Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 117 (2007) 1373–1403
stochastic processes and their applications
www.elsevier.com/locate/spa

Large deviations for weighted empirical mean with outliers

M. Maïda[a], J. Najim[b,*], S. Péché[c]

[a] *Université de Paris-Sud, Equipe "Probabilités-Statistiques", bâtiment 425, 91405 Orsay Cedex, France*
[b] *CNRS, Télécom Paris, 46, rue Barrault, 75013 Paris, France*
[c] *Institut Fourier, 100 rue des Maths, BP 74, 38402 St Martin d'Heres, France*

Received 10 May 2006; received in revised form 30 January 2007; accepted 4 February 2007
Available online 14 February 2007

Abstract

We study in this article the large deviations for the weighted empirical mean $L_n = \frac{1}{n}\sum_1^n \mathbf{f}(x_i^n) \cdot Z_i$, where $(Z_i)_{i\in\mathbb{N}}$ is a sequence of $\mathbb{R}^d$-valued independent and identically distributed random variables with some exponential moments and where the deterministic weights $\mathbf{f}(x_i^n)$ are $m \times d$ matrices. Here $\mathbf{f}$ is a continuous application defined on a locally compact metric space $(\mathcal{X}, \rho)$ and we assume that the empirical measure $\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}$ weakly converges to some probability distribution R with compact support $\mathcal{Y}$.

The scope of this paper is to study the effect on the Large Deviation Principle (LDP) of *outliers*, that is elements $x_{i(n)}^n \in \{x_i^n, 1 \le i \le n\}$ such that

$$\liminf_{n\to\infty} \rho(x_{i(n)}^n, \mathcal{Y}) > 0.$$

We show that outliers can have a dramatic impact on the rate function driving the LDP for L_n. We also show that the statement of a LDP in this case requires specific assumptions related to the large deviations of the single random variable $\frac{Z_1}{n}$. This is the main input with respect to a previous work by Najim [J. Najim, A Cramér type theorem for weighted random variables, Electron. J. Probab. 7 (4) (2002) 32 (electronic)].

MSC: primary 60F10; secondary 15A52; 15A18

Keywords: Large deviations; Spherical integrals; Spiked models

* Corresponding address: Ecole Nat. Sup. Telecom., Statistics, 46 rue Barrault, 75013, Paris, France. Tel.: +33 1 45 81 78 51; fax: +33 1 45 81 71 44.

E-mail addresses: Mylene.Maida@math.u-psud.fr (M. Maïda), najim@enst.fr (J. Najim), sandrine.peche@ujf-grenoble.fr (S. Péché).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2007.02.001

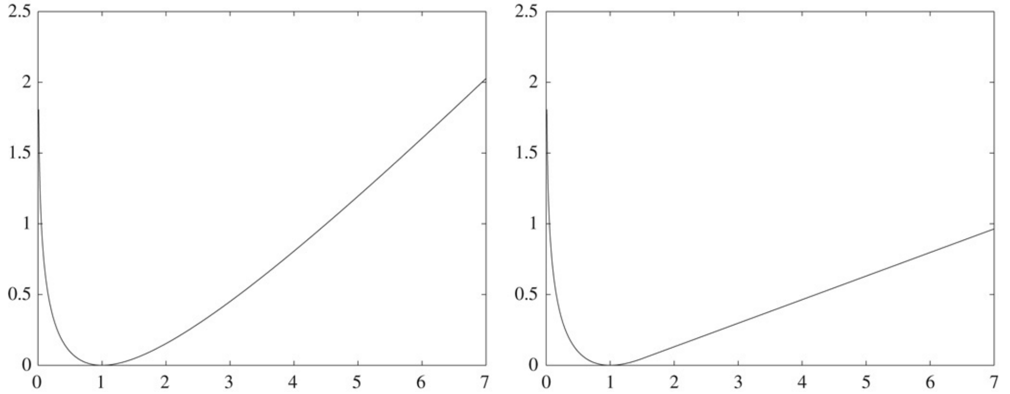

Fig. 1. The rate function of $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ where the X_i's are $\mathcal{N}(0,1)$ Gaussian i.i.d. random variables (left); the rate function of $\frac{1}{n}\sum_{i=1}^{n-1} X_i^2 + \frac{3}{n} X_n^2$ (right). Both rate functions coincide for $x \le \frac{3}{2}$ but the right one is linear for $x > \frac{3}{2}$.

1. Introduction

The model. We study in this article a Large Deviation Principle (LDP) for the weighted empirical mean

$$L_n = \frac{1}{n}\sum_{1}^{n} \mathbf{f}(x_i^n) \cdot Z_i,$$

where $(Z_i)_{i\in\mathbb{N}}$ is a sequence of $\mathbb{R}^d$-valued independent and identically distributed (i.i.d.) random variables satisfying:

$$\mathbb{E}\, e^{\alpha |Z_1|} < \infty \quad \text{for some } \alpha > 0. \tag{1.1}$$

The application $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m\times d}$ is an $m \times d$ matrix-valued continuous function, $(\mathcal{X}, \rho)$ being a locally compact metric space. The term $\mathbf{f}(x) \cdot Z$ denotes the product between matrix $\mathbf{f}(x)$ and vector Z. The set $\{x_i^n, 1 \le i \le n,\ n \ge 1\}$ is an $\mathcal{X}$-valued sequence of deterministic elements such that the empirical measure $\hat{R}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i^n}$ satisfies:

$$\hat{R}_n \xrightarrow[n\to\infty]{weakly} R, \tag{1.2}$$

where R is a probability measure with compact support $\mathcal{Y}$.

We focus in this paper on cases where there are outliers, that is where some of the x_i^n remain far from the support (also called bulk) of R. Loosely speaking, one can think of an outlier as a sequence $(x_{i(n)}^n, n \ge 1)$ satisfying:

$$\liminf_{n\to\infty} \rho(x_{i(n)}^n, \mathcal{Y}) > 0. \tag{1.3}$$

At a large deviation level, such outliers may have a dramatic impact on the shape of the rate function as demonstrated in the simple example of Fig. 1. Although the model under study looks very similar to the LDP studied in [11], the presence of outliers substantially modifies the resulting LDP and may naturally create infinitely many non-exposed points (see the definition in [7] and also Remarks 3.3 and 4.2) for the rate function.

The purpose of this article is to provide clear assumptions (which cover situations where (1.3) can occur) over the set $\{\mathbf{f}(x_i^n), 1 \le i \le n, 1 \le n\}$ and over Z_i under which fairly general LDP results can be proved.

Motivations and related work. Such models are of particular interest in the field of statistical mechanics (spherical spin glasses in [1], spherical integrals in the finite-rank case in [9], etc.) where one has often to establish a LDP for the empirical mean L_n in the case where the random variable Z_i satisfies condition (1.1). In particular, spherical integrals are intimately connected with the study of Deformed Ensembles (see [12] for instance for the definition) in Random Matrix Theory. In dimension one, Z_i is typically the square of a Gaussian random variable. The measure $\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}$ is then a realization of the empirical measure of the eigenvalues associated with a given random matrix model and there are important cases when some of the x_i^n's stay far away from the support of R. Indeed, there has recently been a strong interest in random matrix models (so-called spiked models) where some of the largest eigenvalues lie out of the bulk, that is where the set of limit points of $(x_i^n, 1 \le i \le n,\ n \ge 1)$ can differ from the support of R (see Johnstone [10], Baik et al. [2,3], Péché [12]). These spiked models are of particular interest for statistical applications [10].

The study of the LDP for weighted means was developed by Bercu et al. [5] for Gaussian functionals and considered in greater generality in Najim [11]. In [11], the LDP is stated for L_n under condition (1.1) but in the case where $(x_i^n, 1 \le i \le n, 1 \le n)$ is a subset of $\mathcal{Y}$, the support of the limiting probability measure R. In particular, the framework of [11] does not allow any of the x_i^n's to lie far from the bulk. LDPs involving outliers can be found in Bercu et al. [5], Guionnet and Maïda [9]. For related work concerning quadratic forms of Gaussian processes, we shall also refer the reader to Bercu et al. [4], Gamboa et al. [8], Bryc and Dembo [6] and Zani [15].

Presentation of the results. The purpose of this article is to establish the LDP for the empirical mean L_n under the moment assumption (1.1) and under assumptions which allow the presence of outliers (see (1.3)). Such a LDP will rely on the individual LDP for $\frac{Z_1}{n}$. This is the content of the following assumption.

Assumption A.1. The $\mathbb{R}^d$-valued random variable Z_1 satisfies the following exponential condition:

$$\mathbb{E}\, e^{\alpha |Z_1|} < \infty \quad \text{for some } \alpha > 0,$$

and $\frac{Z_1}{n}$ satisfies the LDP with a good rate function denoted by I.

Note that if $\frac{Z_i}{n}$ does not satisfy a LDP, one can construct counterexamples where L_n does not fulfill a LDP (see for instance [11, Section 2.3]). Finally, two subcases of Assumption A.1 yield two distinct classes of results:

The case where I is convex (Assumption A.2, Section 2.3). This paper is mainly devoted to the study of this case. If I is convex then the assumptions on the sets $C_n^{\mathbf{f}} = \{\mathbf{f}(x_i^n), 1 \le i \le n, 1 \le n\}$ needed to state the LDP for L_n are quite mild. Apart from a standard compacity assumption (Assumption A.3, see Section 2.3), the main assumption over $C_n^{\mathbf{f}}$ (Assumption A.4, Section 2.3) bears on the sole limiting points of $C_n^{\mathbf{f}}$ (in the sense of Painlevé–Kuratowski convergence of sets) and on their role in the LDP. It turns out that Assumption A.4 is an intricate assumption concerning the limiting behaviour of $C_n^{\mathbf{f}}$ and some limiting points of $C_n^{\mathbf{f}}$ involved in the definition of a certain convex domain. This convex domain plays a role in the definition of the rate function of the LDP. As demonstrated by examples in Section 2.2, Assumption A.4 covers a wide variety of models with outliers in the convex case, at least those for which a LDP is to be expected.

Under Assumptions A.1–A.4 and the more classical Assumption A.5 (convergence of $\hat{R}_n$ to R), the empirical mean L_n satisfies the LDP with a good convex rate function (Theorem 3.2). This rate function admits a fairly good representation (in terms of convex features) where the role of the outliers is quiet transparent (Theorem 3.6 and examples in Section 4).

The case where I is not convex. In this case, one can still prove the LDP but the assumptions over $C_n^{\mathbf{f}}$ are much more stringent and the rate function is given by an abstract formula. Moreover, very few insights can be gained by the study of the general formula of the rate function. It seems that the study must be held on a case-by-case analysis.

Outline of the article. In order to study the Large Deviations of L_n, we shall separate outliers from the bulk and accordingly split L_n into two subsums:

$$L_n = \frac{1}{n} \sum_{\{x_i^n \text{ far from the bulk}\}} \mathbf{f}(x_i^n) \cdot Z_i + \frac{1}{n} \sum_{\{x_i^n \text{ near or in the bulk}\}} \mathbf{f}(x_i^n) \cdot Z_i \triangleq \pi_n + \tilde{L}_n.$$

The idea is then to establish separately the LDP for each subsum. This line of proof has been developed in the one-dimensional setting for Gaussian quadratic forms by Bercu et al. [5] and is extended to the multi-dimensional setting in this article.

The paper is organized as follows. Sections 2–4 are devoted to the study of the convex case.

In Section 2, we study the Large Deviations for the following model:

$$\pi_n = \frac{1}{n} \sum_{x_i^n \in C_n} \mathbf{f}(x_i^n) \cdot Z_i \quad \text{where } \frac{\operatorname{card}(C_n)}{n} \xrightarrow[n\to\infty]{} 0. \tag{1.4}$$

The main assumptions related to the set $C_n^{\mathbf{f}} = \{\mathbf{f}(x_i^n);\ x_i^n \in C_n\}$ are stated and the LDP for π_n is established.

In Section 3, the decomposition $L_n = \pi_n + \tilde{L}_n$ where π_n satisfies (1.4) is precisely specified, the LDP for L_n is established and a representation formula is given for the rate function. Section 4 is devoted to examples of LDPs with outliers in the convex case.

A general LDP stated with an abstract rate function is established in the non-convex case in Section 5. In Section 6, a partial study of the rate function is also carried out in the non-convex case in the setting of a specific example.

Comments related to the link between the study of the spherical integral and the LDP of L_n are made in Section 4 (rank-one case) and Section 6 (higher rank).

2. The LDP for the partial mean π_n in the convex case

Let $(C_n)_{n\geq 1}$ be a finite subset of $\mathcal{X}$. This section is devoted to the study of the LDP of

$$\pi_n = \frac{1}{n} \sum_{x_i^n \in C_n} \mathbf{f}(x_i^n) \cdot Z_i \quad \text{where } \frac{\operatorname{card}(C_n)}{n} \xrightarrow[n\to\infty]{} 0,$$

with $\operatorname{card}(C_n)$ standing for the cardinality of the set C_n. It will be proved in Section 3.1 that L_n can be decomposed as $\pi_n + \tilde{L}_n$ with π_n as above.

Remark 2.1. In the case where the random variable Z_1 satisfies

$$\mathbb{E}e^{\alpha|Z_1|} < \infty \quad \text{for all } \alpha \in \mathbb{R}^+, \tag{2.1}$$

the following limit holds true:

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{|\pi_n| > \delta\} = -\infty \quad \text{for all } \delta > 0.$$

Otherwise stated L_n and $\tilde{L}_n$ are exponentially equivalent and π_n does not play any role at a large deviation level. Of course the situation is completely different if (2.1) does not hold.

We first introduce some notations as well as the concepts of inner limit, outer limit and Painlevé–Kuratowski convergence for sets. We then state the assumptions over the sets $C_n^{\mathbf{f}} = \{\mathbf{f}(x_i^n), x_i^n \in C_n\}$ and prove the LDP for π_n.

