ECOLE POLYTECHNIQUE

CENTRE DE MATHÉMATIQUES APPLIQUÉES

UMR CNRS 7641

91128 PALAISEAU CEDEX (FRANCE). Tél: 01 69 33 41 50. Fax: 01 69 33 30 11

http://www.cmap.polytechnique.fr/

A Contraction Principle For Closed Graph Transformations.

Christian Léonard

R.I. N^0 507 *June 2003*

A CONTRACTION PRINCIPLE FOR CLOSED GRAPH TRANSFORMATIONS

CHRISTIAN LÉONARD

Abstract. One proves a contraction principle for a transformation with a closed graph and present two simple applications of this result.

1. Introduction

One proves at Theorem 2.1 a contraction principle for a transformation with a closed graph. The price to pay for this relaxation from the usual Contraction Principle with a continuous transformation is the assumption that the transformed random system is exponentially tight. We present two simple applications of this result. The first one, in the domain of self-normalized sequences, indicates that one should not worry with dividing by zero when this almost never occurs! We hope that Theorem 2.1 may bring some comfort when proving such results. In our second application, one relaxes a topological assumption on the state space when proving large deviations for the empirical measures associated with U-statistics. The state space may only be metric and separable without being complete. Of course, one might not expect striking improvements when using the "closed graph version" of the contraction principle instead of the standard "continuous" version. But maybe this easy result could be useful in some technically intricated situations. (Please, let me know if you find one).

2. The contraction principle

Let $\{X_n\}_{n\geq 1}$ be a collection of random elements taking their values in a topological space $\mathcal{X}$ endowed with some σ-field. We assume that $\{X_n\}$ obeys the Large Deviation Principle (LDP) with rate function $I : \mathcal{X} \to [0, \infty]$. As usual (see [2]), this means that for every closed measurable set F and every open measurable set G

$$\begin{aligned} & \limsup_{n\to\infty} \frac{1}{n} \log I\!P(X_n \in F) \leq -I(F) \\ -I(G) \quad \leq \quad & \liminf_{n\to\infty} \frac{1}{n} \log I\!P(X_n \in G) \end{aligned}$$

where we used the notation $I(A) = \inf_{x\in A} I(x)$.
Let $f : \mathcal{X} \to \mathcal{Y}$ be a transformation from $\mathcal{X}$ to another topological space $\mathcal{Y}$ endowed with some σ-field. We are interested in the large deviation behavior of

$$Y_n \triangleq f(X_n)$$

where it is assumed that Y_n is measurable. Our main result is the following

Date: June 2003.
1991 *Mathematics Subject Classification.* Primary: , Secondary:
Key words and phrases. Large deviations.

Theorem 2.1 (Contraction Principle). *Let us assume that* $f : \mathcal{X} \to \mathcal{Y}$ *has a* closed graph *and that* $\{Y_n\}$ *is exponentially tight in* $\mathcal{Y}$. *Then,* $\{Y_n\}$ *obeys the LDP in* $\mathcal{Y}$ *with the rate function* $J(y) = \inf\{I(x); x \in \mathcal{X}, f(x) = y\}$, $y \in \mathcal{Y}$.

We shall need the following lemma to prove the theorem.

Lemma 2.2. *Let* $f : \mathcal{X} \to \mathcal{Y}$ *be a function with a closed graph. For any compact subset* K *of* $\mathcal{Y}$, $f^{-1}(K) = \{x \in \mathcal{X}; f(x) \in K\}$ *is a closed subset of* $\mathcal{X}$.

In order to apply the present contraction principle in the context of U-statistics, it is not assumed that the topologies are first countable. As a consequence, the next proof is given in terms of nets rather than sequences. For more details about nets, see for instance ([3], Sections IV.2 and IV.3). When working with closed and compact sets, the recipe is "Use nets as you use sequences".

