[14] H. W. Kuhn and J. W. Tucker, "Nonlinear programming," in *Proc. 2nd Berkeley Symp. Mathematical Statistics and Probability*. Berkeley, CA, 1950, pp. 481–492.

[15] N. Linial and N. Nisan, "Approximate inclusion-exclusion," *Combinatorica*, vol. 10, no. 4, pp. 349–365, 1990.

[16] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 1995, p. 46.

[17] D. Psaltis, R. Snapp, and S. Venkatesh, "On the finite sample performance of the nearest neighbor classifier," *IEEE Trans. Inform. Theory*, vol. 40, pp. 820–837, May 1994.

[18] W. Rogers and T. Wagner, "A finite-sample distribution-free performance bound for local discrimination rules," *Ann. Statist.*, vol. 6, pp. 506–514, 1978.

[19] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley.

[20] T. J. Wagner, "Convergence of the nearest neighbor rule," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 566–571, Sept. 1971.

Generalization of Gärtner–Ellis Theorem

Po-Ning Chen, *Member, IEEE*

***Abstract*—A generalization of the Gärtner–Ellis Theorem for *arbitrary random sequences* is established. It is shown that the conventional formula of the *large deviation rate function*, based on the moment generating function techniques, fails to describe the general (possibly *nonconvex*) large deviation rate for an *arbitrary* random sequence. An (*nonconvex*) extension formula obtained by *twisting* the conventional large deviation rate function around a continuous functional is therefore proposed. As a result, a new Gärtner–Ellis upper bound is proved. It is demonstrated by an example that a tight upper bound on the large deviation rate of an arbitrary random sequence can be obtained by choosing the right continuous functional, even if the true large deviation rate is not convex. Also proved is a parallel extension of the Gärtner–Ellis lower bound with the introduction of a new notion of *Gärtner–Ellis set* within which the upper bound coincides with the lower bound (for countably many points).**

***Index Terms*—Arbitrary random sequence, exponent, Gärtner–Ellis theorem, information spectrum, large deviations.**

I. Introduction

A general formula for the capacity of arbitrary single-user channels without feedback had been established by Verdú and Han in 1994 [2]. In their paper, the channel capacity was shown to be the supremum of input–output *inf-information rates* over all input processes, where the *inf-information rate* is defined as the *liminf in probability* of the normalized information density. This result was based on two key results: Feinstein's lemma [3] for the direct coding theorem and Verdú and Han's theorem [2, Theorem 4] for the converse coding theorem. The former provides a universal upper bound on average channel coding errors for every input process, and the latter gives a lower bound on the same quantity for the uniform input process over a reliable code sequence. While Feinstein's lemma was used in [2], a standard random coding argument can also be used for the achievability proof [5].

Specifically, the upper bound from Feinstein's lemma takes the form

$$P_e(n,R) \le \Pr\left[\frac{1}{n} i_{X^nW^n}(X^n;Y^n) \le R+\gamma\right] + e^{-\gamma n} \tag{1.1}$$

for every $\gamma > 0$ and every input distribution P_{X^n} on the input alphabet $\mathcal{X}^n$, where $P_e(n,R)$ represents the attainable average channel coding error, as a function of code length n and code rate R, and

$$i_{X^nW^n}(x^n;y^n) \triangleq \log \frac{P_{W^n}(y^n \mid x^n)}{P_{Y^n}(y^n)}$$

is the *information density* for a given channel transition probability $P_{W^n}(\cdot \mid x^n)$ and the output statistics P_{Y^n} due to the input P_{X^n}. The lower bound given by Verdú and Han has the shape

$$P_e(n,R) \ge \Pr\left[\frac{1}{n} i_{X^nW^n}(X^n;Y^n) \le R-\gamma\right] - e^{-\gamma n} \tag{1.2}$$

for any $\gamma > 0$ and every uniform input over a reliable code. These two bounds are shown to provide a *good* approximation to the maximum code rate R under the condition that the limsup of $P_e(n,R)$ in n equals zero [2]. By definition, this maximum code rate is the *channel capacity* C.

Comparing these two bounds with the result of the Gärtner–Ellis Theorem [1, p. 15], we see that the exponential rate (with respect to n) of the first term on the right-hand side of (1.1) or (1.2) can actually be carried out by letting the *sequence of random variables*, considered by Gärtner and Ellis, to be the *information density*. As a result of (1.1), the exponent of $P_e(n,R)$ in n under fixed $R \in (0,C)$ is bounded from below by both the magnitude of the *large deviation rate function* of the information density around $R+\gamma$ and the constant γ. We, therefore, pose a question "Whether the large deviation rate function for information density *still* provides a *good* approximation to the ultimate exponential dependence (in block length n) of $P_e(n,R)$ under fixed $R \in (0,C)$ for arbitrary single-user channels."

In studying this problem, we first observe that the information density for an arbitrary channel now becomes *arbitrary* in its statistics. Hence, the first step in this investigation is to generalize the large deviation rate function for arbitrary random sequences. Using the limsup and liminf of the log-moment generating functions, a simple extension of the Gärtner–Ellis Theorem for an arbitrary random sequence is established. However, such an extension, at times, is shown to yield a loose bound on the large deviation rate of an arbitrary random sequence, especially when the large deviation rate of the arbitrary random sequence is *not* convex (cf. Example 2.1). Since the large deviation rate function always leads to a convex function, chances of having a tight bound on the large deviation rate of an arbitrary random sequence along this line seem rare. Motivated by this, we then focus on finding a *nonconvex* expression for the large deviation rate.

The proof of the Gärtner–Ellis Theorem is, in fact, based on the Heine–Borel Theorem, which states that a finite subcover on a compact set exists for (uncountably many) open covers. The open covers in their proof take the form of $\{x \in \Re : \theta x - \bar{\varphi}(\theta) > a\}$ for $\theta \in \Re$ and $a \in \Re$. ($\sup_{\theta\in\Re}[\theta x - \bar{\varphi}(\theta)]$ is the *sup-large deviation rate function*, which will be defined in the next section.) These covers remain "open" when the argument x is replaced by any continuous function $h(x)$ over the real line. Such findings lead to a new extension of the Gärtner–Ellis Theorem. Examples will be given to demonstrate that by properly choosing a continuous function, a tight bound on the large deviation rate of an arbitrary random sequence can be obtained, even if it is not convex.

This correspondence is organized as follows. The extensions of Gärtner–Ellis upper and lower bounds are covered in Sections II and III, respectively. Examples will be given, following the theorems. In

Manuscript received April 16, 1999; revised April 22, 2000. This work was supported by the National Science Council in Taiwan, R.O.C., under the project code NSC 86-2213-E-009-115-T.

The author is with the Department of Communication Engineering, National Chiao Tung University, Taiwan, R.O.C. (e-mail: poning@cc.nctu.edu.tw).

Communicated by T. E. Fuja, Associate Editor At Large.
Publisher Item Identifier S 0018-9448(00)09685-1.

Section IV, properties of the (twisted) sup- and inf-large deviation rate functions are examined. Concluding remarks appear in Section V.

Throughout the correspondence, $\{Z_n\}_{n=1}^{\infty}$ denotes an infinite sequence of *arbitrary* random variables.

II. Extension of Gärtner–Ellis Upper Bounds

Definition 2.1: Define

$$\varphi_n(\theta) \triangleq \frac{1}{n}\log E[\exp\{\theta Z_n\}] \quad \text{and} \quad \bar{\varphi}(\theta) \triangleq \limsup_{n\to\infty} \varphi_n(\theta).$$

The *sup-large deviation rate function* of an arbitrary random sequence $\{Z_n\}_{n=1}^{\infty}$ is defined as

$$\bar{I}(x) \triangleq \sup_{\{\theta\in\Re:\bar{\varphi}(\theta)>-\infty\}} [\theta x - \bar{\varphi}(\theta)]. \tag{2.3}$$

The range of the supremum operation in (2.3) is always nonempty since $\bar{\varphi}(0)=0$, i.e., $\{\theta\in\Re : \bar{\varphi}(\theta) > -\infty\} \neq \emptyset$. Hence, $\bar{I}(x)$ is always defined. With the above definition, the first extension theorem of Gärtner–Ellis[1] can be proposed as follows.