2.1. Notations

Denote by $\mathcal{B}(\mathcal{Z})$ the Borel sigma-field of a given topological space $\mathcal{Z}$ (usually $\mathbb{R}^d$, $\mathbb{R}^m$, $\mathbb{R}^{m\times d}$ or $\mathcal{X}$). Denote by $|\cdot|$ a norm on any finite-dimensional vector space ($\mathbb{R}^d$, $\mathbb{R}^m$ or $\mathbb{R}^{m\times d}$). In the sequel, we use bold letters **a, b, y**, etc. to denote $m \times d$ matrices. We denote by $\langle\cdot,\cdot\rangle$ the scalar product in any finite-dimensional space and by $\cdot$ the product between vectors and matrices with compatible size. Let A be a subset of $\mathbb{R}^k$. We denote by $\bar{A}$ its closure, by $\mathrm{int}(A)$ its interior, by $\Delta(\cdot \mid A)$ the convex indicator function of the set A and by $\Delta^*(\cdot \mid A)$ its convex conjugate (also called the support function of A), that is:

$$\Delta(\theta \mid A) = \begin{cases} 0 & \text{if } \theta \in A, \\ \infty & \text{else.} \end{cases},$$
$$\Delta^*(y \mid A) = \sup_{\theta\in\mathbb{R}^k} \{\langle y, \theta\rangle - \Delta(\theta \mid A)\} = \sup_{\theta\in A} \langle y, \theta\rangle,$$

where y and θ are in $\mathbb{R}^k$. The following proposition whose proof is straightforward will be of constant use in the sequel.

Proposition 2.1. *Let A be a subset of $\mathbb{R}^k$, then*

$$\Delta^*(\cdot \mid A) = \Delta^*(\cdot \mid \bar{A}).$$

If moreover A is convex with non-empty interior, then

$$\Delta^*(\cdot \mid \mathrm{int}(A)) = \Delta^*(\cdot \mid A) = \Delta^*(\cdot \mid \bar{A}).$$

Let D_n be a sequence of subsets of $\mathbb{R}^{m\times d}$. We define its outer limit (denoted by $D_{\infty,\mathrm{out}}$) and its inner limit (denoted by $D_{\infty,\mathrm{in}}$) by

$$D_{\infty,\mathrm{out}} = \left\{\mathbf{x} \in \mathbb{R}^{m\times d},\ \exists \phi : \mathbb{N} \to \mathbb{N} \text{ increasing}, \exists\, \mathbf{x}_{\phi(n)} \in D_{\phi(n)},\ \mathbf{x}_{\phi(n)} \xrightarrow[n\to\infty]{} \mathbf{x}\right\}$$
$$D_{\infty,\mathrm{in}} = \left\{\mathbf{x} \in \mathbb{R}^{m\times d},\ \exists\, n_0,\ \forall n \geq n_0, \exists\, \mathbf{x}_n \in D_n,\ \mathbf{x}_n \xrightarrow[n\to\infty]{} \mathbf{x}\right\}.$$

The limit D_∞ of the sets (D_n) exists if the outer limit and the inner limit are equal. Set convergence in this sense is known as Painlevé–Kuratowski convergence and in this case, we will denote:

$$D_n \xrightarrow[n\to\infty]{\mathrm{pk}} D_\infty.$$

For more details on Painlevé–Kuratowski convergence of sets, see Rockafellar and Wets [14, Chapter 4].

2.2. A preliminary analysis: Two simple examples

Consider

$$C_n^{\mathbf{f}} = \left\{\mathbf{f}(x_i^n),\, x_i^n \in C_n\right\} \quad \text{where } \frac{\mathrm{card}(C_n)}{n} \to 0.$$

The sets $C_{\infty,\mathrm{in}}^{\mathbf{f}}$ and $C_{\infty,\mathrm{out}}^{\mathbf{f}}$ are respectively the inner and outer limits of $(C_n^{\mathbf{f}})$. In the study of the forthcoming examples, we will focus on the links between the LDP for π_n and the sets $C_{\infty,\mathrm{in}}^{\mathbf{f}}$ and $C_{\infty,\mathrm{out}}^{\mathbf{f}}$. This section is aimed at introducing Assumption A.4 but can be skipped as no further notation is introduced.

2.2.1. Example 1: A simple case where the LDP fails to hold for π_n

Let X be a standard Gaussian random variable and consider $\pi_n = \frac{2+(-1)^n}{n} X^2$. Direct computations yield the LDP for π_{2n} (resp. π_{2n+1}) with good rate function Δ_{even}^* (resp. Δ_{odd}^*) where

$$\Delta_{\mathrm{even}}^*(z) = \begin{cases} z/6 & \text{if } z > 0,\\ \infty & \text{else.}\end{cases} \quad \text{and} \quad \Delta_{\mathrm{odd}}^*(z) = \begin{cases} z/2 & \text{if } z > 0,\\ \infty & \text{else.}\end{cases}$$

Therefore one cannot expect the LDP for $(\pi_n, n \in \mathbb{N})$.

2.2.2. Example 2: The LDP holds after modification of Example 1

Let X and Y be independent standard Gaussian random variables and consider $\pi_n = \frac{2+(-1)^n}{n} X^2 + \frac{4}{n} Y^2$. In this case, π_{2n} and π_{2n+1} satisfy the LDP (by a direct analysis) with the same rate function

$$\Delta^*(z) = \begin{cases} z/8 & \text{if } z > 0,\\ \infty & \text{else.}\end{cases}$$

This yields the LDP for the whole sequence $(\pi_n, n \in \mathbb{N})$ with rate function Δ^*.

Despite the erratic behaviour of $\frac{2+(-1)^n}{n} X^2$ (as seen in the previous example), the LDP holds due to presence of the term $\frac{4}{n} Y^2$.

2.2.3. Comparison of the two examples

Denote by

$$\mathcal{D}_y = \{\lambda \in \mathbb{R}, \log \mathbb{E} e^{\lambda y X^2} < \infty\} = (-\infty, (2y)^{-1})$$

where X is a standard Gaussian random variable.

In the case of Section 2.2.1, one can easily check that $C_{2n}^{\mathbf{f}} = \{3\}$ and $C_{2n+1}^{\mathbf{f}} = \{1\}$. Thus $C_{\infty,\mathrm{out}}^{\mathbf{f}} = \{1, 3\}$ while $C_{\infty,\mathrm{in}}^{\mathbf{f}} = \emptyset$. It is straightforward to check that the rate functions driving the LDP of π_{2n} and π_{2n+1} can be expressed as:

$$\Delta_{\mathrm{even}}^*(z) = \sup_{\lambda \in \mathcal{D}_3} \lambda z \quad \text{and} \quad \Delta_{\mathrm{odd}}^*(z) = \sup_{\lambda \in \mathcal{D}_1} \lambda z.$$

The very reason for which the LDP does not hold in this case is that

$$\bigcap_{y \in C_{\infty,\mathrm{out}}^{\mathbf{f}}} \mathcal{D}_y \neq \bigcap_{y \in C_{\infty,\mathrm{in}}^{\mathbf{f}}} \mathcal{D}_y.$$

In the case of Section 2.2.2, $C^{\mathbf{f}}_{2n} = \{3, 4\}$ while $C^{\mathbf{f}}_{2n+1} = \{1, 4\}$. Therefore $C^{\mathbf{f}}_{\infty,\mathrm{out}} = \{1, 3, 4\}$ while $C^{\mathbf{f}}_{\infty,\mathrm{in}} = \{4\}$. Despite the fact that $C^{\mathbf{f}}_{\infty,\mathrm{out}} \neq C^{\mathbf{f}}_{\infty,\mathrm{in}}$, the LDP holds in this case with good rate function given by:

$$\Delta^*(z) = \sup_{\lambda \in \mathcal{D}_4} \lambda z.$$

As we shall see, the underlying reason for which the LDP holds is

$$\bigcap_{y \in C^{\mathbf{f}}_{\infty,\mathrm{out}}} \mathcal{D}_y = \bigcap_{y \in C^{\mathbf{f}}_{\infty,\mathrm{in}}} \mathcal{D}_y \; (= \mathcal{D}_4),$$

and this will be a key-point in the statement of Assumption A.4.

We are now in a position to state the assumptions and the main result.

2.3. Assumptions and main results

Let C_n be a finite subset of $\mathcal{X}$ and recall that

$$C^{\mathbf{f}}_n = \left\{ \mathbf{f}(x^n_i),\; x^n_i \in C_n \right\} \quad \text{where } \frac{\operatorname{card}(C_n)}{n} \xrightarrow[n\to\infty]{} 0.$$

Let $\mathbf{y}$ be an $m \times d$ matrix and denote by

$$\mathcal{D}_{\mathbf{y}} = \left\{ \lambda \in \mathbb{R}^m, \log \mathbb{E}\, \mathrm{e}^{\langle \lambda, \mathbf{y} \cdot Z_1 \rangle} < \infty \right\}. \tag{2.2}$$

We can now state our assumptions.

Assume that Z_1 is an $\mathbb{R}^d$-valued random variable satisfying Assumption A.1 and recall that I is the rate function associated with $\frac{Z_1}{n}$.

Assumption A.2. Let $\mathcal{D}_Z \stackrel{\triangle}{=} \{\theta \in \mathbb{R}^d, \log \mathbb{E}\, \mathrm{e}^{\langle \theta, Z_1 \rangle} < \infty\}$, then

$$I(z) = \Delta^*(z \mid \mathcal{D}_Z).$$

In particular, I is a convex rate function.

Assumption A.3. Let $(D_n)_{n\geq 1}$ be a sequence of non-empty subsets of $\mathbb{R}^{m\times d}$. There exists a compact set $K \subset \mathbb{R}^{m\times d}$ such that $D_n \subset K$ for every $n \geq 1$.

Remark 2.2. This assumption implies in particular that the outer limit $D_{\infty,\mathrm{out}}$ of $(D_n)_{n\geq 1}$ is a non-empty compact set of $\mathbb{R}^{m\times d}$.

Assumption A.4. Let $(D_n)_{n\geq 1}$ be a sequence of subsets of $\mathbb{R}^{m\times d}$. Denote by $D_{\infty,\mathrm{in}}$ and $D_{\infty,\mathrm{out}}$ its inner and outer limits. Then:

$$\bigcap_{\mathbf{y} \in D_{\infty,\mathrm{in}}} \mathcal{D}_{\mathbf{y}} = \bigcap_{\mathbf{y} \in D_{\infty,\mathrm{out}}} \mathcal{D}_{\mathbf{y}}$$

where $\mathcal{D}_{\mathbf{y}}$ is defined by (2.2).

Remark 2.3. If $(D_n)_{n\geq 1}$ fulfills Assumptions A.3 and A.4, then in particular, $D_{\infty,\mathrm{in}}$ is not empty.

We can now state the main result of the section.

Theorem 2.2. *Assume that $(Z_i)_{i\in\mathbb{N}}$ is a sequence of $\mathbb{R}^d$-valued i.i.d. random variables. Assume moreover that Assumptions A.1 and A.2 hold for Z_1. Assume that $(\mathcal{X},\rho)$ is a metric space and let $C_n \subset \mathcal{X}$ be such that*

$$\frac{\mathrm{card}(C_n)}{n} \xrightarrow[n\to\infty]{} 0.$$

Denote by $C_n^{\mathbf{f}} = \{\mathbf{f}(x_i^n), x_i^n \in C_n\}$ where $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m\times d}$ is continuous. Assume that Assumptions A.3 and A.4 hold for the sequence of sets $(C_n^{\mathbf{f}})_{n\in\mathbb{N}}$. Then the random variable

$$\pi_n = \frac{1}{n}\sum_{x_i^n\in C_n} \mathbf{f}(x_i^n)\cdot Z_i$$

satisfies the LDP in $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ with good rate function

$$\Delta^*(z \mid \mathcal{D}) = \sup\{\langle\lambda, z\rangle, \lambda\in\mathcal{D}\} \quad \text{where } \mathcal{D} = \bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{in}}} \mathcal{D}_{\mathbf{y}} = \bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{out}}} \mathcal{D}_{\mathbf{y}}.$$

Remark 2.4 (*On Assumption A.4*). A close look to the proof of Theorem 2.2 shows that the rate function that drives the lower bound of the LDP is the support function of $\bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{in}}} \mathcal{D}_{\mathbf{y}}$ while the rate function that drives the upper bound is the support function of $\bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{out}}} \mathcal{D}_{\mathbf{y}}$. Both rate functions coincide when assuming Assumption A.4. (See also the examples in Section 2.2.)

2.4. Proof of Theorem 2.2

In order to prove Theorem 2.2, we follow the strategy developed in [11], essentially based on an exponential approximation technique. The next proposition is the counterpart of Lemma 5.1 in [11].

Lemma 2.3. *Let $\phi : \mathbb{N}\setminus\{0\} \to \mathbb{N}\setminus\{0\}$ be such that $\frac{\phi(n)}{n} \xrightarrow[n\to\infty]{} 0$. Let (Z_i) be a sequence of $\mathbb{R}^d$-valued random variables satisfying Assumptions A.1 and A.2. Then $\bar{Z}_n^\phi \triangleq \frac{1}{n}\sum_{i=1}^{\phi(n)} Z_i$ satisfies the LDP in $\mathbb{R}^d$ with good rate function given by*

$$I(y) = \Delta^*(y \mid \mathcal{D}_Z)$$

where $\mathcal{D}_Z$ is defined in Assumption A.2.

Proof. Denote by Λ_n^ϕ the log-Laplace transform of $\bar{Z}_n^\phi$, i.e. $\Lambda_n^\phi(\theta) = \log \mathbb{E}\, \mathrm{e}^{\langle\theta, \bar{Z}_n^\phi\rangle}$. Then

$$\frac{1}{n}\Lambda_n^\phi(n\theta) = \frac{\phi(n)}{n}\log\mathbb{E}\,\mathrm{e}^{\langle\theta, Z_i\rangle} \xrightarrow[n\to\infty]{} \Delta(\theta \mid \mathcal{D}_Z).$$

Therefore, the large deviation upper bound holds for $\bar{Z}_n^\phi$ with rate function I by Theorem 2.3.6(a) in [7]. To prove the large deviation lower bound, it is sufficient to prove that

$$-I(y) \le \liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\bar{Z}_n^\phi \in B(y,\varepsilon)\right)$$

where $B(y,\varepsilon) = \{y' \in \mathbb{R}^d, |y'-y| < \varepsilon\}$. Define

$$\tilde{Z}_n^\phi = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=2}^{\phi(n)} Z_i & \text{if } \phi(n)\ge 2,\\ 0 & \text{otherwise.}\end{cases}$$

Then $\{Z_1/n \in B(y, \varepsilon/3)\} \cap \{\tilde{Z}_n^{\phi} \in B(0, \varepsilon/3)\} \subset \{\bar{Z}_n^{\phi} \in B(y, \varepsilon)\}$ which yields

$$\frac{1}{n}\log \mathbb{P}\left(Z_1/n \in B(y, \varepsilon/3)\right) + \frac{1}{n}\log \mathbb{P}\left(\tilde{Z}_n^{\phi} \in B(0, \varepsilon/3)\right) \leq \frac{1}{n}\log \mathbb{P}\left(\bar{Z}_n^{\phi} \in B(y, \varepsilon)\right). \tag{2.3}$$

Exponential Markov inequality yields $\lim_{n\to\infty} \mathbb{P}\{|\tilde{Z}_n^{\phi}| > \varepsilon/3\} = 0$ which readily implies that $\lim_{n\to\infty} \mathbb{P}\{\tilde{Z}_n^{\phi} \in B(0, \varepsilon/3)\} = 1$. Consequently, taking the liminf on both sides of (2.3) and using the lower bound for the single variable $\frac{Z_1}{n}$ yields the desired lower bound. The proof is completed. □

We first consider Theorem 2.2 under an additional assumption.