Proof. Let $\{x_\alpha\}_{\alpha \in \mathcal{I}}$ be a net in $f^{-1}(K)$ such that $x_\alpha \to x \in \mathcal{X}$. We have to show that $f(x) \in K$.
We have $y_\alpha \triangleq f(x_\alpha) \in K$ for all $\alpha \in \mathcal{I}$. As K is compact, there exists a convergent subnet $\{y_\beta\}_{\beta \in \mathcal{J}} : y_\beta \to y \in K$. Hence, $x_\beta \to x$ and $y_\beta = f(x_\beta) \to y$. As the graph of f is closed, we get $y = f(x)$. Hence $f(x) \in K$, which is the desired result. □

Proof of Theorem 2.1. It is assumed that $\{Y_n\}$ is exponentially tight: for all $L \geq 0$, there exists a measurable compact subset K_L of $\mathcal{Y}$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log I\!P(Y_n \notin K_L) \leq -L. \tag{1}$$

Let us begin with the *upper bound.* Let F be a closed subset of $\mathcal{Y}$. For any $L \geq 0$, we have

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log I\!P(Y_n \in F) &\leq \limsup_{n\to\infty} \frac{1}{n} \log[I\!P(Y_n \in F \cap K_L) + I\!P(Y_n \notin K_L)] \\
&\leq \max[\limsup_{n\to\infty} \frac{1}{n} \log I\!P(Y_n \in F \cap K_L), -L]. \qquad (2)
\end{aligned}$$

By the above lemma, $f^{-1}(F \cap K_L)$ is a closed subset $\mathcal{X}$ and the LDP for $\{X_n\}$ leads us to

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log I\!P(Y_n \in F \cap K_L) &= \limsup_{n\to\infty} \frac{1}{n} \log I\!P(X_n \in f^{-1}(F \cap K_L)) \\
&\leq -\inf\{I(x); x \in f^{-1}(F \cap K_L)\} \\
&= -\inf_{y \in F \cap K_L} \inf\{I(x); x \in \mathcal{X}, f(x) = y\} \\
&\leq -\inf_{y \in F} \inf\{I(x); x \in \mathcal{X}, f(x) = y\} \\
&= -\inf_{y \in F} J(y).
\end{aligned}$$

Plugging this estimate in (2) and letting L tend to infinity leads to the desired result.
Let us go on with the *lower bound.* Let G be an open subet of $\mathcal{Y}$. We want to prove

$$\liminf_{n\to\infty} \frac{1}{n} \log I\!P(Y_n \in G) \geq -I(f^{-1}(G)). \tag{3}$$

Let $L \geq 0$, we define

$$U(G, L) \triangleq f^{-1}(G \cup K_L^c).$$

This set is open. Indeed, $U(G, L) = f^{-1}([G^c \cap K_L]^c) = f^{-1}(G^c \cap K_L)^c$ and $f^{-1}(G^c \cap K_L)$ is closed by Lemma 2.2.
The basic estimate is

$$I\!P(Y_n \in G) \geq I\!P(X_n \in U(G, L)) - I\!P(Y_n \notin K_L) \tag{4}$$

which holds since $U(G, L) = f^{-1}(G) \cup f^{-1}(K_L^c)$ so that $I\!P(X_n \in U(G, L)) \leq I\!P(Y_n \in G) + I\!P(Y_n \notin K_L)$.
Before applying the lower bound for $\{X_n\}$ to the open set $U(G, L)$ we need some more estimates. We have

$$I(f^{-1}(K_L)^c) \geq L, \forall L \geq 0. \tag{5}$$

Indeed, (1) is $-L \geq \limsup_{n\to\infty} \frac{1}{n} \log I\!P(X_n \in f^{-1}(K_L)^c)$ where by Lemma 2.2, $f^{-1}(K_L)^c$ is open. Thanks to the lower bound of the LDP for $\{X_n\}$, we get

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log I\!P(Y_n \notin K_L) &= \limsup_{n\to\infty} \frac{1}{n} \log I\!P(X_n \in f^{-1}(K_L)^c) \\
&\geq \liminf_{n\to\infty} \frac{1}{n} \log I\!P(X_n \in f^{-1}(K_L)^c) \\
&\geq -I(f^{-1}(K_L)^c),
\end{aligned}$$

which leads to (5).
We also need to know that if $I(f^{-1}(G)) < \infty$, there exists $L_1(G) \geq 0$ such that for all $L \geq 0$,