Theorem 2.1: For $a, b \in \Re$ and $a \le b$

$$\limsup_{n\to\infty} \frac{1}{n}\log \Pr\left\{\frac{Z_n}{n} \in [a,b]\right\} \le -\inf_{x\in[a,b]} \bar{I}(x).$$

Proof: The proof follows directly from Theorem 2.2 by taking $h(x)=x$, and hence, we omit it. ■

The bound obtained in the above theorem is not in general tight. This can be easily seen by noting that for an *arbitrary* random sequence $\{Z_n\}_{n=1}^{\infty}$, the exponent of $\Pr\{Z_n/n \le b\}$ is not necessarily convex in b, and therefore, cannot be achieved by a *convex* (sup-)large deviation rate function. The next example further substantiates this argument.

Example 2.1: Suppose that $\Pr\{Z_n = 0\} = 1 - e^{-2n}$ and $\Pr\{Z_n = -2n\} = e^{-2n}$. Then, from Definition 2.1, we have

$$\varphi_n(\theta) \triangleq \frac{1}{n}\log E[e^{\theta Z_n}] = \frac{1}{n}\log\left[1 - e^{-2n} + e^{-(\theta+1)\cdot 2\cdot n}\right]$$

and

$$\bar{\varphi}(\theta) \triangleq \limsup_{n\to\infty}\varphi_n(\theta) = \begin{cases} 0, & \text{for } \theta \ge -1 \\ -2(\theta+1), & \text{for } \theta < -1. \end{cases}$$

Hence $\{\theta\in\Re : \bar{\varphi}(\theta) > -\infty\} = \Re$ and

$$\begin{aligned} \bar{I}(x) &= \sup_{\theta\in\Re}[\theta x - \bar{\varphi}(\theta)] \\ &= \sup_{\theta\in\Re}[\theta x + 2(\theta+1)\mathbf{1}\{\theta < -1)\}] \\ &= \begin{cases} -x, & \text{for } -2 \le x \le 0 \\ \infty, & \text{otherwise} \end{cases} \end{aligned}$$

where $\mathbf{1}\{\cdot\}$ represents the indicator function of a set. Consequently, by Theorem 2.1

$$\begin{aligned} \limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in[a,b]\right\} &\le -\inf_{x\in[a,b]}\bar{I}(x) \\ &= \begin{cases} 0, & \text{for } 0 \in [a,b] \\ b, & \text{for } b\in[-2,0] \\ -\infty, & \text{otherwise.} \end{cases} \end{aligned}$$ □

The exponent of $\Pr\{Z_n/n \in [a,b]\}$ in the above example is indeed given by

$$\lim_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in[a,b]\right\} = -\inf_{x\in[a,b]} I^*(x)$$

where

$$I^*(x) = \begin{cases} 2, & \text{for } x = -2 \\ 0, & \text{for } x = 0 \\ \infty, & \text{otherwise.} \end{cases} \tag{2.4}$$

Thus the upper bound obtained in Theorem 2.1 is not tight.

As mentioned earlier, the *looseness* of the upper bound in Theorem 2.1 cannot be improved by simply using a *convex* sup-large deviation rate function. Note that the true exponent (cf. (2.4)) of the above example is not a convex function. We then observe that the convexity of the sup-large deviation rate function is simply because it is a pointwise supremum of a collection of *affine* functions (cf. (2.3)). In order to obtain a better bound that achieves a *nonconvex* large deviation rate, the involvement of *nonaffine* functionals seems necessary. As a result, a new extension of the Gärtner–Ellis theorem is established along this line.

Before introducing the *nonaffine* extension of Gärtner–Ellis upper bound, we define the *twisted sup-large deviation rate function* as follows.

Definition 2.2: Define

$$\varphi_n(\theta;h) \triangleq \frac{1}{n}\log E\left[\exp\left\{n\cdot\theta\cdot h\left(\frac{Z_n}{n}\right)\right\}\right]$$

$$\text{and} \quad \bar{\varphi}_h(\theta) \triangleq \limsup_{n\to\infty}\varphi_n(\theta;h)$$

where $h(\cdot)$ is a given real-valued continuous function. The *twisted sup-large deviation rate function* of an arbitrary random sequence $\{Z_n\}_{n=1}^{\infty}$ with respect to a real-valued continuous function $h(\cdot)$ is defined as

$$\bar{J}_h(x) \triangleq \sup_{\{\theta\in\Re:\bar{\varphi}_h(\theta)>-\infty\}}[\theta\cdot h(x) - \bar{\varphi}_h(\theta)]. \tag{2.5}$$

Similarly to $\bar{I}(x)$, the range of the supremum operation in (2.5) is not empty, and hence, $\bar{J}_h(\cdot)$ is always defined.

Theorem 2.2: Suppose that $h(\cdot)$ is a real-valued continuous function. Then for $a, b\in\Re$ and $a \le b$

$$\limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in[a,b]\right\} \le -\inf_{x\in[a,b]}\bar{J}_h(x).$$

Proof: The proof is divided into two parts. Part 1 proves the result under

$$[a,b]\cap\{x\in\Re : \bar{J}_h(x) < \infty\} \neq \emptyset, \tag{2.6}$$

and Part 2 verifies it under

$$[a,b]\subset\{x\in\Re : \bar{J}_h(x) = \infty\}. \tag{2.7}$$

Since either (2.6) or (2.7) is true, these two parts, together, complete the proof of this theorem.

Part 1: **Assume** $[a,b]\cap\{x\in\Re : \bar{J}_h(x) < \infty\} \neq \emptyset$.

[1]For completeness, the conventional Gärtner–Ellis Theorem in [1, p. 15] is reproduced below.

Theorem (Gärtner–Ellis): If for all $\theta\in\Re$, $\varphi(\theta) = \limsup_{n\to\infty}\varphi_n(\theta) = \liminf_{n\to\infty}\varphi_n(\theta)$ and $[a,b]\cap\{x\in\Re : I(x) < \infty\}\neq\emptyset$, then

$$\limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in[a,b]\right\} \le -\inf_{x\in[a,b]} I(x).$$

If, in addition, $\varphi(\cdot)$ is differentiable on $\{\theta\in\Re : \varphi(\theta) < \infty\}$ and $(a,b)\subset\{x\in\Re : x = \varphi'(\theta) \text{ and } \varphi(\theta) < \infty \text{ for some } \theta\in\Re\}$, then

$$\liminf_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in(a,b)\right\} \ge -\inf_{x\in(a,b)} I(x)$$

where $I(x) \triangleq \sup_{\theta\in\Re}[\theta x - \varphi(\theta)]$.

Define $J^* \triangleq \inf_{x\in[a,b]} \bar{J}_h(x)$. By assumption, $J^* < \infty$. Therefore,

$$[a,b] \subset \{x : \bar{J}_h(x) > J^* - \varepsilon\}$$

for any $\varepsilon > 0$, and

$$\begin{aligned}
&\{x \in \Re : \bar{J}_h(x) > J^* - \varepsilon\} \\
&= \left\{ x \in \Re : \sup_{\{\theta\in\Re : \bar{\varphi}_h(\theta) > -\infty\}} [\theta h(x) - \bar{\varphi}_h(\theta)] > J^* - \varepsilon \right\} \\
&\subset \bigcup_{\{\theta\in\Re : \bar{\varphi}_h(\theta) > -\infty\}} \{x \in \Re : [\theta h(x) - \bar{\varphi}_h(\theta)] > J^* - \varepsilon\}.
\end{aligned}$$