Lemma 2.4. *Under the same assumptions as in Theorem 2.2 and if we assume in addition that*

$$C_n^{\mathbf{f}} \xrightarrow[n\to\infty]{\text{pk}} C_\infty^{\mathbf{f}}, \tag{2.4}$$

then π_n satisfies the LDP in $\mathbb{R}^d$ with good rate function $\Delta^*(\cdot \mid \mathcal{D})$, where $\mathcal{D} = \bigcap_{\mathbf{y}\in C_\infty^{\mathbf{f}}} \mathcal{D}_{\mathbf{y}}$.

Proof of Lemma 2.4 is postponed to Appendix.

We now relax the extra assumption (2.4) and prove Theorem 2.2. The scheme of the proof is the following. We first show, using directly the result in Lemma 2.4, that the lower bound is driven by the support function of the set $\bigcap_{\mathbf{y}\in C_{\infty,\text{in}}^{\mathbf{f}}} \mathcal{D}_{\mathbf{y}}$. We then obtain that the upper bound is driven by the support function of the set $\bigcap_{\mathbf{y}\in C_{\infty,\text{out}}^{\mathbf{f}}} \mathcal{D}_{\mathbf{y}}$, by majorizing the log-Laplace of π_n. Under Assumption A.4, both bounds coincide and we get the full LDP.

Proof of Theorem 2.2. To get the lower bound, we split $C_n^{\mathbf{f}}$ into two disjoint subsets:

$$C_n^{\mathbf{f}} = \mathcal{I}_n^{\mathbf{f}} \cup \mathcal{O}_n^{\mathbf{f}} \quad \text{where } \mathcal{I}_n^{\mathbf{f}} \xrightarrow[n\to\infty]{\text{pk}} C_{\infty,\text{in}}^{\mathbf{f}}. \tag{2.5}$$

Let us sketch the construction of $\mathcal{I}_n^{\mathbf{f}}$. Let $B(z, \frac{1}{m})$ be a ball centered in $z \in C_{\infty,\text{in}}^{\mathbf{f}}$ with radius $\frac{1}{m}$. Since $C_{\infty,\text{in}}^{\mathbf{f}}$ is compact by Assumption A.3, there exists $(z_\ell)_{1\leq \ell\leq L_m}$ such that

$$C_{\infty,\text{in}}^{\mathbf{f}} \subset \bigcup_{\ell=1}^{L_m} B\left(z_\ell, \frac{1}{m}\right) \quad \text{and} \quad B\left(z_\ell, \frac{1}{m}\right) \cap C_{\infty,\text{in}}^{\mathbf{f}} \neq \emptyset \quad \text{for } 1 \leq \ell \leq L_m.$$

The mere definition of $C_{\infty,\text{in}}^{\mathbf{f}}$ yields that there exists $\psi(m)$ such that for all ℓ, $1 \leq \ell \leq L_m$:

$$\forall n \geq \psi(m), \quad \exists \mathbf{f}(x_{i_\ell}^n) \in B\left(z_\ell, \frac{1}{m}\right) \quad \text{with } \mathbf{f}(x_{i_\ell}^n) \in C_n^{\mathbf{f}}.$$

Denote by $\mathcal{A}_{n,m}$ ($n \geq \psi(m)$) such a collection of $\mathbf{f}(x_{i_\ell}^n)$'s. Choose now similarly a collection of balls with radius $\frac{1}{m+1}$ and the related $\psi(m+1)$ with $\psi(m+1) > \psi(m)$, and set

$$\mathcal{I}_n^{\mathbf{f}} = \mathcal{A}_{n,m} \quad \text{if } \psi(m) \leq n < \psi(m+1).$$

With such a definition, it is straightforward to check that $\mathcal{I}_n^{\mathbf{f}} \xrightarrow{\text{pk}} C_{\infty,\text{in}}^{\mathbf{f}}$. We write

$$\pi_n = \frac{1}{n}\sum_{x_i^n \in \mathbf{f}^{-1}(\mathcal{I}_n^{\mathbf{f}})} \mathbf{f}(x_i^n)\cdot Z_i + \frac{1}{n}\sum_{x_i^n \notin \mathbf{f}^{-1}(\mathcal{I}_n^{\mathbf{f}})} \mathbf{f}(x_i^n)\cdot Z_i, \triangleq \pi_n^{\mathcal{I}} + \pi_n^{\mathcal{O}}.$$

The lower bound can be established as in Lemma 2.3. Let us prove that:

$$-\Delta^*\left(z\,\middle|\,\bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{in}}}\mathcal{D}_{\mathbf{y}}\right)\le\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\pi_n\in B(z,\varepsilon)\right). \tag{2.6}$$

Since

$$\{\pi_n^{\mathcal{I}}\in B(z,\varepsilon/3)\}\cap\{\pi_n^{\mathcal{O}}\in B(0,\varepsilon/3)\}\subset\{\pi_n\in B(z,\varepsilon)\},$$

one has

$$\begin{aligned}&\frac{1}{n}\log\mathbb{P}\left(\pi_n^{\mathcal{I}}\in B(z,\varepsilon/3)\right)+\frac{1}{n}\log\mathbb{P}\left(\pi_n^{\mathcal{O}}\in B(0,\varepsilon/3)\right)\\&\quad\le\frac{1}{n}\log\mathbb{P}\left(\pi_n\in B(z,\varepsilon)\right).\end{aligned} \tag{2.7}$$

Exponential Markov inequality yields $\lim_{n\to\infty}\mathbb{P}(|\pi_n^{\mathcal{O}}|>\varepsilon/3)=0$. This in turn implies that $\lim_{n\to\infty}\mathbb{P}\left(\pi_n^{\mathcal{O}}\in B(0,\varepsilon/3)\right)=1$. Since $\pi_n^{\mathcal{I}}$ fulfills assumptions of Lemma 2.4, the following lower bound holds:

$$-\Delta^*\left(z\,\middle|\,\bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{in}}}\mathcal{D}_{\mathbf{y}}\right)\le\frac{1}{n}\log\mathbb{P}\left(\pi_n^{\mathcal{I}}\in B(z,\varepsilon/3)\right). \tag{2.8}$$

Consequently, taking the liminf on both sides of (2.7) and using (2.8) yields the desired lower bound. The proof of the lower bound is completed.

Let us now prove the upper bound. Denote by $\Lambda_n(\lambda)$ the log-Laplace transform of π_n, i.e. $\Lambda_n(\lambda)=\log\mathbb{E}\,\mathrm{e}^{\langle\lambda,\pi_n\rangle}$. In order to prove the upper bound, we estimate the following limit:

$$\frac{1}{n}\Lambda_n(n\lambda)=\frac{1}{n}\sum_{x_i^n\in C_n}\log\mathbb{E}\,\mathrm{e}^{\langle\lambda,\mathbf{f}(x_i^n)\cdot Z_i\rangle}\quad\text{where }\frac{\mathrm{card}(C_n)}{n}\xrightarrow[n\to\infty]{}0.$$

We shall prove that

$$\limsup_{n\to\infty}\frac{1}{n}\Lambda_n(\lambda)\le\Delta\left(\lambda\,\middle|\,\mathrm{int}\left(\bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{out}}}\mathcal{D}_{\mathbf{y}}\right)\right). \tag{2.9}$$

Theorem 4.5.3 in [7] will then yield:

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\pi_n\in F)\le&\;-\inf_{z\in F}\Delta^*\left(z\,\middle|\,\mathrm{int}\left(\bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{out}}}\mathcal{D}_{\mathbf{y}}\right)\right)\\\underset{=}{(a)}&\;-\inf_{z\in F}\Delta^*\left(z\,\middle|\,\bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{out}}}\mathcal{D}_{\mathbf{y}}\right)\end{aligned} \tag{2.10}$$

for any closed set F. Equality (a) follows from Proposition 2.1 and the fact that $\mathrm{int}(\bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{out}}}\mathcal{D}_{\mathbf{y}})$ is a non-empty convex set due to Assumption A.1.

In order to prove (2.9), consider $\lambda \in \mathbb{R}^d$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \Lambda_n(n\lambda) > 0. \tag{2.11}$$

From (2.11), we can successively:

– extract a subsequence n_α from n such that

$$\lim_{n\to\infty} \frac{1}{n_\alpha} \sum_{x_i^{n_\alpha} \in C_{n_\alpha}} \log \mathbb{E} e^{\langle \lambda, \mathbf{f}(x_i^{n_\alpha}) \cdot Z_i \rangle} > 0,$$

– extract a subsequence n_β from n_α such that

$$\lim_{n\to\infty} \mathbb{E} e^{\langle \lambda, \mathbf{f}(x_i^{n_\beta}) \cdot Z_i \rangle} = \infty,$$

– extract a subsequence n_γ from n_β such that

$$\mathbf{f}(x_i^{n_\gamma}) \xrightarrow[n\to\infty]{} \mathbf{y}_0.$$

One can notice in particular that $\mathbf{y}_0 \in C^{\mathbf{f}}_{\infty,\mathrm{out}}$.

Let us now prove that

$$\lambda \notin \mathrm{int}(\mathcal{D}_{\mathbf{y}_0}). \tag{2.12}$$

Assume that (2.12) is not true. Then there exists $p > 1$ such that $p\lambda \in \mathcal{D}_{\mathbf{y}_0}$. Let $\varepsilon > 0$ be arbitrarily small. Then, if n is large enough to ensure that $|\lambda||\mathbf{f}(x_i^{n_\gamma}) - \mathbf{y}_0| \le \varepsilon/q$ where $1/p + 1/q = 1$, one has

$$\mathbb{E}\, e^{\langle \lambda, \mathbf{f}(x_i^{n_\gamma}) \cdot Z \rangle} = \mathbb{E}\, e^{\langle \lambda, \mathbf{y}_0 \cdot Z \rangle} e^{\langle \lambda, (\mathbf{f}(x_i^{n_\gamma}) - \mathbf{y}_0) \cdot Z \rangle} \le \left(\mathbb{E}\, e^{p\langle \lambda, \mathbf{y}_0 \cdot Z \rangle}\right)^{\frac{1}{p}} \left(\mathbb{E}\, e^{\varepsilon |Z|}\right)^{\frac{1}{q}}.$$

This contradicts the fact that

$$\lim_{n\to\infty} \mathbb{E} e^{\langle \lambda, \mathbf{f}(x_i^{n_\gamma}) Z_i \rangle} = \infty.$$

Therefore (2.12) holds and yields that $\lambda \notin \mathrm{int}(\bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{out}}} \mathcal{D}_{\mathbf{y}_0})$. From this, we deduce that

$$\limsup_{n\to\infty} \frac{1}{n} \Lambda_n(n\lambda) > 0 \Rightarrow \lambda \notin \mathrm{int}\left(\bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{out}}} \mathcal{D}_{\mathbf{y}}\right).$$

Stated otherwise:

$$\limsup_{n\to\infty} \frac{1}{n} \Lambda_n(n\lambda) \le \Delta\left(\lambda \mid \mathrm{int}\left(\bigcap_{\mathbf{y}\in C^{\mathbf{f}}_{\infty,\mathrm{out}}} \mathcal{D}_{\mathbf{y}}\right)\right).$$

Therefore, (2.9) is proved and so is (2.10).

Gathering the lower bound (2.6), the upper bound (2.10) and Assumption A.4 yield the full LDP for π_n. $\square$

3. The LDP for the empirical mean and the rate function in the convex case

Our goal is now to get the full LDP for L_n (Theorem 3.2 below). As announced in the outline of the article, the first step is to split the x_i^n's into two different subsets according to whether they live near the support of the limiting measure or whether they are outliers.

3.1. The decomposition $L_n = \pi_n + \tilde{L}_n$

Recall that $(\mathcal{X}, \rho)$ is a metric space.

Proposition 3.1. *Let $A_n = \{x_i^n,\ 1 \le i \le n\}$. Assume that*

$$\hat{R}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i^n} \xrightarrow[n\to\infty]{weakly} R$$

and denote by $\mathcal{Y}$ the support of R. Then there exist subsets B_n and $C_n = A_n \setminus B_n$ such that

(1) $\frac{\mathrm{card}(B_n)}{n} \xrightarrow[n\to\infty]{} 1$,

(2) $\frac{1}{\mathrm{card}(B_n)} \sum_{x_i^n \in B_n} \delta_{x_i^n} \xrightarrow[n\to\infty]{weakly} R$,

(3) $\rho(B_n, \mathcal{Y}) \xrightarrow[n\to\infty]{} 0$ *where $\mathcal{Y}$ is the support of R.*

We will then set

$$\tilde{L}_n = \frac{1}{n}\sum_{x_i^n \in B_n} \mathbf{f}(x_i^n)\cdot Z_i \quad \text{and} \quad \pi_n = \frac{1}{n}\sum_{x_i^n \in C_n} \mathbf{f}(x_i^n)\cdot Z_i.$$

Note that since $\mathrm{card}(B_n) + \mathrm{card}(C_n) = n$, property (1) yields then that $\frac{\mathrm{card}(C_n)}{n} \to 0$ as n goes to infinity.

