An Ergodic Theorem for Random Lagrangians with an Application to Stochastic Programming

Adib Bagh
Department of Economics
University of California, Davis
bagh@math.ucdavis.edu

Michael Casey
Department of Mathematics & Computer Science
University of Puget Sound, Tacoma
mcasey@ups.edu

July 20, 2003

Abstract

We prove an ergodic theorem showing the almost sure epi/hypo-convergence of a sequence of random lagrangians to a limit lagrangian where the random lagrangians are generated by stationary sampling of a probability measure. We apply this theorem to stochastic programming and demonstrate that the outer set-limit of the sequence of the set of saddle points from the sampled problems is a subset of the set of saddle points of the true problem.

AMS Subject Classification: 90C15, 90C47
Keywords: Ergodic theory, stochastic programming, duality, lagrangian, saddle point, epi/hyo-convergence.

1 Introduction

Stochastic programming (SP) is one model of decision making under uncertainty and can be generally formulated as follows:

$$(M)\qquad \begin{array}{l} \text{minimize } E\{f_0(\xi,x)\} = \int_\Xi f_0(\xi,x)\ P(d\xi) \\ \text{subject to} \\ \qquad g_1(x) \le 0 \\ \qquad g_2(x) = 0 \\ \qquad x \in I\!R^n. \end{array}$$

Here $(\Xi, \mathcal{S}, P)$ is a probability space, $f_0 : \Xi \times I\!R^n \to \overline{I\!R}$ is a random lower semicontinuous (lsc) function and for $i = 1, 2$, $g_i : I\!R^n \to I\!R^{m_i}$ is a vector-valued function with

$$g_i(x) = \begin{pmatrix} g_{i1}(x) \\ \cdots \\ g_{im_i}(x) \end{pmatrix},$$

where, for $j = 1, \ldots, m$, g_{ij} is lsc. Recall that $\overline{I\!R} = I\!R \cup \{+\infty\}$.This model is general enough to represent the classical two-stage stochastic program as well as the multi-stage SP which models a sequential decision making problem [9], [7].

In many applications, the probability measure P representing the underlying uncertainty of the problem is not known and must be estimated. One convenient estimator is the empirical measure P^ν based on a random sample $\xi^1, \ldots, \xi^\nu$. Since P^ν is based on a random sample, it is a random measure and results in a random SP:

$$(M^\nu) \qquad \begin{array}{l} \text{minimize } E^\nu\{f_0(\xi, x)\} = \int_\Xi f_0(\xi, x)\ P^\nu(d\xi) = 1/\nu \sum_{k=1}^{\nu} f(\xi^k, x) \\ \text{subject to} \\ \qquad\qquad g_1(x) \le 0 \\ \qquad\qquad g_2(x) = 0 \\ \qquad\qquad x \in I\!R^n. \end{array}$$

When the random sample $\xi^1, \ldots, \xi^\nu$ is independent and identically distributed (iid), several authors [2],[6], [10] obtained various *laws of large numbers* for these random problems. These theorems establish that the random problems epi-converge to the true problem. This implies that if x^ν is a sequence of solutions of the sampled problems, then any cluster point of x^ν is an optimal solution of the true problem. In the more general case when $\xi^1, \ldots, \xi^\nu$ is strongly stationary, Korf and Wets [11] proved an *ergodic theorem* which establishes the same type of behavior as these laws of large numbers.

Whereas the above authors focus on the behavior of primal solutions of M^ν, we consider the asymptotic behavior of both the primal and dual solutions simultaneously, that is to say, we investigate the saddle points of the random approximating problems and establish convergence to saddle points of the true problem. In his Ph.D. thesis[1], Abdulfattah considered this problem when the sampling is iid. In this paper, we relax some of Abdulfattah's conditions, although we restrict our setting to $I\!R^n$, and we consider the case when the sampling is stationary. In order to do this, we consider the lagrangians associated with the true and approximating problems and demonstrate, under appropriate conditions, the almost sure epi/hypo-convergence of the random lagrangians to the true lagrangian; this is the ergodic theorem for random lagrangians and it implies that the outer set-limit of the set of saddle points of the random lagrangians is a subset of the set of saddle points of the true lagrangian.
Towards this end we reformulate problem (M) using a convex and finite penalty function to obtain the problem (we assume no equality constraints and so set