Observe that

$$\bigcup_{\{\theta\in\Re : \bar{\varphi}_h(\theta) > -\infty\}} \{x \in \Re : [\theta x - \bar{\varphi}_h(\theta)] > J^* - \varepsilon\}$$

is a collection of (uncountably infinite) open sets that cover $[a,b]$ which is closed and bounded (and hence compact). By the Heine–Borel theorem, we can find a finite subcover such that

$$[a,b] \subset \bigcup_{i=1}^{k} \{x \in \Re : [\theta_i x - \bar{\varphi}_h(\theta_i)] > J^* - \varepsilon\}$$

and $(\forall 1 \le i \le k)\ \bar{\varphi}_h(\theta_i) < \infty$ (otherwise, the set

$$\{x : [\theta_i x - \bar{\varphi}_h(\theta_i)] > J^* - \varepsilon\}$$

is empty, and can be removed). Also note $(\forall 1 \le i \le k)\ \bar{\varphi}_h(\theta_i) > -\infty$. Consequently,

$$\begin{aligned}
&\Pr\left\{\frac{Z_n}{n} \in [a,b]\right\} \\
&leq \Pr\left\{\frac{Z_n}{n} \in \bigcup_{i=1}^{k} \{x : \theta_i h(x) - \bar{\varphi}_h(\theta_i)] > J^* - \varepsilon\}\right\} \\
&\le \sum_{i=1}^{k} \Pr\left\{h\left(\frac{Z_n}{n}\right) \cdot \theta_i - \bar{\varphi}_h(\theta_i) > J^* - \varepsilon\right\} \\
&= \sum_{i=1}^{k} \Pr\left\{n \cdot h\left(\frac{Z_n}{n}\right) \cdot \theta_i > n\bar{\varphi}_h(\theta_i) + n(J^* - \varepsilon)\right\} \\
&\le \sum_{i=1}^{k} \exp\{n[\bar{\varphi}_n(\theta_i; h) - \bar{\varphi}_h(\theta_i)] - n(J^* - \varepsilon)\}
\end{aligned}$$

where the last step follows from Markov's inequality. Since k is a constant independent of n, and for each integer $i \in [1,k]$

$$\limsup_{n\to\infty} \frac{1}{n} \log(\exp\{n[\bar{\varphi}_n(\theta_i; h) - \bar{\varphi}_h(\theta_i)] - n(J^* - \varepsilon)\}) = -(J^* - \varepsilon)$$

we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{\frac{Z_n}{n} \in [a,b]\right\} \le -(J^* - \varepsilon).$$

Since ε is arbitrary, the proof is completed.

Part 2: **Assume** $[a,b] \subset \{x \in \Re : \bar{J}_h(x) = \infty\}$.

Observe that $[a,b] \subset \{x : \bar{J}_h(x) > L\}$ for any $L > 0$. Following the same procedure as used in Part 1, we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{\frac{Z_n}{n} \in [a,b]\right\} \le -L.$$

Since L can be taken arbitrarily large

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{\frac{Z_n}{n} \in [a,b]\right\} = -\infty = -\inf_{x\in[a,b]} \bar{J}_h(x). \qquad \blacksquare$$

The proof of the above theorem has implicitly used the condition $-\infty < \bar{\varphi}_h(\theta_i) < \infty$ to guarantee that

$$\limsup_{n\to\infty} [\bar{\varphi}_n(\theta_i; h) - \bar{\varphi}_h(\theta_i)] = 0$$

for each integer $i \in [1,k]$. Note that when $\bar{\varphi}_h(\theta_i) = \infty$ (resp., $-\infty$), $\limsup_{n\to\infty} [\bar{\varphi}_n(\theta_i; h) - \bar{\varphi}_h(\theta_i)] = -\infty$ (resp., ∞). This explains why the range of the supremum operation in (2.5) is taken to be $\{\theta \in \Re : \bar{\varphi}_h(\theta) > -\infty\}$, instead of the whole real line.

As indicated in Theorem 2.2, a better upper bound can possibly be found by *twisting* the large deviation rate function around an appropriate (nonaffine) functional on the real line. Such improvement is substantiated in the next example.

Example 2.2: Let us, again, investigate the $\{Z_n\}_{n=1}^{\infty}$ defined in Example 2.1. Take

$$h(x) = \frac{1}{2}(x+2)^2 - 1.$$

Then, from Definition 2.2, we have

$$\begin{aligned}
\varphi_n(\theta; h) &\triangleq \frac{1}{n} \log E[\exp\{n\theta h(Z_n/n)\}] \\
&= \frac{1}{n} \log[\exp\{n\theta\} - \exp\{n(\theta-2)\} + \exp\{-n(\theta+2)\}]
\end{aligned}$$

and

$$\bar{\varphi}_h(\theta) \triangleq \limsup_{n\to\infty} \varphi_n(\theta; h) = \begin{cases} -(\theta+2), & \text{for } \theta \le -1 \\ \theta, & \text{for } \theta > -1. \end{cases}$$

Hence, $\{\theta \in \Re : \bar{\varphi}_h(\theta) > -\infty\} = \Re$ and

$$\bar{J}_h(x) \triangleq \sup_{\theta\in\Re} [\theta h(x) - \bar{\varphi}_h(\theta)] = \begin{cases} -\frac{1}{2}(x+2)^2 + 2, & \text{for } x \in [-4,0] \\ \infty, & \text{otherwise.} \end{cases}$$

Consequently, by Theorem 2.2, we get (2.8) at the top of following page.

For $b \in (-2,0)$ and $a \in [-2-\sqrt{2b-4}, b)$, the upper bound attained in the previous example is strictly less than that given in Example 2.1, and hence, an improvement is obtained. However, for $b \in (-2,0)$ and $a < -2-\sqrt{2b-4}$, the upper bound in (2.8) is actually looser. Accordingly, we combine the two upper bounds from Examples 2.1 and 2.2 to get

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{\frac{Z_n}{n} \in [a,b]\right\} &\le -\max\left\{\inf_{x\in[a,b]} \bar{J}_h(x),\ \inf_{x\in[a,b]} \bar{I}(x)\right\} \\
&= \begin{cases} 0, & \text{for } 0 \in [a,b] \\ \frac{1}{2}(b+2)^2 - 2, & \text{for } b \in [-2,0] \\ -\infty, & \text{otherwise.} \end{cases}
\end{aligned}$$

A *better* bound on the exponent of $\Pr\{Z_n/n \in [a,b]\}$ is thus obtained. As a result, Theorem 2.2 can be further generalized as follows.

Theorem 2.3: For $a, b \in \Re$ and $a \le b$

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{\frac{Z_n}{n} \in [a,b]\right\} \le -\inf_{x\in[a,b]} \bar{J}(x)$$

where $\bar{J}(x) \triangleq \sup_{h\in\mathcal{H}} \bar{J}_h(x)$ and $\mathcal{H}$ is the set of all real-valued continuous functions.

Proof: By redefining $J^* \triangleq \inf_{x\in[a,b]} \bar{J}(x)$ in the proof of Theorem 2.2, and observing that

$$\begin{aligned}
[a,b] &\subset \{x \in \Re : \bar{J}(x) > J^* - \varepsilon\} \\
&\subset \bigcup_{h\in\mathcal{H}} \bigcup_{\{\theta\in\Re : \bar{\varphi}_h(\theta) > -\infty\}} \{x \in \Re : [\theta h(x) - \bar{\varphi}_h(\theta)] > J^* - \varepsilon\}
\end{aligned}$$

$$\limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in[a,b]\right\}\le-\inf_{x\in[a,b]}\bar{J}_h(x)$$
$$=\begin{cases}-\min\left\{-\frac{1}{2}(a+2)^2+2,-\frac{1}{2}(b+2)^2+2\right\}, & \text{for } -4\le a<b\le 0\\ 0, & \text{for } a>0 \text{ or } b<-4\\ -\infty, & \text{otherwise}\end{cases} \tag{2.8}$$

□

the theorem holds under $[a,b]\cap\{x\in\Re:\bar{J}(x)<\infty\}\neq\emptyset$. Similar modifications to the proof of Theorem 2.2 can be applied to the case of $[a,b]\subset\{x\in\Re:\bar{J}(x)=\infty\}$. ■

Example 2.3: Let us again study the $\{Z_n\}_{n=1}^{\infty}$ in Example 2.1 (also in Example 2.2). Suppose $c>1$. Take $h_c(x)=c_1(x+c_2)^2-c$, where

$$c_1\triangleq\frac{c+\sqrt{c^2-1}}{2}\quad\text{and}\quad c_2\triangleq\frac{2\sqrt{c+1}}{\sqrt{c+1}+\sqrt{c-1}}.$$

