Operations Research. Springer-Verlag, New York (2002). ISBN 0-387-95358-2. MR1876437.