Proof (*Construction of B_n*). Let $m \ge 1$ be fixed and denote by $\mathcal{Y}_m$ the $\frac{1}{m}$-blowup of $\mathcal{Y}$, i.e. $\mathcal{Y}_m = \{x \in \mathcal{X}, \rho(x, \mathcal{Y}) < \frac{1}{m}\}$ where $\mathcal{Y}$ is the support of R. Then $\frac{1}{n}\sum_1^n 1_{\mathcal{Y}_m}(x_i^n) \to 1$; in particular there exists $\psi_m \ge 1$ such that for all $n \ge \psi_m$:

$$\left| \frac{1}{n}\sum_{i=1}^{n} 1_{\mathcal{Y}_m}(x_i^n) - 1 \right| < \frac{1}{m}.$$

One can then build recursively a sequence of integers $(\psi_m)_{m\in\mathbb{N}}$ such that $\psi_m < \psi_{m+1}$ (so that $\psi_m \to \infty$ as $m \to \infty$). Set

$$B_n = \{x_i^n \in \mathcal{Y}_m,\ 1 \le i \le n\} \quad \text{for } \psi_m \le n < \psi_{m+1}.$$

We prove (1) and leave the proofs of properties (2) and (3) to the reader.

Let $\varepsilon > 0$ be fixed and take m such that $\frac{1}{m} < \varepsilon$. For such an m, take the corresponding ψ_m and let $n \ge \psi_m$. Then,

$$\left| \frac{\mathrm{card}(B_n)}{n} - 1 \right| = \left| \frac{\sum_{i=1}^{n} 1_{\mathcal{Y}_m}(x_i^n)}{n} - 1 \right| \le \frac{1}{m} < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, property (1) is proved. □

3.2. The LDP for the empirical mean L_n

In order to get the full LDP for $L_n = \tilde{L}_n + \pi_n$, we need to prove the LDP for $\tilde{L}_n$. We will mainly rely on the results in [11]. The following assumption is needed:

Assumption A.5. Assume that $(\mathcal{X}, \rho)$ is a locally compact metric space. The family $(x_i^n, 1 \leq i \leq n, n \geq 1) \subset \mathcal{X}$ satisfies

$$\hat{R}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i^n} \xrightarrow[n\to\infty]{\text{weakly}} R,$$

where R is a probability measure over $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Moreover, the support of R denoted by $\mathcal{Y}$ is a compact set and for every non-empty open set U of $\mathcal{Y}$ (for the induced topology over $\mathcal{Y}$), $R(U) > 0$.

Remark 3.1. The LDP may fail to hold if the last part of Assumption A.5, that is $R(U) > 0$ for U non-empty open set, is not fulfilled. Counter examples, also closely related to Assumption A.1, are developed in [11].

We recall that we denote by $\Lambda(\theta) = \log \mathbb{E}\, \mathrm{e}^{\langle \theta, Z_1 \rangle}$ the log-Laplace transform of Z_1. We introduce the following functional

$$\Gamma(\lambda) = \int_{\mathcal{X}} \Lambda\left(\sum_{k=1}^{m} \lambda_k f_k(x)\right) R(\mathrm{d}x), \tag{3.1}$$

where $\lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$ and f_k denotes the kth row of matrix $\mathbf{f}$. Let Γ^* be the convex conjugate of Γ:

$$\Gamma^*(z) = \sup_{\lambda \in \mathbb{R}^m} \{\langle \lambda, z \rangle - \Gamma(\lambda)\}.$$

We can now state the LDP.

Theorem 3.2. *Let $(Z_i)_{i \in \mathbb{N}}$ be a sequence of $\mathbb{R}^d$-valued i.i.d. random variables where Z_1 satisfies Assumptions A.1 and A.2.*

Consider a triangular array $(x_i^n, 1 \leq i \leq n, n \geq 1) \subset \mathcal{X}$ which fulfills Assumption A.5.

Denote by $C_n^{\mathbf{f}} = \{\mathbf{f}(x_i^n), x_i^n \in C_n\}$ where C_n is a subset of $\{x_i^n, 1 \leq i \leq n\}$ given by Proposition 3.1 and $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m \times d}$ is continuous. Assume that $C_n^{\mathbf{f}}$ satisfies Assumptions A.3 and A.4. Then

$$L_n = \frac{1}{n} \sum_{1}^{n} \mathbf{f}(x_i^n) \cdot Z_i$$

satisfies the LDP in $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ with good rate function

$$I_{\mathbf{f}}(z) = \inf\{\Gamma^*(z_1) + \Delta^*(z_2 \mid \mathcal{D}),\ z_1 + z_2 = z\},$$

where the definition of $\mathcal{D}$ follows from Theorem 2.2.

Proof. Recall the decomposition $L_n = \tilde{L}_n + \pi_n$ where

$$\tilde{L}_n = \frac{1}{n} \sum_{x_i^n \in B_n} \mathbf{f}(x_i^n) \cdot Z_i \quad \text{and} \quad \pi_n = \frac{1}{n} \sum_{x_i^n \in C_n} \mathbf{f}(x_i^n) \cdot Z_i,$$

where the sets B_n and C_n are defined in Section 3.1. Theorem 2.2 yields the LDP for π_n with good rate function $\Delta^*(\cdot \mid \mathcal{D})$. It remains now to prove the LDP for $\tilde{L}_n$. We will rely on Theorem 2.2 in [11] and therefore slightly modify $\tilde{L}_n$ so that it fulfills the assumptions of this theorem.

In fact, it is required in [11] that all the points x_i^n belong to $\mathcal{Y}$, which might not be the case here. We build in the sequel a sequence $(\tau(x_i^n)) \subset \mathcal{Y}$ which approximates the sequence $(x_i^n, x_i^n \in B_n)$. Let $x_i^n \in B_n$ and set

$$\tau(x_i^n) = \begin{cases} x_i^n & \text{if } x_i^n \in \mathcal{Y}, \\ \text{one of the argmin } \{\rho(x, x_i^n),\ x \in \mathcal{Y}\} & \text{else.} \end{cases}$$

Such a minimizer always exists and belongs to $\mathcal{Y}$ since $\mathcal{Y}$ is compact.

Since $\lim_n \sup\{\rho(x, \mathcal{Y}),\ x \in B_n\} = 0$, one has $\sup_{x_i^n \in B_n} \rho(x_i^n, \tau(x_i^n)) \xrightarrow[n\to\infty]{} 0$ and

$$\kappa_n(\mathbf{f}) \triangleq \sup_{x_i^n \in B_n} \{|\mathbf{f}(x_i^n) - \mathbf{f}(\tau(x_i^n))|\} \xrightarrow[n\to\infty]{} 0.$$

Indeed, for n large enough, B_n lies in an ε-blowup of $\mathcal{Y}$, which is compact since $\mathcal{X}$ is locally compact and $\mathbf{f}$ is therefore uniformly continuous on this set.

Now, if we define $\bar{L}_n$ by

$$\bar{L}_n \triangleq \frac{1}{n} \sum_{x_i^n \in B_n} \mathbf{f}(\tau(x_i^n)) \cdot Z_i,$$

then $\tilde{L}_n$ and $\bar{L}_n$ are exponentially equivalent. Indeed,

$$\begin{aligned} \frac{1}{n} \log \mathbb{P}\left(|\tilde{L}_n - \bar{L}_n| > \varepsilon\right) &\le \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{\mathrm{card}(B_n)} |Z_i| > \frac{\varepsilon}{\kappa_n(\mathbf{f})}\right) \\ &\le -\Lambda_{|Z|}^*\left(\frac{\varepsilon}{\kappa_n(\mathbf{f})}\right) \xrightarrow[n\to\infty]{} -\infty \end{aligned}$$

where $\Lambda_{|Z|}^*$ stands for the convex conjugate of the log-Laplace transform of $|Z|$. The measure $\bar{L}_n$ satisfies all the assumptions of Theorem 2.2 in [11]. Therefore, the LDP holds for it with good rate function Γ^*. Finally the exponential equivalence yields the LDP for $\tilde{L}_n$ with the same rate function (see for instance [7, Theorem 4.2.13]).

As the two subsums are independent, the contraction principle yields the LDP for L_n with good rate function $I_{\mathbf{f}}$ given by:

$$I_{\mathbf{f}}(z) = \inf\{\Gamma^*(z_1) + \Delta^*(z_2 \mid \mathcal{D}), z_1 + z_2 = z\}. \quad \square \tag{3.2}$$

3.3. *More insight on the rate function $I_{\mathbf{f}}$*

In the convex case, that is when Assumption A.2 holds, the rate function $I_{\mathbf{f}}$ can be expressed more explicitly. This section is aimed at describing how to perform the inf-convolution (3.2).

We first introduce some definitions from convex analysis (see e.g. [13]). The main result is stated in Theorem 3.6.

Definition 3.3 (*Normal Cone*). Let $\mathcal{C} \subset \mathbb{R}^d$ be a convex set and let $a \in \mathcal{C}$. The normal cone of $\mathcal{C}$ at a, denoted by $N_{\mathcal{C}}(a)$, is defined by:

$$N_{\mathcal{C}}(a) = \{z \in \mathbb{R}^d;\ \langle z, x - a\rangle \le 0, \forall x \in \mathcal{C}\}.$$

Remark 3.2. In particular, if $z \in N_{\mathcal{C}}(a)$ then $\Delta^*(z \mid \mathcal{C}) = \langle z, a\rangle$.

Definition 3.4 (*Relative Interior*). Let $\mathcal{C} \subset \mathbb{R}^d$ be a convex set. Its affine hull, denoted by aff $\mathcal{C}$, is the smallest affine subset of $\mathbb{R}^d$ containing $\mathcal{C}$. The relative interior of $\mathcal{C}$, denoted by ri $\mathcal{C}$, is defined by:

$$\mathrm{ri}\,\mathcal{C} \triangleq \left\{x \in \mathrm{aff}\,\mathcal{C}, \exists \varepsilon > 0 \text{ such that } (x + \varepsilon B(0,1)) \bigcap \mathrm{aff}\,\mathcal{C} \subset \mathcal{C}\right\}.$$

Definition 3.5 (*Subdifferential of a Convex Function*). A vector x^* is said to be a subgradient of a convex function f at a point x if for any z,

$$f(z) \geq f(x) + \langle x^*, z - x\rangle.$$

The subdifferential $\partial f(x)$ of f at x is the set of all subgradients of f at x.

We can now state:

Theorem 3.6. *Under the assumptions of Theorem 3.2, the rate function $I_{\mathbf{f}}$ admits the following representation:*

$$I_{\mathbf{f}}(z) = \sup_{\lambda \in \mathcal{D}} (\langle \lambda, z\rangle - \Gamma(\lambda)), \tag{3.3}$$

where Γ is given by (3.1). Furthermore, for any $z \in$ ri dom $I_{\mathbf{f}}$, we can decompose z as $z = z^* + z_{\mathbf{n}}$, where there exists $\lambda^* \in$ dom $\Gamma \cap \bar{\mathcal{D}}$ such that:

(i) $z^* \in \partial\Gamma(\lambda^*)$ *and*
(ii) $z_{\mathbf{n}} \in N_{\bar{\mathcal{D}}}(\lambda^*)$.

In particular, for any such decomposition,

$$I_{\mathbf{f}}(z) = \Gamma^*(z^*) + \Delta^*(z_{\mathbf{n}} \mid \mathcal{D}).$$

Remark 3.3 (*Non-exposed Points*). Let $z \in$ ri dom $I_{\mathbf{f}}$. Consider the decomposition given by Theorem 3.6, namely $z = z^* + z_{\mathbf{n}}$, then:

$$\forall t \in \mathbb{R}^+, \quad I_{\mathbf{f}}(z^* + t z_{\mathbf{n}}) = \Gamma^*(z^*) + t\langle z_{\mathbf{n}}, \lambda^*\rangle \quad \text{where } z^* \in \partial\Gamma(\lambda^*) \text{ and } z_{\mathbf{n}} \in N_{\mathcal{D}}(\lambda^*).$$

In particular if $z_{\mathbf{n}} \neq 0$, $I_{\mathbf{f}}$ is affine in the direction $\mathbb{R}^+ \ni t \mapsto z^* + t z_{\mathbf{n}}$ and has thus infinitely many non-exposed points (see for instance the example developed in Section 4).

Proof. We first prove (3.3). Theorem 3.2 and Proposition 2.1 yield

$$I_{\mathbf{f}}(z) = \inf_{z = z_1 + z_2} \{\Gamma^*(z_1) + \Delta^*(z_2 \mid \mathcal{D})\}.$$

As $I_{\mathbf{f}}$, Γ and $\Delta(. \mid \bar{\mathcal{D}})$ are convex, proper and lower semicontinuous, we get from Theorem 16.4 in [13] that

$$\begin{aligned} I_{\mathbf{f}}(z) &= \left[\Gamma + \Delta(. \mid \bar{\mathcal{D}})\right]^*(z), \\ &= \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, z\rangle - \Gamma(\lambda) - \Delta(\lambda \mid \bar{\mathcal{D}})\}, \\ &= \sup_{\lambda \in \bar{\mathcal{D}}} \{\langle \lambda, z\rangle - \Gamma(\lambda)\} = \sup_{\lambda \in \mathcal{D}} \{\langle \lambda, z\rangle - \Gamma(\lambda)\}, \end{aligned}$$

and (3.3) is proved. As $I_{\mathbf{f}}$ is convex, so is its domain and we can consider its relative interior ri dom $I_{\mathbf{f}}$. Let $z \in$ ri dom $I_{\mathbf{f}}$, then $I_{\mathbf{f}}(z) < +\infty$ and define F_z by :

$$F_z(x) = \Gamma^*(x) + \Delta^*(z - x \mid \bar{\mathcal{D}}).$$

The properties of Γ^* and $\Delta^*(. \mid \bar{\mathcal{D}})$ yield that F_z is proper, convex and lower semicontinuous; its level sets are compact. In particular, the infimum of F_z is attained over $\mathbb{R}^d$. Let z^* be a point where this infimum is attained, i.e.

$$\inf_{x \in \mathbb{R}^d} F_z(x) = F_z(z^*).$$

In this case,

$$0 \in \partial F_z(z^*).$$

In order to go further in the proof, we shall describe $\partial F_z(z^*)$ in terms of $\partial \Gamma^*$ and $\partial \Delta^*(z - \cdot \mid \bar{\mathcal{D}})$. This is the purpose of the following proposition:

Proposition 3.7. *If* $z \in \operatorname{ri} \operatorname{dom} I_{\mathbf{f}}$, *then for any* x,

$$\partial F_z(x) = \partial \Gamma^*(x) - \partial \Delta^*(z - x \mid \bar{\mathcal{D}}).$$