Then from Definition 2.2, we have

$$\begin{aligned}\varphi_n(\theta;h_c)&\triangleq\frac{1}{n}\log E\left[\exp\left\{n\theta h_c\left(\frac{Z_n}{n}\right)\right\}\right]\\&=\frac{1}{n}\log[(1-p_n)\exp\{n\theta\}+p_n\exp\{-n\theta\}]\\&=\frac{1}{n}\log[\exp\{n\theta\}-\exp\{n(\theta-2)\}+\exp\{-n(\theta+2)\}]\end{aligned}$$

and

$$\bar{\varphi}_{h_c}(\theta)\triangleq\limsup_{n\to\infty}\varphi_n(\theta;h_c)=\begin{cases}-(\theta+2), & \text{for } \theta\le-1\\ \theta, & \text{for } \theta>-1.\end{cases}$$

Hence, $\{\theta\in\Re:\bar{\varphi}_{h_c}(\theta)>-\infty\}=\Re$ and

$$\begin{aligned}\bar{J}_{h_c}(x)&=\sup_{\theta\in\Re}[\theta h_c(x)-\bar{\varphi}_{h_c}(\theta)]\\&=\begin{cases}-c_1(x+c_2)^2+c+1, & \text{for } x\in[-2c_2,0]\\ \infty, & \text{otherwise.}\end{cases}\end{aligned}$$

From Theorem 2.3

$$\bar{J}(x)=\sup_{h\in\mathcal{H}}\bar{J}_h(x)\ge\max\left\{\liminf_{c\to\infty}\bar{J}_{h_c}(x),\bar{I}(x)\right\}=I^*(x)$$

where $I^*(x)$ is defined in (2.4). Consequently,

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in[a,b]\right\}&\le-\inf_{x\in[a,b]}\bar{J}(x)\\&\le-\inf_{x\in[a,b]}I^*(x)\\&=\begin{cases}0, & \text{if } 0\in[a,b]\\ -2, & \text{if } -2\in[a,b] \text{ and } 0\notin[a,b]\\ -\infty, & \text{otherwise}\end{cases}\end{aligned}$$

and a tight upper bound is finally obtained! □

Theorem 2.2 gives us the upper bound on the limsup of $(1/n)\log\Pr\{Z_n/n\in[a,b]\}$. With the same technique, we can also obtain a parallel theorem for the quantity

$$\liminf_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in[a,b]\right\}.$$

Definition 2.3: Define $\underline{\varphi}_h(\theta)\triangleq\liminf_{n\to\infty}\varphi_n(\theta;h)$, where $\varphi_n(\theta;h)$ was defined in Definition 2.2. The *twisted inf-large deviation rate function* of an arbitrary random sequence $\{Z_n\}_{n=1}^{\infty}$ with respect to a real-valued continuous function $h(\cdot)$ is defined as

$$\underline{J}_h(x)\triangleq\sup_{\{\theta\in\Re:\underline{\varphi}_h(\theta)>-\infty\}}[\theta\cdot h(x)-\underline{\varphi}_h(\theta)].$$

Theorem 2.4: For $a,b\in\Re$ and $a\le b$,

$$\liminf_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in[a,b]\right\}\le-\inf_{x\in[a,b]}\underline{J}(x)$$

where $\underline{J}(x)\triangleq\sup_{h\in\mathcal{H}}\underline{J}_h(x)$ and $\mathcal{H}$ is the set of all real-valued continuous functions.

III. Extension of Gärtner–Ellis Lower Bounds

The tightness of the upper bound given in Theorem 2.2 naturally relies on the validity of

$$\limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in(a,b)\right\}\ge-\inf_{x\in(a,b)}\bar{J}_h(x) \tag{3.9}$$

which is an extension of the Gärtner–Ellis lower bound. The above inequality, however, is not in general true for all choices of a and b (cf. Case A of Example 3.4). It, therefore, becomes significant to find those (a,b) within which the extended Gärtner–Ellis lower bound holds.

Definition 3.4: Define the *sup-Gärtner–Ellis set* with respect to a real-valued continuous function $h(\cdot)$ as

$$\bar{\mathcal{G}}_h\triangleq\bigcup_{\{\theta\in\Re:\bar{\varphi}_h(\theta)>-\infty\}}\bar{\mathcal{G}}(\theta;h)$$

where

$$\bar{\mathcal{G}}(\theta;h)\triangleq\left\{x\in\Re:\limsup_{t\downarrow0}\frac{\bar{\varphi}_h(\theta+t)-\bar{\varphi}_h(\theta)}{t}\le h(x)\le\liminf_{t\downarrow0}\frac{\bar{\varphi}_h(\theta)-\bar{\varphi}_h(\theta-t)}{t}\right\}.$$

Let us briefly remark on the *sup-Gärtner–Ellis set* defined above. It is self-explanatory in its definition that $\bar{\mathcal{G}}_h$ is always defined for any real-valued function $h(\cdot)$. Furthermore, it can be derived that the *sup-Gärtner–Ellis set* is reduced to

$$\bar{\mathcal{G}}_h\triangleq\bigcup_{\{\theta\in\Re:\bar{\varphi}_h(\theta)>-\infty\}}\{x\in\Re:\bar{\varphi}'_h(\theta)=h(x)\}$$

if the derivative $\bar{\varphi}'_h(\theta)$ exists for all θ. Observe that the condition $h(x)=\bar{\varphi}'_h(\theta)$ is exactly the equation for finding the θ that achieves $\bar{J}_h(x)$, which is obtained by taking the derivative of $[\theta h(x)-\bar{\varphi}_h(\theta)]$.

This somehow hints that the *sup-Gärtner–Ellis set* is a collection of those points at which the exact sup-large deviation rate is achievable.

We now state the main theorem in this section.

Theorem 3.5: Suppose that $h(\cdot)$ is a real-valued continuous function. Then if $(a,b) \subset \bar{\mathcal{G}}_h$

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{ \frac{Z_n}{n} \in \mathcal{J}_h(a,b) \right\} \geq - \inf_{x\in(a,b)} \bar{J}_h(x)$$

where

$$\mathcal{J}_h(a,b) \triangleq \{y \in \Re : h(y) = h(x) \text{ for some } x \in (a,b)\}.$$

Proof: Let $F_n(\cdot)$ denote the distribution function of Z_n. Define its *extended twisted distribution* around the real-valued continuous function $h(\cdot)$ as

$$\begin{aligned} dF_n^{(\theta;h)}(\cdot) &\triangleq \frac{\exp\{n\,\theta\,h(x/n)\}\,dF_n(\cdot)}{E[\exp\{n\,\theta\,h(Z_n/n)\}]} \\ &= \frac{\exp\{n\,\theta\,h(x/n)\}\,dF_n(\cdot)}{\exp\{n\varphi_n(\theta;h)\}}. \end{aligned}$$

Let $Z_n^{(\theta:h)}$ be the random variable having $F_n^{(\theta;h)}(\cdot)$ as its probability distribution. Let

$$J^* \triangleq \inf_{x\in(a,b)} \bar{J}_h(x).$$

Then for any $\varepsilon > 0$, there exists $v \in (a,b)$ with $\bar{J}_h(v) \leq J^* + \varepsilon$.

Now the continuity of $h(\cdot)$ implies that

$$\mathcal{B}(v,\delta) \triangleq \{x \in \Re : |h(x) - h(v)| < \delta\} \subset \mathcal{J}_h(a,b)$$

for some $\delta > 0$. Also, $(a,b) \subset \bar{\mathcal{G}}_h$ ensures the existence of θ satisfying

$$\limsup_{t\downarrow 0} \frac{\bar{\varphi}_h(\theta+t) - \bar{\varphi}_h(\theta)}{t} \leq h(v) \leq \liminf_{t\downarrow 0} \frac{\bar{\varphi}_h(\theta) - \bar{\varphi}_h(\theta-t)}{t}$$

which, in turn, guarantees the existence of $t = t(\delta) > 0$ satisfying

$$\frac{\bar{\varphi}_h(\theta+t) - \bar{\varphi}_h(\theta)}{t} \leq h(v) + \frac{\delta}{4} \quad \text{and} \quad h(v) - \frac{\delta}{4} \leq \frac{\bar{\varphi}_h(\theta) - \bar{\varphi}_h(\theta-t)}{t}. \tag{3.10}$$