$$L \geq L_1(G) \Longrightarrow I(U(G, L)) = I(f^{-1}(G)). \tag{6}$$

Indeed, $I(U(G, L)) = \min[I(f^{-1}(G)), I(f^{-1}(K_L)^c)]$ since $U(G, L) = f^{-1}(G) \cup f^{-1}(K_L)^c$, and (6) follows with (5).
Let us show now that (3) holds. If $I(f^{-1}(G)) = \infty$, there is nothing to prove. Let us suppose from now on that $I(f^{-1}(G)) < \infty$. As $U(G, L)$ is open, by the lower bound for $\{X_n\}$ and (6), we obtain $\liminf_{n\to\infty} \frac{1}{n} \log I\!P(X_n \in U(G, L)) \geq -I(U(G, L)) = -I(f^{-1}(G))$ for all $L \geq L_1(G)$. Hence, for all $\delta > 0$, there exist $n_1(\delta, G) \geq 1$ such that $I\!P(X_n \in U(G, L)) \geq \exp(-n[I(f^{-1}(G)) + \delta])$, as soon as $n \geq n_1(\delta, G)$.
On the other hand, by (1) there exists $n_2(L) \geq 1$ such that $I\!P(Y_n \notin K_L) \leq \exp(-\frac{nL}{2})$, as soon as $n \geq n_2(L)$. One can choose $0 < \delta \leq 1$ without loss. Therefore, there exists $L_2(G) \geq 0$ such that for all $n \geq 1, 0 < \delta \leq 1$ and all $L \geq L_2(G)$, $\exp(\frac{-nL}{2}) \leq \frac{1}{2}\exp(-n[I(f^{-1}(G)) + \delta])$.
Taking (4) into account, it comes out that for all $n \geq \max(n_1(\delta, G), n_2(L_*(G)))$ with $L_*(G) = \max(L_1(G), L_2(G))$ and all $0 < \delta \leq 1$ we have $I\!P(Y_n \in G) \geq \frac{1}{2}\exp(-n[I(f^{-1}(G)) + \delta])$ which leads to the desired result. □

3. Applications

In this section, we give two simple applications of Theorem 2.1.

3.1. **Self-normalized sequences.** About self-normalized sequences and their applications in statistics, see for instance [1] and [4]. Let us consider the simplest self-normalized sequence:

$$Y_n = \frac{\frac{1}{n}\sum_{i=1}^n Z_i}{\left(\frac{1}{n}\sum_{i=1}^n |Z_i|^p\right)^{1/p}}$$

where $1 \leq p < \infty$ and (Z_i) is an iid sequence of real valued centered random variables. We decide that $Y_n = 0$ if either $U_n \triangleq \frac{1}{n}\sum_{i=1}^n Z_i$ or $V_n \triangleq [\frac{1}{n}\sum_{i=1}^n |Z_i|^p]^{1/p}$ vanishes.
We have $Y_n = f(X_n)$ with $X_n = (U_n, V_n)$ and $f(u,v) = u/v$ if $v \neq 0$ and $f(u,0) = 0$ (as a convention). Clearly, f has a closed graph. On the other hand, by Jensen's inequality $Y_n \in [-1,1]$ so that it is exponentially tight. Therefore, using Theorem 2.1 one can derive the LDP for $\{Y_n\}$ from the LDP for (U_n, V_n) without worrying about dividing by zero (which of course is not a real trouble since V_n almost never vanishes).

3.2. U**-statistics.** Let S be a separable metric space endowed with its Borel σ-field $\mathcal{B}(S)$ and $(Z_i)_{i\geq 1}$ a sequence of S-valued random variables. Suppose that the collection of the empirical measures

$$X_n = \frac{1}{n}\sum_{i=1}^n \delta_{Z_i} \in \mathcal{P}(S), n \geq 1$$

obeys the LDP in the space $\mathcal{P}(S)$ of all probability measures on S equipped with the usual topology of weak convergence with the rate function I. For instance, if the Z_i's are independent and identically distributed with the law $P \in \mathcal{P}(S)$, I is the relative entropy with respect to P. We are interested in the large deviations of the U-statistics

$$X_n^{(2)} = \frac{1}{n(n-1)} \sum_{i,j=1; i\neq j}^n \delta_{(Z_i,Z_j)} \in \mathcal{P}(S^2), n \geq 2$$