Proof of Proposition 3.7. Define f_z to be the function given by $f_z(x) = \Delta^*(z - x \mid \bar{\mathcal{D}})$. Note in particular that $F_z(x) = \Gamma^*(x) + f_z(x)$. Since $I_{\mathbf{f}}(z) = \inf_{z=z_1+z_2}\{\Gamma^*(z_1) + \Delta^*(z_2 \mid \mathcal{D})\}$, the sum of the epigraphs of Γ^* and Δ^* are equal to the epigraph of $I_{\mathbf{f}}$. This immediately implies that

$$\operatorname{dom} I_{\mathbf{f}} = \operatorname{dom} \Gamma^* + \operatorname{dom} \Delta^*(\cdot \mid \bar{\mathcal{D}}).$$

These sets being convex, Corollary 6.6.2 in [13] yields

$$\operatorname{ri} \operatorname{dom} I_{\mathbf{f}} = \operatorname{ri} \operatorname{dom} \Gamma^* + \operatorname{ri} \operatorname{dom} \Delta^*(\cdot \mid \bar{\mathcal{D}}).$$

Let $z \in \operatorname{ri} \operatorname{dom} I_{\mathbf{f}}$, then there exists $y \in \operatorname{ri} \operatorname{dom} \Gamma^*$ such that $z - y \in \operatorname{ri} \operatorname{dom} \Delta^*(\cdot \mid \bar{\mathcal{D}})$. This is equivalent to the fact that $y \in \operatorname{ri} \operatorname{dom} f_z(x)$ and therefore

$$\operatorname{ri} \operatorname{dom} \Gamma^* \cap \operatorname{ri} \operatorname{dom} f_z \neq \emptyset. \tag{3.4}$$

Theorem 23.8 in [13] whose main assumption is fulfilled by (3.4) yields then

$$\begin{aligned} \partial F_z(x) &= \partial \Gamma^*(x) + \partial f_z(x) \\ &= \partial \Gamma^*(x) - \partial \Delta^*(z - x \mid \bar{\mathcal{D}}) \end{aligned}$$

and Proposition 3.7 is proved. □

Let us now go back to the proof of Theorem 3.6. By Proposition 3.7,

$$\partial F_z(z^*) = \partial \Gamma^*(z^*) - \partial \Delta^*(z - z^* \mid \bar{\mathcal{D}}).$$

Since $0 \in \partial F_z(z^*)$, there exists $\lambda^* \in \partial \Gamma^*(z^*)$ such that $\lambda^* \in \partial \Delta^*(z - z^* \mid \bar{\mathcal{D}})$. By applying Theorem 23.5 in [13], one obtains

$$\lambda^* \in \partial \Gamma^*(z^*) \Leftrightarrow z^* \in \partial \Gamma(\lambda^*)$$

which in particular implies that $\lambda^* \in \operatorname{dom} \Gamma$. Moreover,

$$\begin{aligned} -\lambda^* \in \partial \Delta^*(z - z^* \mid \bar{\mathcal{D}}) &\Leftrightarrow z - z^* \in \partial \Delta(\lambda^* \mid \bar{\mathcal{D}}) \\ &\Leftrightarrow z - z^* \in N_{\bar{\mathcal{D}}}(\lambda^*), \end{aligned}$$

which in particular implies that $\lambda^* \in \bar{\mathcal{D}}$.

By denoting $z_{\mathbf{n}} = z - z^*$, one obtains the decomposition stated in Theorem 3.6. It remains to prove that:

$$I_{\mathbf{f}}(z) = \Gamma^*(z^*) + \Delta^*(z_{\mathbf{n}} \mid \bar{\mathcal{D}}).$$

We have:

$$\begin{aligned} I_{\mathbf{f}}(z) &= \sup_{\lambda \in \bar{\mathcal{D}}} \{\langle \lambda, z \rangle - \Gamma(\lambda)\} \\ &\geq \langle \lambda^*, z^* \rangle - \Gamma(\lambda^*) + \langle \lambda^*, z_{\mathbf{n}} \rangle = \Gamma^*(z^*) + \Delta^*(z_{\mathbf{n}} \mid \mathcal{D}). \end{aligned}$$

On the other hand,

$$\begin{aligned} I_{\mathbf{f}}(z) &= \sup_{\lambda \in \bar{\mathcal{D}}} \{\langle \lambda, z \rangle - \Gamma(\lambda)\} \\ &\leq \sup_{\lambda \in \bar{\mathcal{D}}} \{\langle \lambda, z^* \rangle - \Gamma(\lambda)\} + \sup_{\lambda \in \bar{\mathcal{D}}} \langle \lambda, z_{\mathbf{n}} \rangle = \Gamma^*(z^*) + \langle \lambda^*, z_{\mathbf{n}} \rangle, \end{aligned}$$

and Theorem 3.6 is proved. □

4. An example of LDP in the convex case

To illustrate the range of Theorems 3.2 and 3.6, we study in detail the following model:

$$L_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{f}(x_i^n) \cdot Z_i \quad \text{where } \mathbf{f}(x) = \begin{pmatrix} 1 & 0 \\ 0 & x \end{pmatrix} \quad \text{and} \quad Z_i = \begin{pmatrix} X_i^2 \\ X_i^2 \end{pmatrix}, \tag{4.1}$$

the sequence $(X_i)_{i \in \mathbb{N}}$ being a sequence of i.i.d. $\mathcal{N}(0, 1)$ Gaussian random variables and $(x_i^n)_{n \in \mathbb{N}}$ being a sequence of real numbers satisfying

$$\hat{R}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i^n} \to R.$$

We assume moreover that the support $\mathcal{Y}$ of R is given by $\mathcal{Y} = [m, M]$ and that

$$\sup_{1 \leq i \leq n} x_i^n \xrightarrow[n \to \infty]{} x_{\max} > M \quad \text{and} \quad \inf_{1 \leq i \leq n} x_i^n \xrightarrow[n \to \infty]{} x_{\min} < m.$$

Our goal is to establish the LDP for L_n and to describe as explicitly as possible the related rate function $I_{\mathbf{f}}$.

Remark 4.1. This example can be seen as the extension to the dimension 2 of the example studied in [5]. Indeed, under the same assumptions, Bercu et al. study the LDP for the following empirical mean $\frac{1}{n} \sum_{i=1}^{n} x_i^n X_i^2$.

Proposition 4.1 below is devoted to the description of the rate function. We first need the following notations. For $(\xi, \xi') \in \mathbb{R}^2$, set

$$\Gamma(\xi, \xi') = -\frac{1}{2} \int \log(1 - 2\xi - 2x\xi') R(\mathrm{d}x), \tag{4.2}$$

and denote by Γ^* the convex conjugate of Γ (the expression for Γ follows from a Gaussian integration and from formula (3.1)). Define H to be the Hilbert transform of R, that is

$$H(t) = \int \frac{R(\mathrm{d}x)}{t - x} \quad \text{for } t \in [m, M]^c.$$

Set

$$H_{\min} = H(x_{\min}) \quad \text{and} \quad \alpha_{\min} = x_{\min} - \frac{1}{H_{\min}};$$
$$H_{\max} = H(x_{\max}) \quad \text{and} \quad \alpha_{\max} = x_{\max} - \frac{1}{H_{\max}}.$$

Note that under the assumption that $x_{\min} < m$ and $x_{\max} > M$, $H_{\min}$ is a well-defined negative number while $H_{\max}$ is a well-defined positive number. In particular $x_{\min} < \alpha_{\min}$ and $\alpha_{\max} < x_{\max}$. Moreover, the following inequalities hold true:

$$m < \alpha_{\min} \leq \int x R(\mathrm{d}x) \quad \text{and} \quad \int x R(\mathrm{d}x) \leq \alpha_{\max} < M.$$

In particular, $\alpha_{\min} \leq \alpha_{\max}$. In order to describe the rate function related to the LDP of L_n, we introduce the following domains:

$$\begin{aligned}
\mathcal{D}_\infty &= \{(x, y) \in \mathbb{R}^2, x \leq 0 \text{ or } y \geq x_{\max} x \text{ or } y \leq x_{\min} x\} \\
\mathcal{D}_{(I_{\mathbf{f}} = \Gamma^*)} &= \{(x, y) \in \mathbb{R}^2, x > 0 \text{ and } \alpha_{\min} x \leq y \leq \alpha_{\max} x\} \\
\mathcal{D}^+_{linear} &= \{(x, y) \in \mathbb{R}^2,\ x > 0 \text{ and } \alpha_{\max} x < y \leq x_{\max} x\} \\
\mathcal{D}^-_{linear} &= \{(x, y) \in \mathbb{R}^2,\ x > 0 \text{ and } x_{\min} x \leq y < \alpha_{\min} x\}.
\end{aligned}$$

These domains are represented in Fig. 3 (right). We can now state the following result.

Proposition 4.1. *The empirical mean L_n defined in (4.1) satisfies the LDP in $\mathbb{R}^2$ with good rate function $I_{\mathbf{f}}$ given by*

(1) *If $(x, y) \in \mathcal{D}_\infty$ then $I_{\mathbf{f}}(x, y) = +\infty$,*
(2) *If $(x, y) \in \mathcal{D}_{(I_{\mathbf{f}} = \Gamma^*)}$ then $I_{\mathbf{f}}(x, y) = \Gamma^*(x, y)$,*
(3) *If $(x, y) \in \mathcal{D}^+_{linear}$ then*

$$I_{\mathbf{f}}(x, y) = \Gamma^*\left(H_{\max}(x_{\max} x - y), \alpha_{\max} H_{\max}(x_{\max} x - y)\right) + \frac{1}{2}\left((1 - H_{\max} x_{\max})\, x + H_{\max} y\right),$$

(4) *If $(x, y) \in \mathcal{D}^-_{linear}$ then*

$$I_{\mathbf{f}}(x, y) = \Gamma^*\left(H_{\min}(x_{\min} x - y), \alpha_{\min} H_{\min}(x_{\min} x - y)\right) + \frac{1}{2}\left((1 - H_{\min} x_{\min}) x + H_{\min} y\right).$$

Remark 4.2. Let $x_0 > 0$ be fixed and consider the ray:

$$y^-(x) = x_{\min} x + (\alpha_{\min} - x_{\min}) x_0, \quad x \geq x_0.$$

Then

$$I_{\mathbf{f}}(x, y^-(x)) = \Gamma^*(x_0, \alpha_{\min} x_0) + \frac{1}{2}(x - x_0).$$

In particular, there are infinitely many non-exposed points for $I_{\mathbf{f}}$ along the ray $((x, y^-(x)); x \geq x_0)$. The same can be shown along the ray

$$y^+(x) = x_{\max} x + (\alpha_{\max} - x_{\max}) x_0; \quad x \geq x_0.$$

Proof of Proposition 4.1. The LDP will be established as soon as assumptions of Theorem 3.2 are fulfilled. It is straightforward to check Assumptions A.1–A.3 and A.5. In order to check Assumption A.4, we rely on the following lemma:

Lemma 4.2. *For every $x \in [x_{\min}, x_{\max}]$, one has:*

$$\mathcal{D}_{\mathbf{f}(x_{\min})} \cap \mathcal{D}_{\mathbf{f}(x_{\max})} \subset \mathcal{D}_{\mathbf{f}(x)}.$$

Proof of Lemma 4.2. Let $(\xi, \xi') \in \mathcal{D}_{\mathbf{f}(x_{\min})} \cap \mathcal{D}_{\mathbf{f}(x_{\max})}$. This implies that $(\xi, x_{\min}\xi') \in \mathcal{D}_{Z_1}$ and $(\xi, x_{\max}\xi') \in \mathcal{D}_{Z_1}$. Every $x \in [x_{\min}, x_{\max}]$ can be written as a convex combination of $x_{\min}$ and $x_{\max}$: $x = ax_{\min} + bx_{\max}$, where $a + b = 1$, a, b being non-negative. By convexity of $\mathcal{D}_{Z_1}$, $(\xi, x\xi') = a(\xi, x_{\min}\xi') + b(\xi, x_{\max}\xi') \in \mathcal{D}_{Z_1}$. Therefore $(\xi, \xi') \in \mathcal{D}_{\mathbf{f}(x)}$. □

We can now check Assumption A.4. The mere definition of $x_{\min}$ and $x_{\max}$ implies that both $x_{\min}$ and $x_{\max}$ belong to $C^{\mathbf{f}}_{\infty,\text{out}}$ and $C^{\mathbf{f}}_{\infty,\text{in}}$ and that both $C^{\mathbf{f}}_{\infty,\text{out}}$ and $C^{\mathbf{f}}_{\infty,\text{in}}$ are included in $[x_{\min}, x_{\max}]$. In particular, the set $\mathcal{D}$ is well defined and is given by:

$$\mathcal{D} = \bigcap_{\{x,\ \mathbf{f}(x) \in C^{\mathbf{f}}_{\infty,\text{out}}\}} \mathcal{D}_{\mathbf{f}(x)} \stackrel{(a)}{=} \mathcal{D}_{\mathbf{f}(x_{\min})} \cap \mathcal{D}_{\mathbf{f}(x_{\max})} \stackrel{(b)}{=} \bigcap_{\{x, \mathbf{f}(x) \in C^{\mathbf{f}}_{\infty,\text{in}}\}} \mathcal{D}_{\mathbf{f}(x)}$$

where (a) and (b) follow from Lemma 4.2. An easy computation yields

$$\mathcal{D} = \{(\xi, \xi') \in \mathbb{R}^2;\ 1 - 2\xi - 2x_{\min}\xi' > 0 \text{ and } 1 - 2\xi - 2x_{\max}\xi' > 0\}. \tag{4.3}$$

The LDP is therefore established by applying Theorem 3.2 and the rate function is given by:

$$I_{\mathbf{f}}(z) = \inf_{z = z_1 + z_2} \{\Gamma^*(z_1) + \Delta^*(z_2 | \mathcal{D})\},$$

with $\mathcal{D}$ as above and Γ as defined in (3.1). Formula (4.2) yields:

$$\operatorname{dom} \Gamma = \{(\xi, \xi') \in \mathbb{R}^2;\ 1 - 2\xi - 2x\xi' > 0 \text{ for all } x \in [m, M]\},$$

and therefore

$$\operatorname{dom} \Gamma = \{(\xi, \xi') \in \mathbb{R}^2;\ 1 - 2\xi - 2m\xi' > 0 \text{ and } 1 - 2\xi - 2M\xi' > 0\}. \tag{4.4}$$

Fig. 2 shows dom Γ and $\mathcal{D}$ for particular choices of the parameters.