We then derive

$$\begin{aligned} \Pr\left\{ \frac{Z_n}{n} \in \mathcal{J}_h(a,b) \right\} &\geq \Pr\left\{ \frac{Z_n}{n} \in \mathcal{B}(v,\delta) \right\} \\ &= \Pr\left\{ \left| h\left(\frac{Z_n}{n}\right) - h(v) \right| < \delta \right\} \\ &= \int_{\{x\in\Re:|h(x/n)-h(v)|<\delta\}} dF_n(x) \\ &= \int_{\{x\in\Re:|h(x/n)-h(v)|<\delta\}} \\ &\quad \times \exp\left\{ n\varphi_n(\theta;h) - n\theta h\left(\frac{x}{n}\right) \right\} dF_n^{(\theta;h)}(x) \\ &\geq \exp\{n\varphi_n(\theta;h) - n\theta h(v) - n|\theta|\delta\} \\ &\quad \times \int_{\{x\in\Re:|h(x/n)-h(v)|<\delta\}} dF_n^{(\theta;h)}(x) \\ &= \exp\{n\varphi_n(\theta;h) - n\theta h(v) - n|\theta|\delta\} \\ &\quad \times \Pr\left\{ \frac{Z_n^{(\theta;h)}}{n} \in \mathcal{B}(v,\delta) \right\} \end{aligned}$$

which implie

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{ \frac{Z_n}{n} \in \mathcal{J}_h(a,b) \right\} &\geq -[\theta h(v) - \bar{\varphi}_h(\theta)] - |\theta|\,\delta \\ &\quad + \limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{ \frac{Z_n^{(\theta;h)}}{n} \in \mathcal{B}(v,\delta) \right\} \\ &= -\bar{J}_h(v) - |\theta|\delta + \limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{ \frac{Z_n^{(\theta;h)}}{n} \in \mathcal{B}(v,\delta) \right\} \\ &\geq -J^* - \varepsilon - |\theta|\delta + \limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{ \frac{Z_n^{(\theta;h)}}{n} \in \mathcal{B}(v,\delta) \right\}. \end{aligned}$$

Since both δ and ε can be made arbitrarily small, it remains to show

$$\limsup_{n\to\infty} \frac{1}{n} \log \Pr\left\{ \frac{Z_n^{(\theta;h)}}{n} \in \mathcal{B}(v,\delta) \right\} = 0. \tag{3.11}$$

To show (3.11), we first note that

$$\begin{aligned} \Pr\left\{ h\left(\frac{Z_n^{(\theta;h)}}{n} \right) \geq h(v) + \delta \right\} &= \Pr\left\{ e^{nth\left(Z_n^{(\theta;h)}/n\right)} \geq e^{nth(v)+nt\delta} \right\} \\ &\leq e^{-nth(v)-nt\delta} \int_{\Re} e^{nth(x/n)} dF_n^{(\theta;h)} \\ &= e^{-nth(v)-nt\delta} \int_{\Re} e^{nth(x/n)+n\theta h(x/n)-n\varphi_n(\theta;h)} dF_n(x) \\ &= e^{-nth(v)-nt\delta-n\varphi_n(\theta;h)+n\varphi_n(\theta+t;h)}. \end{aligned}$$

Similarly,

$$\begin{aligned} \Pr\left\{ h\left(\frac{Z_n^{(\theta;h)}}{n} \right) \leq h(v) - \delta \right\} &= \Pr\left\{ e^{-nth\left(Z_n^{(\theta;h)}/n\right)} \geq e^{-nth(v)+nt\delta} \right\} \\ &\leq e^{nth(v)-nt\delta} \int_{\Re} e^{-nth(x/n)} dF_n^{(\theta;h)} \\ &= e^{nth(v)-nt\delta} \int_{\Re} e^{-nth(x/n)+n\theta h(x/n)-n\varphi_n(\theta;h)} dF_n(x) \\ &= e^{nth(v)-nt\delta-n\varphi_n(\theta;h)+n\varphi_n(\theta-t;h)}. \end{aligned}$$

Now by definition of limsup

$$\varphi_n(\theta+t;h) \leq \bar{\varphi}_h(\theta+t) + \frac{t\delta}{4} \quad \text{and} \quad \varphi_n(\theta-t;h) \leq \bar{\varphi}_h(\theta-t) + \frac{t\delta}{4} \tag{3.12}$$

for sufficiently large n; and

$$\varphi_n(\theta;h) \geq \bar{\varphi}_h(\theta) - \frac{t\delta}{4} \tag{3.13}$$

for infinitely many n. Hence, there exists a subsequence $\{n_1, n_2, n_3, \ldots\}$ such that for all n_j, (3.12) and (3.13) hold. Consequently, for all j

$$\begin{aligned}\frac{1}{n_j}\log\Pr\left\{\frac{Z_{n_j}^{(\theta;h)}}{n_j}\notin\mathcal{B}(v,\delta)\right\} &\le \frac{1}{n_j}\log\Big(2\\ &\quad\times\max\Big[e^{-n_jth(v)-n_jt\delta-n_j\varphi_{n_j}(\theta;h)+n_j\varphi_{n_j}(\theta+t;h)},\\ &\quad\times\ e^{n_jth(v)-n_jt\delta-n_j\varphi_{n_j}(\theta;h)+n_j\varphi_{n_j}(\theta-t;h)}\Big]\Big)\\ &= \frac{1}{n_j}\log 2+\max\big\{\big[-th(v)+\varphi_{n_j}(\theta+t;h),th(v)\\ &\qquad+\varphi_{n_j}(\theta-t;h)\big]\big\}-\varphi_{n_j}(\theta;h)-t\delta\\ &\le \frac{1}{n_j}\log 2+\max\{[-th(v)+\bar\varphi_h(\theta+t),th(v)\\ &\quad+\bar\varphi_h(\theta-t)]\}-\bar\varphi_h(\theta)-\frac{t\delta}{2}=\frac{1}{n_j}\log 2\\ &\quad+t\cdot\max\left\{\left[\frac{\bar\varphi_h(\theta+t)-\bar\varphi_h(\theta)}{t}-h(v),h(v)\right.\right.\\ &\qquad\left.\left.-\frac{\bar\varphi_h(\theta)-\bar\varphi_h(\theta-t)}{t}\right]\right\}-\frac{t\delta}{2}\\ &\le \frac{1}{n_j}\log 2-\frac{t\delta}{4}\end{aligned}\tag{3.14}$$

where (3.14) follows from (3.10). The proof is then completed by obtaining

$$\liminf_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n^{(\theta;h)}}{n}\notin\mathcal{B}(v,\delta)\right\}\le-\frac{t\delta}{4}$$

which immediately guarantees the validity of (3.11). ∎

Next, we use an example to demonstrate that by choosing the right $h(\cdot)$, we can completely characterize the exact (nonconvex) sup-large deviation rate $\bar I^*(x)$ for all $x\in\Re$.

Example 3.4: Suppose $Z_n=X_1+\cdots+X_n$, where $\{X_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.) Gaussian random variables with mean 1 and variance 1 if n is even, and with mean -1 and variance 1 if n is odd. Then the exact large deviation rate formula $\bar I^*(x)$ that satisfies for all $a<b$

$$\begin{aligned}-\inf_{x\in[a,b]}\bar I^*(x)&\ge\limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in[a,b]\right\}\\ &\ge\limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in(a,b)\right\}\ge-\inf_{x\in(a,b)}\bar I^*(x)\end{aligned}$$

is

$$\bar I^*(x)=\frac{(|x|-1)^2}{2}.\tag{3.15}$$

Case A: $h(x)=x$.

For the affine $h(\cdot)$, $\varphi_n(\theta)=\theta+\theta^2/2$ when n is even, and $\varphi_n(\theta)=-\theta+\theta^2/2$ when n is odd. Hence, $\bar\varphi(\theta)=|\theta|+\theta^2/2$, and

$$\begin{aligned}\bar{\mathcal{G}}_h&=\left(\bigcup_{\theta>0}\{v\in\Re: v=1+\theta\}\right)\bigcup\left(\bigcup_{\theta<0}\{v\in\Re: v=-1+\theta\}\right)\\ &\quad\bigcup\{v\in\Re: 1\le v\le-1\}\\ &=(1,\infty)\cup(-\infty,-1).\end{aligned}$$

Therefore, Theorem 3.5 cannot be applied to any a and b with $(a,b)\cap[-1,1]\neq\emptyset$.