$g_1 = g$ in what follows) M_θ:

$$\text{Minimize } E\{f_0(\xi, x)\} + \theta(g(x)) \text{ over } \mathbb{R}^n,$$

where $\theta : \mathbb{R}^m \to \mathbb{R}_+$. We can associate the following lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ with the problem M_θ :

$$L(x, y) = \int_\Xi f_0(\xi, x)P(d\xi) + < g(x), y > -\theta^*(y),$$

where θ^* is the conjugate of θ.
In this case, sampling produces lagrangians of the form:

$$L^\nu(x, y) = \frac{1}{\nu}\Sigma_{k=1}^{\nu} f_o(\xi^k, x) - < g(x), y > -\theta^*(y).$$

Observe that the lagrangian L^ν is dependent upon the random sample $\xi^1, \ldots, \xi^\nu$ and is therefore itself random. Our ergodic theorem will give sufficient conditions for the epi/hypo-convergence of L^ν to L, which implies the following stability result: Let (x^ν, y^ν) be saddle points for L^ν. If (x^{ν_m}, y^{ν_m}) is a subsequence of (x^ν, y^ν) that converges almost surely to some point (x, y), then (x, y) is a saddle point of L.
The paper is organized as follows. Section 2 contains a discussion of epi/hypo-analysis, a body of variational results specific to lagrangians. In section 3, we review the probabilistic setting of our problem. In section 4, we modify the results of Korf and Wets in order to apply them to random lagrangians and in the process prove a new ergodic theorem for the average problem for random lsc functions. Section 5 contains our main results: the ergodic theorems for random lagrangians. The first ergodic theorem applies to the problem of finding the saddle points of the epi/hypo-expectation of the random lagrangian L:

$$\underset{\boldsymbol{R}^n, \boldsymbol{R}^m}{\text{argminimax}} \;\text{e/h-}EL(x, y).$$

This problem we call the *conjugate problem* or the *average problem* for random lagrangians. The second ergodic thereom applies to the standard problem where we are interested in the saddle points of the expectation of L:

$$\underset{\boldsymbol{R}^n, \boldsymbol{R}^m}{\text{argminimax}}\, EL(x, y).$$

In section 6 we apply our theorem to stochastic programming.

2 Epi/Hypo-Analysis

We presuppose the reader to be familiar with some basic notions of variational analysis, e.g. epi-graph, epi-convergence and outer and inner set limits. For a discussion of these and other concepts we refer the reader to [13].
The conjugate of f, also called the Legendre-Fenchel transform of f, is

$$f^*(x^*) = \sup_{x \in \boldsymbol{R}^n} \{< x, x^* > -f(x)\}.$$

Epi-addition and epi-multiplication are well known tools in optimization that are often used to obtain functions with regularity properties [13, Chap 1, sec. H]. The definition of the epi-sum of f and g, denoted $f +_e g$, is

$$f +_e g(x) = \inf\{f(u) + g(v) | u + v = x\}.$$

Epi-multiplication of f by $\alpha > 0$, denoted by $\alpha \underset{e}{*} f$, is defined as

$$\alpha \underset{e}{*} f(x) = \alpha f(\alpha^{-1} x).$$

In general, a convex optimization problem can be formulated as

$$\inf_{x \in \mathbb{R}^n} f(x),$$

with

$$f(x) = \begin{cases} f_0(x) & \text{if } x \in C \subset \mathbb{R}^n \\ +\infty & \text{otherwise.} \end{cases}$$

where f_0 is real-valued and convex, and C is a convex subset of $\mathbb{R}^n$. We embed the problem (P) into a parameterized family of problems. To this end we introduce a convex perturbation function $F : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ so that

$$F(x, 0) = f(x).$$

We now form the lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$:

$$L(x, y) = \inf_{y^* \in \mathbb{R}^m} \{F(x, y^*) - < y^*, y >\}.$$

We denote by $\underset{\mathbb{R}^n, \mathbb{R}^m}{\operatorname{argminimax}} L$ the saddle points of L. These are points $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}), \forall x \in \mathbb{R}^n, \forall y \in \mathbb{R}^m.$$