We first prove Proposition 4.1(1). In order to prove this statement, it is equivalent to determine the domain of $I_{\mathbf{f}}$. We use the fact that

$$\operatorname{dom} I_{\mathbf{f}} = \operatorname{dom} \Gamma^* + \operatorname{dom} \Delta^*(\cdot \mid \mathcal{D})$$

and focus on the two domains of the right-hand side. One can check that

$$\operatorname{dom} \Gamma^* = \{(x, y) \in \mathbb{R}^2;\ x > 0 \text{ and } mx \le y \le Mx\},$$
$$\operatorname{dom} \Delta^*(\cdot \mid \mathcal{D}) = \{(x, y) \in \mathbb{R}^2;\ x \ge 0 \text{ and } x_{\min}x \le y \le x_{\max}x\}.$$

Therefore

$$\operatorname{dom} I_{\mathbf{f}} = \{(x, y) \in \mathbb{R}^2;\ x > 0 \text{ and } x_{\min}x < y < x_{\max}x\}. \tag{4.5}$$

Note in particular that in this case, ri dom $I_{\mathbf{f}} = \operatorname{dom} I_{\mathbf{f}}$.

The three domains dom Γ^*, dom $\Delta^*(\cdot \mid \mathcal{D})$ and dom $I_{\mathbf{f}}$ are represented on Fig. 3.

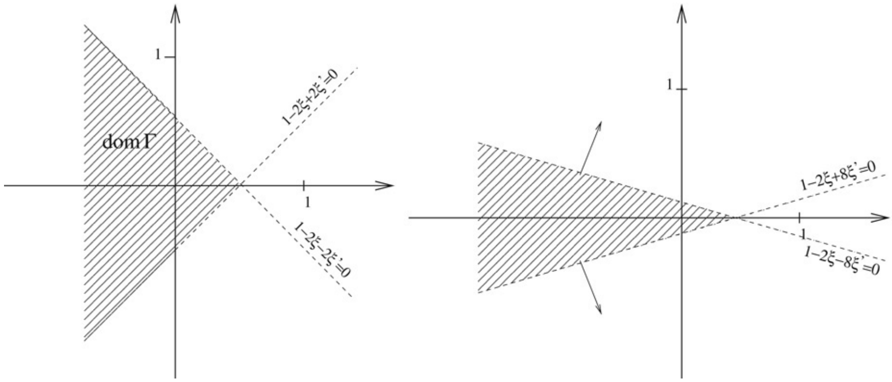

Fig. 2. In this figure are represented dom Γ for $m = -1$ and $M = 1$ (left) and $\mathcal{D}$ for $x_{\min} = -4$ and $x_{\max} = 4$ (right). In the picture of $\mathcal{D}$, we figured also some of the normal cones to $\bar{\mathcal{D}}$, whose directions are represented by the arrows.

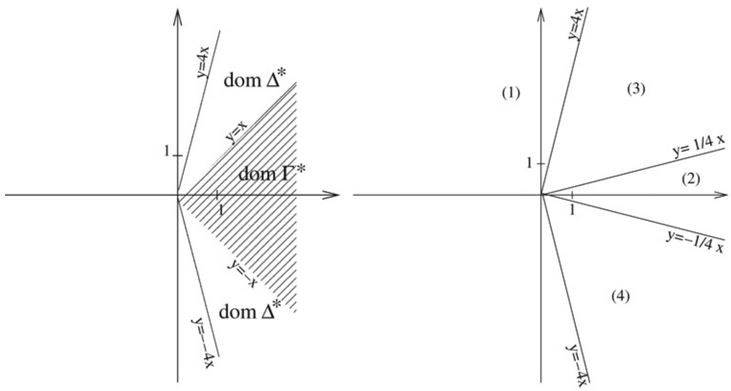

Fig. 3. The left picture represents dom Γ^* (hatched cone) and dom $\Delta^*(\cdot \mid \mathcal{D})$ (delimited by the two half-lines $y = 4x$ and $y = -4x$). The right picture represents the four zones of $\mathbb{R}^2$ where $I_{\mathbf{f}}$ has a particular expression. Zone (1) (resp. (2), (3) and (4)) represents $\mathcal{D}_\infty$ (resp. $\mathcal{D}_{(I_{\mathbf{f}}=\Gamma^*)}$, $\mathcal{D}^+_{linear}$ and $\mathcal{D}^-_{linear}$). We kept the same values of the parameters as in Fig. 2 and chose $H_{\max} = -H_{\min} = 2/5$.

We now prove Proposition 4.1(2). Theorem 3.6 yields:

$$I_{\mathbf{f}}(z) = \sup_{\lambda \in \bar{\mathcal{D}}} \{\langle \lambda, z \rangle - \Gamma(\lambda)\}.$$

If one considers $g_z(\lambda) = \langle \lambda, z \rangle - \Gamma(\lambda)$, one can check that for $z \in \operatorname{dom} \Gamma^*$, an element $\bar{\lambda} = (\bar{\xi}, \bar{\xi}')$ realizing that the supremum of g_z satisfies the condition

$$\alpha - \frac{1}{H(\alpha)} = \frac{y}{x}, \quad \text{with } \alpha = \frac{1 - 2\bar{\xi}}{2\bar{\xi}'}.$$

Therefore $\bar{\lambda} \in \operatorname{dom} \Gamma \cap \bar{\mathcal{D}}$ if and only if $\frac{y}{x} \in [\alpha_{\min}, \alpha_{\max}]$ and in this case $I_{\mathbf{f}}(z) = \Gamma^*(z)$.

We now turn to the proof of Proposition 4.1(3).

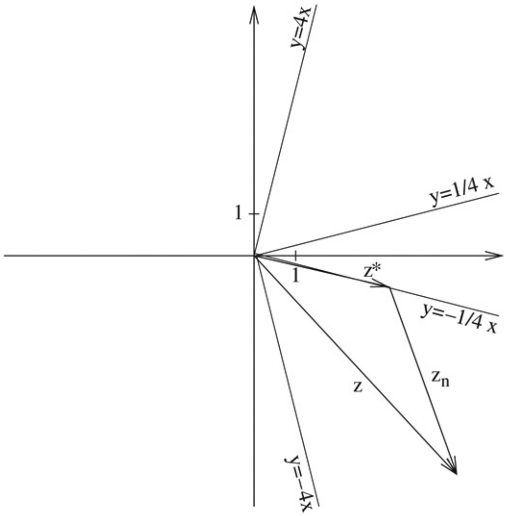

Fig. 4. For a $z = (x, y)$ such that $x_{\min}x < y < \alpha_{\min}x$, we decompose $z = z^* + z_n$ with z^* such that $y^* = \alpha_{\min}x^*$ and $z_n = t(1, x_{\min})$, for a $t > 0$.

From Theorem 3.6, we just need to exhibit a decomposition $z = z^* + z_{\mathbf{n}}$, where $z^* \in \partial\Gamma(\lambda^*)$ and $z_{\mathbf{n}} \in N_{\bar{\mathcal{D}}}(\lambda^*)$ for some $\lambda^* \in \operatorname{dom}\Gamma \cap \bar{\mathcal{D}}$. In this case, the value of $I_{\mathbf{f}}(z)$ is given by $I_{\mathbf{f}}(z) = \Gamma^*(z^*) + \langle\lambda^*, z_{\mathbf{n}}\rangle$. One can check that $\operatorname{dom}\Gamma \cap \bar{\mathcal{D}}$ can be split into three subsets : the interior of $\mathcal{D}$, and the two half-lines $\{1-2\xi-2x_{\min}\xi' = 0, \xi < 1/2\}$ and $\{1-2\xi-2x_{\max}\xi' = 0, \xi < 1/2\}$. The normal cones to $\bar{\mathcal{D}}$ are then easy to determine:

– if $(\xi, \xi') \in \operatorname{int}\mathcal{D}$, then $N_{\bar{\mathcal{D}}}(\xi, \xi') = \{(0, 0)\}$,
– if $\xi < 1/2$ and $1 - 2\xi - 2x_{\min}\xi' = 0$, then $N_{\bar{\mathcal{D}}}(\xi, \xi') = \{t(1, x_{\min}), t \geq 0\}$,
– if $\xi < 1/2$ and $1 - 2\xi - 2x_{\max}\xi' = 0$, then $N_{\bar{\mathcal{D}}}(\xi, \xi') = \{t(1, x_{\max}), t \geq 0\}$.

These normal cones are represented by the arrows on Fig. 2 (right).

We can now conclude the proof of the third point of the proposition. If we choose

$$\lambda^* = \left(\frac{1}{2} - \frac{x_{\min}}{y - x_{\min}x}, \frac{1}{y - x_{\min}x}\right),$$
$$z^* = (H_{\min}(x_{\min}x - y), (x_{\min}H_{\min} - 1)(x_{\min}x - y)),$$
$$z_{\mathbf{n}} = z - z^*,$$

it is easy to check that this decomposition fulfills the required properties, i.e. $z^* \in \partial\Gamma(\lambda^*)$ and $z_{\mathbf{n}} \in N_{\bar{\mathcal{D}}}(\lambda^*)$ for some $\lambda^* \in \operatorname{dom}\Gamma \cap \bar{\mathcal{D}}$. Therefore,

$$I_{\mathbf{f}}(z) = \Gamma^*(z^*) + \langle\lambda^*, z_n\rangle = \Gamma^*(z^*) + \frac{1}{2}(x + H_{\min}(y - x_{\min}x)).$$

The decomposition $z = z^* + z_{\mathbf{n}}$ can be seen on Fig. 4.

The proof of Proposition 4.1(4) is very similar and is left to the reader. □

Remarks on the LDP and the spherical integral. We conclude this section with remarks related to the prime motivation of this study, namely the study of the asymptotics of spherical integrals.

We recall from [9] that the goal is to get the asymptotics of

$$I_n(A_n, B_n) = \int \mathrm{e}^{N\ \mathrm{Trace}(A_n U B_n U^*)} \mathrm{d}m_n(U), \tag{4.6}$$

where A_n and B_n are two real diagonal matrices and m_n is the Haar measure on the orthogonal group. Obtaining the asymptotic expansion of such integrals has major applications in statistics for instance. Indeed, the asymptotic expansion for the joint eigenvalue density of some deformed Wigner matrices can readily be deduced from the above integral.

In the case where A_n is of rank one, with a unique non-zero eigenvalue denoted by θ and where $B_n = \mathrm{diag}(x_i^n, 1 \le i \le n)$ where $\frac{1}{n}\sum \delta_{x_i^n}$ converges, the spherical integral can be written as

$$I_n(A_n, B_n) = \mathbb{E} \exp \left(n\theta \frac{\sum_{i=1}^n x_i^n X_i^2}{\sum_{i=1}^n X_i^2} \right), \tag{4.7}$$

where $\mathbb{E}$ is the expectation under the standard N-dimensional Gaussian measure.

A natural strategy to tackle the asymptotics of I_n is then to establish the LDP for the empirical measure L_n as studied in the previous example and to apply Varadhan's lemma to get the asymptotics of I_n (see [9, Theorem 6]).

Besides the fact that we fully recover the LDP result of [9], we believe that the representation of the rate function (Theorem 3.6) sheds new light on the role played by the largest and lowest eigenvalues in the asymptotics of the rank-one spherical integral: The very reason comes from the fact that the individual rate function of the particle $\frac{1}{n}\begin{pmatrix} x_1^2 \\ X_1^2 \end{pmatrix}$ fulfills the convexity assumption (Assumption A.2). This is in particular illustrated in Lemma 4.2.

In the forthcoming section, we study the LDP in the non-convex case, that is when Assumption A.2 is not fulfilled. This will lead to partial results in the study of the asymptotics of the spherical integral beyond the rank-one case.

5. The LDP in the non-convex case

There are several models which fulfill Assumption A.1 with a non-convex rate function. Take for instance the simple model $Z_1 = (X_1^2, Y_1^2, X_1 Y_1)$ where X_1 and Y_1 are independent standard Gaussian random variables. Denote by $\mathcal{C} = \{(x, y, z) \in \mathbb{R}^3,\ z = -\sqrt{xy} \text{ or } z = \sqrt{xy}\}$, then $\frac{Z_1}{n}$ satisfies the LDP with good rate function

$$I(x, y, z) = \frac{x}{2} + \frac{y}{2} + \Delta(z \mid \mathcal{C}) \quad \text{where } \Delta(z \mid \mathcal{C}) = \begin{cases} 0 & \text{if } z \in \mathcal{C} \\ \infty & \text{else} \end{cases},$$

which is highly non-convex. We will see that this kind of models arises in the study of spherical integrals and may give rise to interesting phenomena.

We give in this section an assumption over the set $A_n = \{x_i^n \in \mathcal{X}, 1 \le i \le n\}$ which ensures the LDP for L_n to hold. Although quite stringent, this assumption encompasses interesting models as we shall see. We then state the LDP.

Recall that $\mathcal{Y}$ is the support of the limiting probability R.

Assumption A.6. Assume that $\mathcal{X} \subset \mathbb{R}^p$ for a given integer p. Denote by $A_n = \{x_i^n \in \mathcal{X}, 1 \le i \le n\}$. Then there exists an integer T such that:

$$A_n = \tilde{A}_n \cup \bigcup_{\ell=1}^{T} \{x_{i_\ell}^n\}$$

where $\rho(\tilde{A}_n, \mathcal{Y})$ goes to zero as $n \to \infty$ while for $1 \le \ell \le T$,

$$x_{i_\ell}^n \xrightarrow[n\to\infty]{} x_\ell^\infty,$$

where the x_ℓ^∞'s do not belong to $\mathcal{Y}$.

Remark 5.1. Assumption A.6 implies that there exist a finite number of outliers $x_{i_\ell}^n$ that remain outside the support $\mathcal{Y}$ and that converge pointwise to a limit x_ℓ^∞.