By deriving

$$\bar I(x)=\sup_{\theta\in\Re}\{x\theta-\bar\varphi(\theta)\}=\begin{cases}\dfrac{(|x|-1)^2}{2}, & \text{for } |x|>1\\ 0, & \text{for } |x|\le 1\end{cases}$$

we obtain for any $a\in(-\infty,1)\cup(1,\infty)$

$$\begin{aligned}\lim_{\varepsilon\downarrow0}\limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in(a-\varepsilon,a+\varepsilon)\right\}\\ \ge-\lim_{\varepsilon\downarrow0}\inf_{x\in(a-\varepsilon,a+\varepsilon)}\bar I(x)=-\frac{(|a|-1)^2}{2}\end{aligned}$$

which can be shown tight by Theorem 2.2 (or directly by (3.15)). Note that the above inequality does not hold for any $a\in(-1,1)$. To fill the gap, a different $h(\cdot)$ must be employed.

Case B: $h(x)=|x-a|$ for $-1<a<1$.

For n even

$$\begin{aligned}E&\left[e^{n\theta h(Z_n/n)}\right]\\ &=E\left[e^{n\theta|Z_n/n-a|}\right]\\ &=\int_{-\infty}^{na}e^{-\theta x+n\theta a}\frac{1}{\sqrt{2\pi n}}e^{-(x-n)^2/(2n)}\,dx\\ &\quad+\int_{na}^{\infty}e^{\theta x-n\theta a}\frac{1}{\sqrt{2\pi n}}e^{-(x-n)^2/(2n)}\,dx\\ &=e^{n\theta(\theta-2+2a)/2}\int_{-\infty}^{na}\frac{1}{\sqrt{2\pi n}}e^{-[x-n(1-\theta)]^2/(2n)}\,dx\\ &\quad+e^{n\theta(\theta+2-2a)/2}\int_{na}^{\infty}\frac{1}{\sqrt{2\pi n}}e^{-[x-n(1+\theta)]^2/(2n)}\,dx\\ &=e^{n\theta(\theta-2+2a)/2}\cdot\Phi((\theta+a-1)\sqrt{n})\\ &\quad+e^{n\theta(\theta+2-2a)/2}\cdot\Phi((\theta-a+1)\sqrt{n})\end{aligned}$$

where $\Phi(\cdot)$ represents the unit Gaussian cumulative distribution function (cdf).

Similarly, for n odd

$$\begin{aligned}E\left[e^{n\theta h(Z_n/n)}\right]&=e^{n\theta(\theta+2+2a)/2}\cdot\Phi((\theta+a+1)\sqrt{n})\\ &\quad+e^{n\theta(\theta-2-2a)/2}\cdot\Phi((\theta-a-1)\sqrt{n}).\end{aligned}$$

Observe that for any $b\in\Re$

$$\lim_{n\to\infty}\frac{1}{n}\log\Phi(b\sqrt{n})=\begin{cases}0, & \text{for } b\ge0\\ -\dfrac{b^2}{2}, & \text{for } b<0.\end{cases}$$

Hence

$$\bar\varphi_h(\theta)=\begin{cases}-\dfrac{(|a|-1)^2}{2}, & \text{for } \theta<|a|-1\\ \dfrac{\theta[\theta+2(1-|a|)]}{2}, & \text{for } |a|-1\le\theta<0\\ \dfrac{\theta[\theta+2(1+|a|)]}{2}, & \text{for } \theta\ge0.\end{cases}$$

Therefore, we get the expressions at the top of following page. We then apply Theorem 3.5 to obtain

$$\begin{aligned}\lim_{\varepsilon\downarrow0}\limsup_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in(a-\varepsilon,a+\varepsilon)\right\}\\ \ge-\lim_{\varepsilon\downarrow0}\inf_{x\in(a-\varepsilon,a+\varepsilon)}\bar J_h(x)\\ =-\lim_{\varepsilon\downarrow0}\frac{(\varepsilon-1+|a|)^2}{2}=-\frac{(|a|-1)^2}{2}.\end{aligned}$$

Note that the above lower bound is valid for any $a\in(-1,1)$, and can be shown tight, again, by Theorem 2.2 (or directly by (3.15)).

$$\bar{\mathcal{G}}_h = \left(\bigcup_{\theta>0}\{x\in\Re : |x-a| = \theta+1+|a|\}\right)\bigcup\left(\bigcup_{\theta<0}\{x\in\Re : |x-a| = \theta+1-|a|\}\right)$$
$$= (-\infty, a-1-|a|)\cup(a-1+|a|, a+1-|a|)\cup(a+1+|a|,\infty)$$

and

$$\bar{J}_h(x) = \begin{cases} \frac{(|x-a|-1+|a|)^2}{2}, & \text{for } a-1+|a|<x<a+1-|a| \\ \frac{(|x-a|-1-|a|)^2}{2}, & \text{for } x>a+1+|a| \text{ or } x<a-1-|a| \\ 0, & \text{otherwise.} \end{cases} \tag{3.16}$$

Finally, by combining the results of Cases A and B, the true large deviation rate of $\{Z_n\}_{n\geq 1}$ is completely characterized. □

Remarks:

- One of the problems in applying the extended Gärtner–Ellis Theorems is the difficulty in choosing an appropriate real-valued continuous function (not to mention the finding of the optimal one in the sense of Theorem 2.3). From the previous example, we observe that the resultant $\bar{J}_h(x)$ is in fact equal to the lower convex contour[2] (with respect to $h(\cdot)$) of $\min_{\{y\in\Re: h(y)=h(x)\}}\bar{I}^*(x)$. Indeed, if the lower convex contour of $\min_{\{y\in\Re: h(y)=h(x)\}}\bar{I}^*(x)$ equals $\bar{I}^*(x)$ for some x lying in the interior of $\bar{\mathcal{G}}_h$, we can apply Theorems 2.2 and 3.5 to establish the large deviation rate at this point. From the above example, we somehow sense that taking $h(x)=|x-a|$ is advantageous in characterizing the large deviation rate at $x=a$. As a consequence of such choice of $h(\cdot)$, $\bar{J}_h(x)$ will shape like the lower convex contour of $\min\{\bar{I}^*(x–a), \bar{I}^*(a-x)\}$ in $h(x)=|x-a|$. Hence, if a lies in $\bar{\mathcal{G}}_h$, $\bar{J}_h(a)$ can surely be used to characterize the large deviation rate at $x=a$ (as it does in Case B of Example 3.4).
- The assumptions required by the conventional Gärtner-Ellis lower bound [1, p. 15] are

  1) $\varphi(\theta)=\bar{\varphi}(\theta)=\underline{\varphi}(\theta)$ exists;
  2) $\varphi(\theta)$ is differentiable on its domain; and
  3) $(a,b)\subset\{x\in\Re : x=\varphi'(\theta) \text{ for some } \theta\}$.

  The above assumptions are somewhat of limited use for arbitrary random sequences, since they do not in general hold. For example, the condition of $\bar{\varphi}(\theta)\neq\underline{\varphi}(\theta)$ is violated in Example 3.4.
- By using the limsup and liminf operators in our extension theorem, the sup-Gärtner–Ellis set is always defined without any requirement on the log-moment generating functions. The sup-Gärtner–Ellis set also clearly indicates the range in which the Gärtner–Ellis lower bound holds. In other words, $\bar{\mathcal{G}}_h$ is a subset of the union of all (a,b) for which the Gärtner–Ellis lower bound is valid. This is concluded in the following equation:

$$\bar{\mathcal{G}}_h \subset \bigcup\left\{(a,b) : \limsup_{n\to\infty}\frac{1}{n}\log\Pr\left[\frac{Z_n}{n}\in\mathcal{J}_h(a,b)\right] \geq -\inf_{x\in(a,b)}\bar{J}_h(x)\right\}.$$

  To verify whether or not the above two sets are equal merits further investigation.