The convex parent of L is the function $F : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ defined as

$$F(x, y^*) = \sup_{y \in \mathbb{R}^m} \{L(x, y) + < y, y^* >\}.$$

The concave parent of L is $G : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ defined as

$$G(x^*, y) = \inf_{x \in \mathbb{R}^n} \{L(x, y) - < x, x^* >\}.$$

We say that the bi-function L is closed, if $F^* = -G$ and $(-G)^* = F$. When dealing with a function of two variables such as $\Phi(x, y)$, we will use Φ^x and Φ^y to indicate conjugation with respect the first and second variables respectively. We will use Φ^* to indicate conjugation with respect to both variables at the

same time.
Let L_1 and L_2 be two lagrangians. Their epi/hypo-sum is

$$L_1 +_{e/h} L_2(x, y) = \inf_{u_1+u_2=x} \sup_{v_1+v_2=y} \{L_1(u_1, v_1) + L_2(u_2, v_2)\}.$$

We define the epi/hypo-product of the scalar $\alpha > 0$ and L as

$$\alpha \underset{e/h}{*} L(x, y) = \alpha L(\alpha^{-1}x, \alpha^{-1}y).$$

We also define a notion of convergence that is useful when we are approximating lagrangians. We say a sequence of lagrangians L^ν epi/hypo-converge to a lagrangian L, and we write $L^\nu \xrightarrow{e/h} L$, [4], if
(a) $\forall (x, y)$ and $\forall x^\nu \to x$, $\exists y^\nu \to y$ such that

$$\liminf_{\nu\to\infty} L^\nu(x^\nu, y^\nu) \geq L(x, y) \text{ and}$$

(b) $\forall (x, y)$ and $\forall y^\nu \to y$, $\exists x^\nu \to x$ such that

$$\limsup_{\nu\to\infty} L^\nu(x^\nu, y^\nu) \leq L(x, y).$$

Epi/hypo-convergence induces convergence of saddle points.

Theorem 2.1 [3, Theorem 2.5]**.** *Let $\{L^\nu, L : I\!R^n \times I\!R^m \to \overline{I\!R}\}$ be a collection of bivariate functions such that*

$$L^\nu \xrightarrow{e/h} L.$$

Then

$$\limsup_{\nu\to\infty} \operatorname*{argminimax}_{\boldsymbol{R}^n, \boldsymbol{R}^m} L^\nu \subseteq \operatorname*{argminimax}_{\boldsymbol{R}^n, \boldsymbol{R}^m} L.$$

Recall the definition of epi-convergence[13, Proposition 7.2]: A sequence of functions $\{f, f^\nu : I\!R^n \to \overline{I\!R}\}$ epi-converges to $f : I\!R^n \to \overline{I\!R}$, written $f^\nu \xrightarrow{e} f$, if for all $x \in I\!R^n$,

$$(i) \forall x^\nu \to x, \liminf f^\nu(x^\nu) \geq f(x),$$

$$(ii) \exists x^\nu \to x, \limsup f^\nu(x^\nu) \leq f(x).$$

The following theorem relates epi/hypo-convergence of the lagrangians to the epi-convergence of their parents.

Theorem 2.2 [3, Theorem 2.4]**.** *Let* $\{L^\nu, L : I\!R^n \times I\!R^m \to \overline{I\!R}\}$ *be a collection of closed convex-concave lagrangians with convex and concave parents* $\{F^\nu, F : I\!R^n \times I\!R^m \to \overline{I\!R}\}$ *and* $\{G^\nu, G : I\!R^n \times I\!R^m \to \overline{I\!R}\}$ *respectively. Then the following are equivalent*

(i) $F^\nu \xrightarrow{e} F$
(ii) $-G^\nu \xrightarrow{e} -G$
(iii) $L^\nu \xrightarrow{e/h} L$.