Theorem 5.1. *Assume that $(Z_i)_{i\in\mathbb{N}}$ is a sequence of $\mathbb{R}^d$-valued i.i.d. random variables where Z_1 satisfies Assumption A.1. Assume that Assumptions A.5 and A.6 hold for the sequence $(x_i^n, 1 \le i \le n, n \ge 1)$. Then*

$$L_n = \frac{1}{n} \sum_{1}^{n} \mathbf{f}(x_i^n) \cdot Z_i$$

satisfies the LDP in $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ with good rate function

$$I_{\mathbf{f}}(z) = \inf \left\{ \Gamma^*(z_0) + \sum_{\ell=1}^{T} I(y_\ell); z_0 + \sum_{\ell=1}^{T} \mathbf{f}(x_\ell^\infty) \cdot y_\ell = z \right\}. \tag{5.1}$$

Proof. Recall that $A_n = \tilde{A}_n \cup \bigcup_{\ell=1}^{T} \{x_{i_\ell}^n\}$ by Assumption A.6 and write:

$$L_n = \frac{1}{n} \sum_{x_i^n \in \tilde{A}_n} \mathbf{f}(x_i^n) \cdot Z_i + \frac{1}{n} \sum_{\ell=1}^{T} \mathbf{f}(x_{i_\ell}^n) \cdot Z_{i_\ell}.$$

One can prove the LDP for $\frac{1}{n} \sum_{x_i^n \in \tilde{A}_n} \mathbf{f}(x_i^n) \cdot Z_i$ as in the proof of Theorem 3.2 (which relies on an adaptation of Theorem 2.1 in [11] and does not involve the convexity of I). On the other hand, $\sum_{\ell=1}^{T} \frac{\mathbf{f}(x_{i_\ell}^n) \cdot Z_{i_\ell}}{n}$ is exponentially equivalent to $\sum_{\ell=1}^{T} \frac{\mathbf{f}(x_\ell^\infty) \cdot Z_{i_\ell}}{n}$ which satisfies the LDP with good rate function

$$J(z) = \inf \left\{ \sum_{\ell=1}^{T} I(y_\ell), \sum_{\ell=1}^{T} \mathbf{f}(x_\ell^\infty) \cdot y_\ell = z \right\}.$$

Since $\frac{1}{n} \sum_{x_i^n \in \tilde{A}_n} \mathbf{f}(x_i^n) \cdot Z_i$ and $\frac{1}{n} \sum_{\ell=1}^{T} \mathbf{f}(x_{i_\ell}^n) \cdot Z_{i_\ell}$ are independent, the LDP holds with good rate function $I_{\mathbf{f}}$ given by (5.1). Proof of Theorem 5.1 is completed. $\square$

6. An example of LDP in the non-convex case: Influence of the second largest eigenvalue

6.1. Presentation of the example

In this section, we shall study a simple model which underlines the differences between the LDP in the convex case and the LDP in the non-convex one. Consider the set $A_n = \{x_i^n,\ 1 \le$

$i \leq n\}$ where $x_1^n = \kappa_1$, $x_2^n = \kappa_2$ and $x_i^n = 1$ for $i \geq 3$. Assume the following:

$$1 < \kappa_2 < \kappa_1.$$

One can think of the x_i^n as the eigenvalues of an $n \times n$ matrix and one can check that

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i^n} \xrightarrow[n\to\infty]{} \delta_1$$

while κ_1 and κ_2 are two outliers.

In the sequel, we study the influence of the second largest eigenvalue κ_2 over the rate function of a given LDP in a convex and non-convex case. We prove that the second largest eigenvalue has no influence on the rate function that drives the LDP in the convex case (Proposition 6.1) while this eigenvalue has an impact on the LDP in the non-convex case (Proposition 6.2). We finally go back to spherical integrals and make some concluding remarks.

Denote by $\mathbf{f}$ the following matrix-valued function:

$$\mathbf{f}(x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let us now introduce the random variables that we will consider.

6.2. The convex model

Consider a family of $\mathbb{R}^3$-valued random variables $(Z_i)_{i\geq 1}$ satisfying Assumptions A.1 and A.2. Denote by

$$\begin{aligned} L_n(Z) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{f}(x_i^n) \cdot Z_i \\ &= \frac{1}{n}\mathbf{f}(\kappa_1) \cdot Z_1 + \frac{1}{n}\mathbf{f}(\kappa_2) \cdot Z_2 + \frac{1}{n}\sum_{i=3}^{n} \mathbf{f}(x_i^n) \cdot Z_i \\ &\triangleq \pi_n^1(Z) + \pi_n^2(Z) + \tilde{L}_n(Z) \end{aligned}$$

and by $\bar{L}_n(Z) \triangleq \pi_n^1(Z) + \tilde{L}_n(Z)$.

One can apply Theorem 3.2 to $L_n(Z)$ and $\bar{L}_n(Z)$ which therefore satisfy LDPs with given rate functions that we denote respectively by I_Z and $\bar{I}_Z$.

Proposition 6.1. *The rate functions I_Z and $\bar{I}_Z$ related to the LDPs of $L_n(Z)$ and $\bar{L}_n(Z)$ are equal.*

Remark 6.1. This proposition underlines the fact that the second largest eigenvalue does not have any influence on the rate function of the LDP.

Proof. Let

$$Z_i = \begin{pmatrix} U_i \\ V_i \\ W_i \end{pmatrix} \quad \text{then } \mathbf{f}(x) \cdot Z_i = \begin{pmatrix} U_i \\ V_i \\ xU_i \\ xV_i \\ W_i \end{pmatrix}.$$

For $\lambda \in \mathbb{R}^5$, denote by

$$\Lambda(\lambda) = \ln \mathbb{E} e^{\langle \lambda, \mathbf{f}(1) \cdot Z \rangle},$$
$$\Lambda_i(\lambda) = \ln \mathbb{E} e^{\langle \lambda, \mathbf{f}(\kappa_i) \cdot Z \rangle}, \quad i \in \{1, 2\}.$$

Consider also the associated domains:

$$\mathcal{D}_0 = \{\lambda \in \mathbb{R}^5;\ \Lambda(\lambda) < \infty\},$$
$$\mathcal{D}_i = \{\lambda \in \mathbb{R}^5;\ \Lambda_i(\lambda) < \infty\}, \quad i \in \{1, 2\}.$$

Remark that

$$\lambda = (\alpha, \beta, \gamma, \delta, \theta) \in \mathcal{D}_i \Leftrightarrow \lambda_i = (\alpha, \beta, \kappa_i \gamma, \kappa_i \delta, \theta) \in \mathcal{D}_0, \quad i \in \{1, 2\}. \tag{6.1}$$

From Theorem 3.2, we know that

$$I_Z(z) = \sup_{\lambda \in \mathcal{D}_0 \cap \mathcal{D}_1 \cap \mathcal{D}_2} \{\langle \lambda, z \rangle - \Lambda(\lambda)\} \quad \text{and} \quad \bar{I}_Z(z) = \sup_{\lambda \in \mathcal{D}_0 \cap \mathcal{D}_1} \{\langle \lambda, z \rangle - \Lambda(\lambda)\}.$$

We now prove that $\lambda \in \mathcal{D}_0 \cap \mathcal{D}_1$ implies that $\lambda \in \mathcal{D}_2$. Let $\lambda = (\alpha, \beta, \gamma, \delta, \theta) \in \mathcal{D}_0 \cap \mathcal{D}_1$.

From (6.1),

$$\lambda \in \mathcal{D}_1 \Rightarrow \lambda_1 = (\alpha, \beta, \kappa_1 \gamma, \kappa_1 \delta, \theta) \in \mathcal{D}_0.$$

Moreover, as $1 < \kappa_2 < \kappa_1$, κ_2 can be written as $\kappa_2 = a + b\kappa_1$, with a, b non-negative and $a + b = 1$. Due to the convexity of $\mathcal{D}_0$, we have that $a\lambda + b\lambda_1 \in \mathcal{D}_0$. On the other hand,

$$a\lambda + b\lambda_1 = (\alpha, \beta, \kappa_2 \gamma, \kappa_2 \delta, \theta),$$

so that $\lambda \in \mathcal{D}_2$ by (6.1). Therefore,

$$\begin{aligned} I_Z(z) &= \sup_{\lambda \in \mathcal{D}_0 \cap \mathcal{D}_1 \cap \mathcal{D}_2} \{\langle \lambda, z \rangle - \Lambda(\lambda)\} \\ &= \sup_{\lambda \in \mathcal{D}_0 \cap \mathcal{D}_1} \{\langle \lambda, z \rangle - \Lambda(\lambda)\} = \bar{I}_Z(z) \end{aligned}$$

and the proof of Proposition 6.1 is completed. □

6.3. *The non-convex model*

Let $(X_i)_{i \geq 1}$ and $(Y_i)_{i \geq 1}$ be two independent families of i.i.d. standard Gaussian random variables and consider the i.i.d. $\mathbb{R}^3$-valued random variables

$$\check{Z}_i = \begin{pmatrix} X_i^2 \\ Y_i^2 \\ X_i Y_i \end{pmatrix}.$$

We shall study the LDP of

$$\begin{aligned} L_n(\check{Z}) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{f}(x_i^n)\cdot \check{Z}_i \\ &= \frac{1}{n}\begin{pmatrix} X_1^2 \\ Y_1^2 \\ \kappa_1 X_1^2 \\ \kappa_1 Y_1^2 \\ X_1 Y_1 \end{pmatrix} + \frac{1}{n}\begin{pmatrix} X_2^2 \\ Y_2^2 \\ \kappa_2 X_2^2 \\ \kappa_2 Y_2^2 \\ X_2 Y_2 \end{pmatrix} + \frac{1}{n}\sum_{i=3}^{n}\begin{pmatrix} X_i^2 \\ Y_i^2 \\ X_i^2 \\ Y_i^2 \\ X_i Y_i \end{pmatrix} \\ &\triangleq \pi_n^1(\check{Z}) + \pi_n^2(\check{Z}) + \tilde{L}_n(\check{Z}). \end{aligned}$$

As above, we also introduce $\bar{L}_n(\check{Z}) = \pi_n^1(\check{Z}) + \tilde{L}_n(\check{Z})$.

The non-convex model satisfies assumptions of Theorem 5.1. Therefore, both $L_n(\check{Z})$ and $\bar{L}_n(\check{Z})$ satisfy the LDP with given rate functions that we denote respectively by $I_{\check{Z}}$ and $\bar{I}_{\check{Z}}$.

We shall prove the following:

Proposition 6.2. *Let $\kappa_1 < 2\kappa_2 - 1$. The rate function $I_{\check{Z}}$ that drives the LDP for $L_n(\check{Z})$ differs from the rate function $\bar{I}_{\check{Z}}$ that drives the LDP for $\bar{L}_n(\check{Z})$.*

Remark 6.2. Proposition 6.2 illustrates the influence of the second largest eigenvalue on the rate function of the LDP in the non-convex case. Note that the condition $\kappa_1 < 2\kappa_2 - 1$ is merely technical and yields to easier computations.

Proof. In order to prove Proposition 6.2, we shall prove that there exists some point $z^\star$ such that

$$I_{\check{Z}}(z^\star) < \infty \quad \text{while } \bar{I}_{\check{Z}}(z^\star) = \infty.$$

Denote by $z = (x, y, x', y', r)$ and by $\mathcal{A}$ the convex set

$$\mathcal{A} = \{z \in \mathbb{R}^5; x > 0, y > 0, x' = x, y' = y, r^2 \le xy\}.$$

Then Cramér's theorem yields the LDP for $\tilde{L}_n(\check{Z})$ with good rate function

$$\Gamma^*(z) = \frac{x+y}{2} - \frac{1}{2}\log(xy - r^2) + \Delta(z \mid \mathcal{A}).$$

Denote by $\mathcal{B}_\kappa$ the following non-convex set:

$$\mathcal{B}_\kappa = \{z \in \mathbb{R}^5; x > 0, y > 0, x' = \kappa x, y' = \kappa y, |r| = \sqrt{xy}\}.$$

One can prove that $\pi_n^1(\check{Z})$ and $\pi_n^2(\check{Z})$ satisfy the LDP with respective rate functions

$$I_1(z) = \frac{x+y}{2} + \Delta(z \mid \mathcal{B}_{\kappa_1}) \quad \text{and} \quad I_2(z) = \frac{x+y}{2} + \Delta(z \mid \mathcal{B}_{\kappa_2}).$$

The contraction principle then yields

$$\begin{aligned} I_{\check{Z}}(z) &= \inf_{z_0+z_1+z_2=z} \{\Gamma^*(z_0) + I_1(z_1) + I_2(z_2)\} \\ \bar{I}_{\check{Z}}(z) &= \inf_{z_0+z_1=z} \{\Gamma^*(z_0) + I_1(z_1)\}. \end{aligned}$$

Let $z^\star = (1, 1, \kappa_2, \kappa_2, 0)$ then we shall prove that

$$I_{\check{Z}}(z^\star) < \infty \quad \text{while } \bar{I}_{\check{Z}}(z^\star) = \infty. \tag{6.2}$$

This will complete the proof of Proposition 6.2.

In the sequel, we use the notation $z_i = (x_i, y_i, x'_i, y'_i, r_i)$ with $i \in \{0, 1, 2\}$. From the definition of $\bar{I}_{\check{Z}}$, one can easily check that $\bar{I}_{\check{Z}}(z^\star)$ is finite iff the following system of equations:

$$\begin{cases} x_0 + x_1 = 1 \\ y_0 + y_1 = 1 \\ x_0 + \kappa_1 x_1 = \kappa_2 \\ y_0 + \kappa_1 y_1 = \kappa_2 \\ x_1 y_1 < x_0 y_0 \end{cases} \tag{6.3}$$

has a solution such that $x_0 > 0$, $y_0 > 0$, $x_1 > 0$ and $y_1 > 0$. From easy computations, such a solution should satisfy

$$x_0 = \frac{\kappa_1 - \kappa_2}{\kappa_1 - 1} = y_0. \tag{6.4}$$

On the other hand, the last equation of (6.3) implies that $(1 - x_0)^2 < x_0^2$, that is $x_0 > \frac{1}{2}$. As we have assumed that $\kappa_1 < 2\kappa_2 - 1$, this is not compatible with (6.4) and

$$\bar{I}_{\check{Z}}(z^\star) = \infty.$$

We now prove that $I_{\check{Z}}(z^\star) < \infty$. The mere definition of $I_{\check{Z}}$ yields that $I_{\check{Z}}(z^\star) < \infty$ iff there exists a solution to the following system

$$\begin{cases} x_0 + x_1 + x_2 = 1 \\ y_0 + y_1 + y_2 = 1 \\ x_0 + \kappa_1 x_1 + \kappa_2 x_2 = \kappa_2 \\ y_0 + \kappa_1 y_1 + \kappa_2 y_2 = \kappa_2 \\ r_0^2 + \epsilon_1 x_1 y_1 + \epsilon_2 x_2 y_2 = 0 \end{cases} \tag{6.5}$$

satisfying $x_0 > 0$, $y_0 > 0$, $x_1 > 0$, $y_1 > 0$, $x_2 > 0$, $y_2 > 0$, $\epsilon_{1,2} = \pm 1$ and $r_0^2 \le x_0 y_0$.