As $X_n^{(2)} = \frac{n}{n-1}X_n^{\otimes 2} - \frac{1}{n-1}\frac{1}{n}\sum_{i=1}^n \delta_{(Z_i,Z_i)}$, one way to derive the LDP for $\{X_n^{(2)}\}$ from the LDP for $\{X_n\}$ is to prove a LDP for

$$Y_n = X_n^{\otimes 2}$$

and then show that Y_n and $X_n^{(2)}$ are exponentially equivalent. It appears that $Y_n = f(X_n)$ with

$$f : \nu \in \mathcal{P}(S) \mapsto \nu^{\otimes 2} \in \mathcal{P}(S^2)$$

where the product space S^2 is equipped with the product topology $\mathcal{B}(S)^{\otimes 2}$. We have assumed that S is a metric separable space to insure the identity: $\mathcal{B}(S)^{\otimes 2} = \mathcal{B}(S^2)$. Our aim in the following lines is to present a proof of the LDP for $\{Y_n\}$, based on Theorem 2.1.
First of all, let us note that if the separable metric space S is also *complete*, f is a continuous function (see Lemma 3.1 below. Hence, the LDP for $\{Y_n\}$ follows from the usual Contraction Principle. At Proposition 3.3, we are going to extend this result to a separable metric space S in the case where $\{Y_n\}$ is exponentially tight.

Lemma 3.1. *If S is a Polish space, then f is continuous.*

Proof. As S is a separable metric space, $\mathcal{P}(S)$ and $\mathcal{P}(S^2)$ are metric spaces and it is enough to build the proof with sequences. As S is Polish, the set $C_b(S)^{\otimes 2}$ of all tensor products $u \otimes v(s,t) = u(s)v(t)$ of continuous bounded functions u and v on S is convergence determining for the topology of weak convergence on $\mathcal{P}(S^2)$. Let (ν_n) be a convergent sequence: $\nu_n \to \nu \in \mathcal{P}(S)$. For all $u \otimes v \in C_b(S)^{\otimes 2}$, $\lim_n \int_{S^2} u \otimes v \, d\nu_n^{\otimes 2} = \lim_n \int_S u \, d\nu_n \int_S v \, d\nu_n = \int_S u \, d\nu \int_S v \, d\nu = \int_{S^2} u \otimes v \, d\nu^{\otimes 2}$. This proves the desired result. □

Lemma 3.2. *If S is a separable metric space, then f has a closed graph.*

Proof. If S is only a separable metric space, $C_b(S)^{\otimes 2}$ may not be convergence determining anymore, but it is still a determining set. Let us take (ν_n) a convergent sequence: $\nu_n \to \nu \in \mathcal{P}(S)$ such that the limit $\mu = \lim_n f(\nu_n) = \lim_n \nu_n^{\otimes 2}$ exists in $\mathcal{P}(S^2)$. For all $u \otimes v \in C_b(S)^{\otimes 2}$, we have $\lim_n \int_{S^2} u \otimes v \, d\nu_n^{\otimes 2} = \int_{S^2} u \otimes v \, d\mu$. As in the above lemma, we also obtain $\lim_n \int_{S^2} u \otimes v \, d\nu_n^{\otimes 2} = \int_{S^2} u \otimes v \, d\nu^{\otimes 2}$. Therefore, $\lim_n f(\nu_n) = \mu = \nu^{\otimes 2}$, which is the desired result. □

Proposition 3.3. *Let S be a separable metric space such that $\{X_n\}$ obeys the LDP in $\mathcal{P}(S)$ with rate function I. Let us assume in addition that*

(i) *for all $n, k \geq 1$ and $\delta > 0$, there exists a compact subset $K \subset S$ such that $\mathbb{P}(X_n(K^c) > 1/k) \leq \delta$, and*
(ii) *for all $k \geq 1$ and $A \geq 0$, there exists a compact subset $K \subset S$ such that $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X_n(K^c) > 1/k) \leq -A$.*

Then, $\{X_n^{\otimes 2}\}$ obeys the LDP in $\mathcal{P}(S^2)$ with rate function $J(\mu) = I(\nu)$ if $\mu = \nu^{\otimes 2} \in \mathcal{P}(S^2)$ and $J(\mu) = +\infty$ otherwise.