The next lemma shows that in order to establish epi/hypo-convergence of L^ν to L on $I\!R^n \times I\!R^m$, it is sufficient to show it on the special set $\mathcal{R}$, a dense subset of $I\!R^n \times I\!R^m$.

Lemma 2.3. *Let* $\mathcal{R}_1$ *be the projection onto* $I\!R^n \times I\!R^m$ *of a countable dense subset of* $\operatorname{epi} F$ *where* F *is the convex parent of the closed largrangian* L*. Let* $\mathcal{R}_2$ *be the projection onto* $I\!R^n \times I\!R^m$ *of a countable dense subset of* $\text{e-}\liminf F^\nu$ *where* F^ν *is the convex parent of* L^ν*. To show epi/hypo-convergence of* L^ν *to* L *on* $I\!R^n \times I\!R^m$*, it is sufficient to show it on the set* $\mathcal{R}$ *where* $\mathcal{R} = \mathcal{R}_1 \bigcup \mathcal{R}_2$*.*

Proof. By [3, Theorem 2.4], $L^\nu \xrightarrow{e/h} L$ if and only if $F^\nu \xrightarrow{e} F$ where F^ν and F are the convex parents of L^ν and L, respectively. Lemma 2.6 in [11], implies that in fact if $F^\nu \xrightarrow{e} F$ on $\mathcal{R}$ then $F^\nu \xrightarrow{e} F$ on all of $I\!R^n \times I\!R^m$. Now assume that $L^\nu \xrightarrow{e/h} L$ on $\mathcal{R}$. Then $F^\nu \xrightarrow{e} F$ on $\mathcal{R}$. The conclusion follows. □

We say a bi-function L satisfies hypothesis (H), if there exits $y_0, y_1 \in I\!R^m$ such that

(i) $x \longmapsto L(x, y_0)$ is coercive
(ii) $x \longmapsto L(x, y_1)$ is proper,

and for every y

(iii) $L(\cdot, y)$ is lsc.

Remark 1: Recall first that a function f is proper, if $-\infty < f \not\equiv +\infty$.
Remark 2: Any proper closed lagrangian is equivalent to (has the same saddle points as) a lagrangian that satisfies part (iii) of condition H [4].
Some of the properties of epi/hypo-addition, as well as the relation between the epi/hypo-sum of lagrangians and the epi-sum of their convex parents, are given by the following theorem:

Theorem 2.4[1, Proposition 4.3]**.** *Given three convex-concave bi-functions* $L_i : I\!R^n \times I\!R^m \to \overline{I\!R}, i = 1, 2, 3$ *satisfying condition* H*, we have:*

(i) $(L_1 +_{e/h} L_2) = (L_2 +_{e/h} L_1)$
(ii) $\forall\lambda > 0,\ \lambda \underset{e/h}{*} (L_1 +_{e/h} L_2) = (\lambda \underset{e/h}{*} L_1) + (\lambda \underset{e/h}{*} L_2)$
(iii) $L_1 +_{e/h} (L_2 +_{e/h} L_3) = (L_1 +_{e/h} L_2) +_{e/h} L_3$.
(iv) $[-(L_1 +_{e/h} L_2)]^{*_y}(x, y^*) = [(-L)^{*_y}(\cdot, y^*) +_{e_1} (-L_2)^{*_y}(\cdot, y^*)](x)$
(v) $\forall\lambda > 0,\ (\lambda L_1)^{*_y} = \lambda \underset{e}{*} L_1^{*_y}$.

We will later need the following result regarding the joint epi-convergence of convex functions.