- Modifying the proof of Theorem 3.5, we can also establish a lower bound for

$$\liminf_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in\mathcal{J}_h(a,b)\right\}.$$

Definition 3.5: Define the *inf-Gärtner–Ellis set* with respect to a real-valued continuous function $h(\cdot)$ as

$$\underline{\mathcal{G}}_h \triangleq \bigcup_{\{\theta\in\Re:\underline{\varphi}_h(\theta)>-\infty\}}\underline{\mathcal{G}}(\theta;h)$$

where

$$\underline{\mathcal{G}}(\theta;h)\triangleq\left\{x\in\Re : \limsup_{t\downarrow 0}\frac{\underline{\varphi}_h(\theta+t)-\underline{\varphi}_h(\theta)}{t} \leq h(x) \leq \liminf_{t\downarrow 0}\frac{\underline{\varphi}_h(\theta)-\underline{\varphi}_h(\theta-t)}{t}\right\}.$$

Theorem 3.6: Suppose that $h(\cdot)$ is a real-valued continuous function. Then if $(a,b)\subset\underline{\mathcal{G}}_h$

$$\liminf_{n\to\infty}\frac{1}{n}\log\Pr\left\{\frac{Z_n}{n}\in\mathcal{J}_h(a,b)\right\}\geq-\inf_{x\in(a,b)}\underline{J}_h(x).$$

- One of the important usages of the large deviation rate functions is to find the Varadhan's asymptotic integration formula of $\lim_{n\to\infty}(1/n)\log E[\exp\{\theta Z_n\}]$ for a given random sequence $\{Z_n\}_{n=1}^\infty$. To be specific, it is equal [4, Theorem 2.1.10] to

$$\lim_{n\to\infty}\frac{1}{n}\log E[\exp\{\theta Z_n\}] = \sup_{\{x\in\Re: I(x)<\infty\}}[\theta x - I(x)]$$

  if

$$\lim_{L\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log\left[\int_{[x\in\Re:\theta x\geq L]}\exp\{\theta x\}\,dP_{Z_n}(x)\right]=-\infty.$$

  The above result can also be extended using the same idea as applied to the Gärtner–Ellis theorem.

Theorem 3.7: If

$$\lim_{L\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log\left[\int_{[x\in\Re:\theta h(x)\geq L]}\exp\{\theta h(x)\}\,dP_{Z_n}(x)\right]=-\infty$$

then

$$\limsup_{n\to\infty}\frac{1}{n}\log E\left[\exp\left\{n\theta h\left(\frac{Z_n}{n}\right)\right\}\right] = \sup_{\{x\in\Re:\bar{J}_h(x)<\infty\}}[\theta h(x)-\bar{J}_h(x)]$$

[2]We define that the lower convex contour of a function $f(\cdot)$ with respect to $h(\cdot)$ is the largest $g(\cdot)$ satisfying that $g(h(x))\leq f(x)$ for all x, and for every x, y and for all $\lambda\in[0,1]$, $\lambda g(h(x))+(1-\lambda)g(h(y))\geq g(\lambda h(x)+(1-\lambda)h(y))$.

and

$$\liminf_{n\to\infty} \frac{1}{n} \log\left[\exp\left\{n\theta h\left(\frac{Z_n}{n}\right)\right\}\right] = \sup_{\{x\in\Re:\underline{J}_h(x)<\infty\}} [\theta h(x) - \underline{J}_h(x)].$$

Proof: This can be obtained by modifying the proofs of Lemmas 2.1.7 and 2.1.8 in [4]. ■

We close the section by remarking that the result of the above theorem can be reformulated as

$$\bar{J}_h(x) = \sup_{\{\theta\in\Re:\bar{\varphi}_h(\theta)>-\infty\}} [\theta h(x) - \bar{\varphi}_h(\theta)]$$
$$\text{and}\quad \bar{\varphi}_h(\theta) = \sup_{\{x\in\Re:\bar{J}_h(x)<\infty\}} [\theta h(x) - \bar{J}_h(x)]$$

which is an extension of the *Legendre–Fenchel Transform pair*. A similar conclusion applies to $\underline{J}_h(x)$ and $\underline{\varphi}_h(\theta)$.

IV. Properties of (Twisted) Sup- and Inf-Large Deviation Rate Functions

Property 4.1: Let $\bar{I}(x)$ and $\underline{I}(x)$ be the sup- and inf-large deviation rate functions of an infinite sequence of arbitrary random variables $\{Z_n\}_{n=1}^{\infty}$, respectively. Denote $m_n = (1/n)E[Z_n]$. Let $\bar{m} \triangleq \limsup_{n\to\infty} m_n$ and $\underline{m} \triangleq \liminf_{n\to\infty} m_n$. Then

1) $\bar{I}(x)$ and $\underline{I}(x)$ are both convex;
2) $\bar{I}(x)$ is continuous over $\{x \in \Re : \bar{I}(x) < \infty\}$. Likewise, $\underline{I}(x)$ is continuous over $\{x \in \Re : \underline{I}(x) < \infty\}$;
3) $\bar{I}(x)$ gives its minimum value 0 at $\underline{m} \le x \le \bar{m}$;
4) $\underline{I}(x) \ge 0$. But $\underline{I}(x)$ does not necessary give its minimum value at both $x = \bar{m}$ and $x = \underline{m}$.

Proof:

1) $\bar{I}(x)$ is the pointwise supremum of a collection of affine functions. Therefore, it is convex. Similar argument can be applied to $\underline{I}(x)$.

2) A convex function on the real line is continuous everywhere on its domain and hence the property holds.

3) and 4) The proofs follow immediately from Property 4.2 by taking $h(x) = x$. ■

Since the twisted sup/inf-large deviation rate functions are not necessarily convex, a few properties of sup/inf-large deviation functions do not hold for general twisted functions.

Property 4.2: Suppose that $h(\cdot)$ is a real-valued continuous function. Let $\bar{J}_h(x)$ and $\underline{J}_h(x)$ be the corresponding twisted sup- and inf-large deviation rate functions, respectively. Denote

$$m_n(h) \triangleq E[h(Z_n/n)].$$

Let

$$\bar{m}_h \triangleq \limsup_{n\to\infty} m_n(h) \quad\text{and}\quad \underline{m}_h \triangleq \liminf_{n\to\infty} m_n(h).$$

Then

1) $\bar{J}_h(x) \ge 0$, with equality holds if $\underline{m}_h \le h(x) \le \bar{m}_h$.
2) $\underline{J}_h(x) \ge 0$, but $\underline{J}_h(x)$ does not necessary give its minimum value at both $x = \bar{m}_h$ and $x = \underline{m}_h$.

Proof:

1) For all $x \in \Re$,

$$\bar{J}_h(x) \triangleq \sup_{\{\theta\in\Re:\bar{\varphi}_h(\theta)>-\infty\}} [\theta\cdot h(x) - \bar{\varphi}_h(\theta)] \ge 0\cdot h(x) - \bar{\varphi}_h(0) = 0.$$

By Jensen's inequality

$$\begin{aligned}\exp\{n\varphi_n(\theta;h)\} &= E[\exp\{n\cdot\theta\cdot h(Z_n/n)\}]\\ &\ge \exp\{n\cdot\theta\cdot E[h(Z_n/n)]\}\\ &= \exp\{n\cdot\theta\cdot m_n(h)\}\end{aligned}$$

which is equivalent to

$$\theta\cdot m_n(h) \le \varphi_n(\theta;h).$$

After taking the limsup and liminf of both sides of the above inequalities, we obtain

- for $\theta \ge 0$

$$\theta\bar{m}_h \le \bar{\varphi}_h(\theta) \tag{4.17}$$

and

$$\theta\cdot\underline{m}_h \le \underline{\varphi}_h(\theta) \le \bar{\varphi}_h(\theta); \tag{4.18}$$

- for $\theta < 0$

$$\theta\underline{m}_h \le \bar{\varphi}_h(\theta) \tag{4.19}$$

and

$$\theta\cdot\bar{m}_h \le \underline{\varphi}_h(\theta) \le \bar{\varphi}_h(\theta). \tag{4.20}$$

Expressions (4.17) and (4.20) imply $\bar{J}_h(x) = 0$ for those x satisfying $h(x) = \bar{m}_h$, and (4.18) and (4.19) imply $\bar{J}_h(x) = 0$ for those x satisfying $h(x) = \underline{m}_h$. For

$$x \in \{x : \underline{m}_h \le h(x) \le \bar{m}_h\}$$

$$\theta\cdot h(x) - \bar{\varphi}_h(\theta) \le \theta\cdot\bar{m}_h - \bar{\varphi}_h(\theta) \le 0, \qquad \text{for } \theta \ge 0$$

and

$$\theta\cdot h(x) - \bar{\varphi}_h(\theta) \le \theta\cdot\underline{m}_h - \bar{\varphi}_h(\theta) \le 0, \qquad \text{for } \theta < 0.$$

Hence, by taking the supremum over $\{\theta \in \Re : \bar{\varphi}_h(\theta) > -\infty\}$, we obtain the desired result.