The requirement (i) means that the law of X_n is tight, for all n.

Proof. For any $\nu \in \mathcal{P}(S)$ and any $K \in \mathcal{B}(S)$, $\nu^{\otimes 2}([K \times K]^c) \leq 2\nu(K^c)$. So that (i) and (ii) are satisfied for $\{X_n^{\otimes 2}\}$ in $\mathcal{P}(S^2)$ whenever they are satisfied for $\{X_n\}$. By Lemma 3.4 below, $\{X_n^{\otimes 2}\}$ is exponentially tight in $\mathcal{P}(S^2)$. Since f has a closed graph by Lemma 3.2, one can apply Theorem 2.1 to get the desired result. □

Lemma 3.4. *Let $\{X_n\}$ be a collection of random elements of $\mathcal{P}(S)$ which satisfies the requirements (i) and (ii) of Proposition 3.3. Then, it is exponentially tight in $\mathcal{P}(S)$.*

Proof. Thanks to (ii), for all $A \geq 0, k \geq 1$, there exists K a compact subset and $N \geq 1$ such that for all $n \geq N$, $\mathbb{P}(X_n(K^c) > 1/k) \leq e^{-nA}$. With (i), one obtains a similar control for the N first X_n's, so that for all $A \geq 0, k \geq 1$, there exists $K_{A,k}$ a compact subset such that for all $n \geq 1$, $\mathbb{P}(X_n(K_{A,k}^c) > 1/k) \leq e^{-nA}$. Let us consider $\mathcal{K}_{\mathcal{A}} \triangleq \cap_{k\geq 1}\{\nu \in \mathcal{P}(S); \nu(K_{A(k),k}) \geq 1 - 1/k\}$. By the direct statement of Prokhorov's theorem (in a metric space), it is a compact subset of $\mathcal{P}(S)$. Choosing $A(k) = A + k$, one gets

$$\begin{aligned} \mathbb{P}(X_n \notin \mathcal{K}_A) &\leq \sum_{k\geq 1} \mathbb{P}(X_n(K_{A+k,k}^c) > 1/k) \\ &\leq e^{-nA} \sum_{k\geq 1} e^{-k} \leq e^{-nA} \end{aligned}$$

which completes the proof of the lemma. □

As an immediate consequence of Proposition 3.3 and Sanov's theorem, we obtain the following

Corollary 3.5. *Let (Z_i) be a sequence of independent identically distributed radom elements with values in separable metric space S. Let us assume that the common law $P \in \mathcal{P}(S)$ of the Z_i's is a tight measure. Then, the collection of U-statistics $\{X_n^{(2)}\}$ obeys the LDP in $\mathcal{P}(S^2)$ with the rate function $J(\mu) = I(\nu \mid P)$ if $\mu = \nu^{\otimes 2}$ and $+\infty$ otherwise, where $I(\nu \mid P)$ is the relative entropy of $\nu \in \mathcal{P}(S)$ with respect to P.*

The point is that S is not supposed to be complete, but P is assumed to be tight (this is automatically fulfilled when the sepearble metric space S is also complete).

References

[1] A. Dembo, Q.M. Shao. Self-normalized large deviations in vectors spaces. *Progress in probability. Proceeding Oberwolfach*, 43 (1998), 27-32.

[2] A. Dembo, O. Zeitouni. *Large Deviations Techniques and Applications*, Applications of Mathematics 38, Springer (1998).

[3] M. Reed, B. Simon. *Methods of Modern Mathematical Physics, I: Functional Analysis*, Academic Press (1980).

[4] Q.M. Shao. Recent developments in self-normalized limit theorems. In B. Szyskowicz, editor, *Asymptotic methods in probability and statistics*, 25 Elsevier (1998), 467-480.

Modal-X, Université Paris 10, Bât. G, 200 av. de la République. 92001 Nanterre Cedex France

CMAP, École Polytechnique, 91128 Palaiseau Cedex, France
E-mail address: christian.leonard@polytechnique.fr