We can easily check that this system admits the following solution:

$$\begin{aligned} x_0 &= y_0 = \frac{\kappa_1 - \kappa_2}{\kappa_1 + \kappa_2 - 2}, \\ x_1 &= y_1 = \frac{\kappa_2 - 1}{\kappa_1 + \kappa_2 - 2} = x_2 = y_2, \\ \epsilon_1 &= -\epsilon_2 = -1 \quad \text{and} \quad r_0 = 0. \end{aligned}$$

Therefore, (6.2) is proved. $\square$

6.4. Links with the spherical integral beyond the rank-one case

When one wants to study the asymptotics of the spherical integral in the case when the matrix A_n in (4.6) is of finite rank larger than one, one is led to study the Large Deviations for empirical means which do not fulfill the convexity assumption (Assumption A.2). For example, in the rank-two case, the related empirical mean to look at is given by:

$$L_n^{(2)} = \frac{1}{n}\sum \mathbf{f}^{(2)}(x_i^n)\cdot Z_i, \quad \text{with } Z_i = \begin{pmatrix} X_i^2 \\ Y_i^2 \\ X_iY_i \end{pmatrix} \quad \text{and} \quad \mathbf{f}^{(2)}(x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & 1 \\ 0 & 0 & x \end{pmatrix}$$

and Theorem 5.1 applies whenever Assumption A.6 is fulfilled. It is then an easy application of Varadhan's lemma to get the convergence of the spherical integrals in the rank-two case (and analogously for an arbitrary finite rank). The example studied in Section 6.3 supports the feeling (although in a very indirect way) that the asymptotics of the spherical integral in this case should depend not only on the largest eigenvalue (as proved in the rank-one case in [9]) but also on the second largest eigenvalue and maybe on other ones, the number of which is related to the rank of A_n. Unfortunately, the very intricate formula of the rate function associated with the LDP in the non-convex case gives little clue on how to relate the asymptotics of the spherical integral to the largest eigenvalues beyond the rank-one case.

Appendix. Proof of Lemma 2.4

Proof. Let $\varepsilon > 0$ be fixed. Note that $C_\infty^{\mathbf{f}} \neq \emptyset$ by Assumption A.3. Since $C_\infty^{\mathbf{f}}$ exists by (2.4) and is compact by Assumption A.3, there exists a finite number of $m \times d$ matrices $(\mathbf{a}_1, \dots, \mathbf{a}_p)$ such that

$$C_\infty^{\mathbf{f}} \subset \bigcup_{k=1}^{p} B(\mathbf{a}_k, \varepsilon) \quad \text{where } B(\mathbf{a}_k, \varepsilon) = \{\mathbf{y} \in \mathbb{R}^{m\times d}, |\mathbf{y} - \mathbf{a}_k| < \varepsilon\}.$$

From the cover $(B(\mathbf{a}_k, \varepsilon), 1 \le k \le p)$, one can easily build a partition $(\Gamma_k, 1 \le k \le p')$ where $p' \le p$ with the following properties:

- $C_\infty^{\mathbf{f}} \subset \cup_{k=1}^{p'} \Gamma_k$,
- $\sup\{|\mathbf{x} - \mathbf{x}'|, (\mathbf{x}, \mathbf{x}') \in \Gamma_k^2\} \le 2\varepsilon$,
- $\mathrm{int}(\Gamma_k) \cap C_\infty^{\mathbf{f}} \neq \emptyset$ for $1 \le k \le p'$ (in particular $\mathrm{int}(\Gamma_k) \neq \emptyset$).

Let $\mathbf{b}_{k,\varepsilon}$ be an element of $\mathrm{int}(\Gamma_k) \cap C_\infty^{\mathbf{f}}$. Denote by

$$\mathbf{f}^\varepsilon(x) = \sum_{k=1}^{p'} \mathbf{b}_{k,\varepsilon} 1_{\Gamma_k}(\mathbf{f}(x)), \quad x \in \mathcal{X} \quad \text{and} \quad \mathcal{D}^\varepsilon = \bigcap_{k=1}^{p} \mathcal{D}_{\mathbf{b}_{k,\varepsilon}}.$$

We will prove in the sequel the following facts:

(1) The partial weighted empirical mean π_n^ε defined by

$$\pi_n^\varepsilon = \frac{1}{n} \sum_{x_i^n \in C_n} \mathbf{f}^\varepsilon(x_i^n) \cdot Z_i$$

satisfies the LDP with good rate function $\Delta^*(z \mid \mathcal{D}^\varepsilon) = \sup\{\langle z, \lambda\rangle, \lambda \in \mathcal{D}^\varepsilon\}$.

(2) The family of random variables $(\pi_n^\varepsilon, \varepsilon > 0)$ is an exponential approximation of (π_n), i.e.

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{|\pi_n^\varepsilon - \pi_n| > \delta\} = -\infty, \quad \forall \delta > 0.$$

(3) Finally, the family $(\pi_n, n \geq 1)$ satisfies the LDP with good rate function $\Delta^*(z \mid \mathcal{D})$.

Let us first prove fact (1).

$$\pi_n^\varepsilon = \frac{1}{n} \sum_{x_i^n \in C_n} \mathbf{f}^\varepsilon(x_i^n) \cdot Z_i = \frac{\mathbf{b}_{1,\varepsilon}}{n} \cdot \sum_{\{x_i^n, \mathbf{f}(x_i^n) \in \Gamma_1\}} Z_i + \cdots + \frac{\mathbf{b}_{p',\varepsilon}}{n} \cdot \sum_{\{x_i^n, \mathbf{f}(x_i^n) \in \Gamma_{p'}\}} Z_i.$$

Since the sets (Γ_k) are disjoints, the partial empirical means $\frac{1}{n}\sum_{x_i^n \in \mathbf{f}^{-1}(\Gamma_k)} Z_i$ are independent. Denote by $\phi_k(n)$ the cardinality of the set $\{x_i^n, \mathbf{f}(x_i^n) \in \Gamma_k\}$. One has to check that

$$\lim_{n\to\infty} \frac{\phi_k(n)}{n} = 0 \quad \text{and} \quad \phi_k(n) \geq 1 \quad \text{for } n \text{ large enough.}$$

Since $\phi_k(n) \leq \operatorname{card}(C_n)$, the first point is proved. Recall now that $\operatorname{int}(\Gamma_k) \cap C_\infty^{\mathbf{f}} \neq 0$. Thus condition (2.4) yields that for n large enough, there always exist points of $C_n^{\mathbf{f}}$ that belong to Γ_k. In particular, $\phi_k(n) \geq 1$ eventually. Therefore, Lemma 2.3 yields the LDP for $\frac{1}{n}\sum_{x_i^n \in \mathbf{f}^{-1}(\Gamma_k)} Z_i$ with good rate function $I(y)$.

A straightforward application of the contraction principle [7, Theorem 4.2.1] yields the LDP for π_n^ε with good rate function

$$\Delta_\varepsilon^*(z) = \inf\left\{\sum_{k=1}^{p'} \Delta^*(y_k \mid \mathcal{D}_Z), \sum_{k=1}^{p'} \mathbf{b}_{k,\varepsilon} \cdot y_k = z\right\}.$$

We prefer the following representation which expresses the rate function Δ_ε^* as an inf-convolution:

$$\Delta_\varepsilon^*(z) = \inf\left\{\sum_{k=1}^{p'} \Delta^*(z_k \mid \mathcal{D}_{\mathbf{b}_{k,\varepsilon}}),\ \sum_{k=1}^{p'} z_k = z\right\}. \tag{A.1}$$

The rate function Δ_ε^* is lower semi-continuous therefore [13, Theorem 16.4] yields:

$$\begin{aligned}
\Delta_\varepsilon^* &= \sup_{\lambda \in \mathbb{R}^d}\left\{\langle\lambda, z\rangle - \sum_{1\leq k\leq p'} \Delta\left(z \mid \bar{\mathcal{D}}_{\mathbf{b}_{k,\varepsilon}}\right)\right\} \\
&= \sup_{\lambda \in \mathbb{R}^d}\left\{\langle\lambda, z\rangle - \Delta\left(z \,\middle|\, \bigcap_{1\leq k\leq p'} \bar{\mathcal{D}}_{\mathbf{b}_{k,\varepsilon}}\right)\right\} \\
&= \Delta^*\left(z \,\middle|\, \bigcap_{1\leq k\leq p'} \bar{\mathcal{D}}_{\mathbf{b}_{k,\varepsilon}}\right) \overset{(a)}{=} \Delta^*\left(z \,\middle|\, \bigcap_{1\leq k\leq p'} \mathcal{D}_{\mathbf{b}_{k,\varepsilon}}\right) = \Delta^*(z \mid \mathcal{D}_\varepsilon)
\end{aligned}$$

where (a) follows from Proposition 2.1. Fact (1) is proved.

Let us now prove fact (2). We have

$$\left|\pi_n^\varepsilon - \pi_n\right| \leq \frac{1}{n} \sum_{x_i^n \in C_n} \left|\mathbf{f}^\varepsilon(x_i^n) - \mathbf{f}(x_i^n)\right| |Z_i|.$$

By the definition of $\mathbf{f}^\varepsilon$, if $\mathbf{f}(x_i^n) \in \Gamma_k$ then $\mathbf{f}^\varepsilon(x_i^n) = \mathbf{b}_{k,\varepsilon}$ and $|\mathbf{f}(x_i^n) - \mathbf{b}_{k,\varepsilon}| \le 2\varepsilon$. Therefore $|\pi_n^\varepsilon - \pi_n| \le \frac{2\varepsilon}{n} \sum_{x_i^n \in C_n} |Z_i|$ and

$$\begin{aligned}\mathbb{P}\left\{|\pi_n^\varepsilon - \pi_n| > \delta\right\} &\le \mathbb{P}\left(\frac{1}{n}\sum_{x_i^n \in C_n} |Z_i| > \frac{\delta}{2\varepsilon}\right)\\ &\le \exp\left(-\frac{n\delta\kappa}{2\varepsilon}\right)\left(\mathbb{E}\,\mathrm{e}^{\kappa|Z_i|}\right)^{\mathrm{card}(C_n)}\end{aligned}$$

where $\kappa > 0$ is such that $\mathbb{E}\,\mathrm{e}^{\kappa|Z_i|} < \infty$. Therefore

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{|\pi_n^\varepsilon - \pi_n| > \delta\right\} \le -\frac{\kappa\delta}{2\varepsilon} \xrightarrow[\varepsilon\to 0]{} -\infty,$$

which proves the exponential equivalence. Fact (2) is proved.

We now prove fact (3). Since $(\pi_n^\varepsilon, \varepsilon > 0)$ is an exponential approximation of π_n, Theorem 4.2.16(a) in [7] implies that π_n satisfies a weak LDP with rate function given by:

$$\Upsilon(z) = \sup_{\delta>0} \liminf_{\varepsilon\to 0} \inf_{z'\in B(z,\delta)} \Delta_\varepsilon^*(z') \overset{(\star)}{=} \sup_{\delta>0} \limsup_{\varepsilon\to 0} \inf_{z'\in B(z,\delta)} \Delta_\varepsilon^*(z'),$$

where $(\star)$ is a by-product of the proof of [7, Theorem 4.2.16] (see Eq. 4.2.19 for instance). This precisely means that Υ is the epigraphical limit of Δ_ε^* (see [14, Chapter 7] for details). In order to prove that $\Upsilon = \Delta^*(\cdot \mid \mathcal{D})$, we first note that

$$\mathcal{D}^\varepsilon \xrightarrow[\varepsilon\to 0]{\mathrm{pk}} \overline{\mathcal{D}}.$$

A corollary [14, Corollary 11.35(a)] of Wijsman's theorem [14, Theorem 11.34] immediately yields:

$$\Upsilon(z) = \Delta^*(z \mid \overline{\mathcal{D}}) = \operatorname*{epi-lim}_{\varepsilon\to 0} \Delta^*(z \mid \mathcal{D}^\varepsilon), \tag{A.2}$$

where epi-lim denotes the epigraphical limit. Since $\Delta^*(z \mid \overline{\mathcal{D}}) = \Delta^*(z \mid \mathcal{D})$ by Proposition 2.1, we have $\Upsilon = \Delta^*(\cdot \mid \mathcal{D})$. Fact (3) is thus proved and so is Lemma 2.4. □

References

[1] G. Ben Arous, A. Dembo, A. Guionnet, Aging of spherical spin glasses, Probab. Theory Related Fields 120 (1) (2001) 1–67.

[2] J. Baik, G. Ben Arous, S. Péché, Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices, Ann. Probab. 33 (5) (2005) 1643–1697.

[3] J. Baik, J.W. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, J. Multivariate Anal. 97 (6) (2006) 1382–1408.

[4] B. Bercu, F. Gamboa, M. Lavielle, Sharp large deviations for Gaussian quadratic forms with applications, ESAIM Probab. Stat. 4 (2000) 1–24.

[5] B. Bercu, F. Gamboa, A. Rouault, Large deviations for quadratic functionals of stationary Gaussian processes, Stochastic Process. Appl. 71 (1997) 75–90.

[6] W. Bryc, A. Dembo, Large deviations for quadratic functionals of Gaussian processes, J. Theoret. Probab. 10 (1997) 307–332.

[7] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, second ed., Springer Verlag, New York, 1998.

[8] F. Gamboa, A. Rouault, M. Zani, A functional large deviations principle for quadratic forms of Gaussian stationary processes, Statist. Probab. Lett. 43 (1999) 299–308.

[9] A. Guionnet, M. Maïda, A Fourier view on the R-transform and related asymptotics of spherical integrals, J. Funct. Anal. 222 (2) (2005) 435–490.
[10] I.M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2) (2001) 295–327.
[11] J. Najim, A Cramér type theorem for weighted random variables, Electron. J. Probab. 7 (4) (2002) 32 (electronic).
[12] S. Péché, The largest eigenvalue of small rank perturbations of hermitian random matrices, Probab. Theory Related Fields 134 (1) (2006) 127–174.
[13] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1970.
[14] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, Springer, 1998.
[15] M. Zani, Grandes déviations pour des fonctionnelles issues de la statistique des processus, Ph.D. Thesis, Université Paris-Sud, 1999.