2) The nonnegativity of $\underline{J}_h(x)$ can be similarly proved as $\bar{J}_h(x)$.

For Case A of Example 3.4, we have $\bar{m} = 1$, $\underline{m} = -1$, and $\underline{\varphi}(\theta) = -|\theta| + \theta^2/2$. Therefore,

$$\underline{I}(x) = \sup_{\theta\in\Re}\{x\theta - \underline{\varphi}(\theta)\} = \frac{(|x|+1)^2}{2}$$

for which $\underline{I}(-1) = \underline{I}(1) = 2$ and $\min_{x\in\Re}\underline{I}(x) = \underline{I}(0) = 1/2$. Consequently, $\underline{I}(x)$ neither equals zero nor gives its minimum value at both $x = \bar{m}$ and $x = \underline{m}$. ■

V. Concluding Remarks

Our study on the large deviation rates for arbitrary random sequences has yielded new Gärtner–Ellis lower and upper bounds. No assumption on the statistics of the random sequence is required in these two bounds. The newly defined *Gärtner–Ellis set* has been shown to be (a subset of) the range under which our Gärtner–Ellis bounds are tight (for countably many points).

Two issues are still open in this study. The first one concerns a systematic methodology for finding a series of continuous functions for which the large deviation rates can be completely characterized. Our example somehow suggest that the convex continuous functions that bottom at the targeted range could be a proper choice. The second issue questions whether or not the Gärtner–Ellis set is the largest one in which the Gärtner–Ellis lower bound holds. It is our conjecture that the answer is affirmative. In [6], Poor and Verdú have provided an upper bound on the channel reliability of arbitrary single-user channels, which is of the form

$$-\limsup_{n\to\infty} \frac{1}{n} \sup_{X^n} \log \Pr\left\{\frac{1}{n} i_{X^n W^n}(X^n, Y^n) \le R\right\}. \tag{5.21}$$

They conjectured that (5.21) is in fact tight. It would also be interesting to evaluate (5.21) using the twisted large deviation rate function, and see if any twisted functional can provide improvement on the existing channel reliability bounds.

REFERENCES

[1] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation*. New York: Wiley, 1990.
[2] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.
[3] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 2–22, Jan. 1954.
[4] J.-D. Deuschel and D. W. Stroock, *Large Deviations*. San Diego, CA: Academic, 1989.
[5] T. S. Han, *Information-Spectrum Methods in Information Theory*. Tokyo, Japan: Baifukan, 1998.
[6] H. V. Poor and S. Verdú, "A lower bound on the error probability of multihypothesis testing," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1992–1994, Nov. 1995.

Lower Bounds of the Minimal Eigenvalue of a Hermitian Positive-Definite Matrix

Weiwei Sun

***Abstract*—In this correspondence, we present several lower bounds of the minimal eigenvalue of a class of Hermitian positive-definite matrices, which improve the previous bounds given by Dembo [1] and Ma and Zarowski [4].**

***Index Terms*—Eigenvalue bounds, Hermitian positive definite.**

I. INTRODUCTION

The study of lower bounds of the minimal eigenvalue of a Hermitian matrix is of wide interest in many fields [1]–[3]. Here we consider the $n \times n$ Hermitian positive-definite matrix R_n defined in a partition form by

$$R_n = \begin{bmatrix} R_{n-1} & b \\ b^H & c \end{bmatrix} \tag{1.1}$$

where $b \in \boldsymbol{C}^{n-1}$ is an $(n-1)$-dimensional complex vector, R_{n-1} is an $(n-1) \times (n-1)$ matrix, $c > 0$ is a scalar value, and the superscript H denotes Hermitian transposition. Lower bounds of the minimal eigenvalue of the R_n in (1.1) have been studied by Dembo [1] and Ma and Zarowski [4] in terms of the minimal eigenvalue of R_{n-1}.

Manuscript received April 1, 1999; revised April 18, 2000. This work was supported in part by the City University of Hong Kong under Research Grant 7000929.

The author is with the Department of Mathematics, University of Hong Kong, Kowloon, Hong Kong, China (e-mail: maweiw@math.cityu.edu.hk).

Communicated by T. E. Fuja, Associate Editor At Large.
Publisher Item Identifier S 0018-9448(00)09681-4.

The motivation of this correspondence is to present several new bounds for the minimal eigenvalues. The approach is based on a matrix series and the local monotonicity of the minimal eigenvalue. The previous lower bounds can be obtained as special cases. The new bounds presented here are more accurate than both Dembo's bound and Ma and Zarowski's bound.

II. PREVIOUS RESULTS

Let $\lambda_j, j = 1, 2, \ldots, n$, be the eigenvalues of R_n and

$$0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n.$$

Dembo's result is given below.

Theorem A (Dembo [1]): Let η_1 be a lower bound of the minimal eigenvalue of R_{n-1} and η_{n-1} be an upper bound of the maximal eigenvalue of R_{n-1}. Then

$$\underline{\lambda}_A := \frac{c+\eta_1}{2} - \frac{\sqrt{(c+\eta_1)^2 - 4(c\eta_1 - b^H b)}}{2} \le \lambda_1 \tag{2.1}$$

and

$$\bar{\lambda}_A := \frac{c+\eta_{n-1}}{2} - \frac{\sqrt{(c+\eta_{n-1})^2 - 4(c\eta_{n-1} - b^H b)}}{2} \ge \lambda_1.$$

The bounds given in the above theorem are easy to calculate. The result has been extended to positive-semidefinite doubly symmetric matrices and applied for some Toeplitz problems in [1]. It has been noted in [4] that the lower bound in (2.1) can be negative for some Hermitian positive-definite matrices. An example of susch a matrix, given in [4], is

$$R_3 = \begin{bmatrix} 1+\epsilon & 1 & 1 \\ 1 & 1+\epsilon & 1 \\ 1 & 1 & 1+\epsilon \end{bmatrix}$$

which is Hermitian positive-definite for $\epsilon > 0$ and $\lambda_1 = \epsilon$. By Theorem A, we obtain the bound $\underline{\lambda}_A = -1 + \epsilon$.

An improved bound has been given by Ma and Zarowski [4]. Their main result is summarized in the following theorem.

Theorem B (Ma and Zarowski [4]):

$$\lambda_B := \frac{c+\eta_1}{2} - \frac{\sqrt{(c+\eta_1)^2 - 4\left(c - b^H R_{n-1}^{-1} b\right)\eta_1}}{2} \le \lambda_1. \tag{2.2}$$

Since $c - b^H R_{n-1}^{-1} b = \det(R_n)/\det(R_{n-1})$ where $\det(\cdot)$ denotes the determinant, the bound in (2.2) is positive when R_n is Hermitian positive-definite.

III. LOWER BOUNDS OF THE MINIMAL EIGENVALUE

We consider the characteristic polynomial of the matrix R_n, which is given by

$$\det(R_n - \lambda I) = \det\left(\begin{bmatrix} R_{n-1} - \lambda I & b \\ b^H & c - \lambda \end{bmatrix}\right).$$

Since R_n is Hermitian positive-definite, $\lambda_1 \le \eta_1 \le \lambda_2$. For convenience, first we consider the case

$$\lambda_1 < \eta_1$$

and denote by η_1 the minimal eigenvalue of R_{n-1} instead of a lower bound of the minimal eigenvalue, as in Theorems A and B. Then, $R_{n-1} - \lambda I$ is invertible for $\lambda < \eta_1$. We have

$$\det(R_n - \lambda I) = \det(R_{n-1} - \lambda I)(c - \lambda - b^H(R_{n-1} - \lambda I)^{-1} b).$$