Theorem 2.5. *Consider a collection of convex lsc functions $\{f^\nu : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}\}$. Assume:*

(i) $\forall y$, $f^0(\cdot, y)$ is proper.
(ii) there exists a dense set $D \subseteq \mathbb{R}^m$ such that $\forall y \in D$,

$$f^\nu(\cdot, y) \xrightarrow{e} f^0(\cdot, y).$$

Then,

$$f^\nu(\cdot, \cdot) \xrightarrow{e} f^0(\cdot, \cdot).$$

Before proving this theorem we give two Lemmas.

Lemma 2.6. *Suppose $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is proper, convex and lsc. Then*

$$f^\lambda(x, y) := \inf_{u \in \boldsymbol{R}^n} \{f(u, y) + \tfrac{1}{2\lambda} \parallel u - x \parallel^2\}$$

is convex and continuous (jointly in x and y) for all λ.

Proof. The fact f^λ is convex is a direct result of proposition 2.22 in [13]. The fact that, a fixed y, $f^\lambda(\cdot, y)$ is proper implies that f^λ is finite valued over $\mathbb{R}^n \times \mathbb{R}^m$, and hence it is continuous.

Lemma 2.7. *[13, Proposition 7.37]) Suppose that the sequence $\{h^\nu : \mathbb{R}^n \to \overline{\mathbb{R}}\}$ epi-converges to $h^0 : \mathbb{R}^n \to \overline{\mathbb{R}}$, where h^0 is proper, convex, and lsc. Let*

$$h^{\lambda,\nu}(x) = \inf_{u \in \boldsymbol{R}^n} \{h^\nu(u) + \frac{1}{2\lambda} ||u - x||^2\}$$

$$h^{\lambda,0}(x) = \inf_{u \in \boldsymbol{R}^n} \{h^0(u) + \frac{1}{2\lambda} ||u - x||^2\},$$

Then,

$$h^{\lambda,\nu} \to h^{\lambda,0}$$

pointwise.

Finally, the proof of theorem 2.5:
Proof. The "limsup" part of epi-convergence is clear. To prove the " liminf" part, consider the following functions on $\mathbb{R}^n \times \mathbb{R}^m$:

$$f^{\lambda,\nu}(x, y) = \inf_{u \in \boldsymbol{R}^n} \{f^\nu(u, y) + \frac{1}{2\lambda} ||u - x||^2\}$$

$$f^{\lambda,0}(x, y) = \inf_{u \in \boldsymbol{R}^n} \{f^0(u, y) + \frac{1}{2\lambda} ||u - x||^2\},$$

where $|| \cdot ||$ is the norm in $\mathbb{R}^n$. Then, by lemma 2.6, $f^{\lambda,\nu} \xrightarrow{p} f^{\lambda,0}$ on $\mathbb{R}^n \times D$. Condition (i) implies that $\operatorname{int dom} f^{\lambda,0} \neq \emptyset$. Moreover, $f^{\lambda,0}$ is lsc by lemma 2.5.

Therefore, $f^{\lambda,\nu} \xrightarrow{e} f^{\lambda,0}$ on $\mathbb{R}^n \times \mathbb{R}^m$ by [13, theorem 7.17]. Hence, $\forall (x,y) \in \mathbb{R}^n \times \mathbb{R}^m$, $\forall x^\nu \to x$ and $\forall y^\nu \to y$, we have

$$\liminf_{\nu\to\infty} f^\nu(x^\nu, y^\nu) \geq \liminf_{\nu\to\infty} f^{\lambda,\nu}(x^\nu, y^\nu) \geq f^{\lambda,0}(x,y). \tag{1}$$

Hence, by taking the limit of (1) as $\lambda \to 0$, we obtain [13, theorem 1.25]

$$\liminf_{\nu\to\infty} f^\nu(x^\nu, y^\nu) \geq f^0(x,y).$$

□

Note that assumption (ii) is satisfied by setting $D = \mathcal{R}$.

3 Probabilistic Framework

We consider a complete probability space $(\Xi, \mathcal{S}, P)$. A function $f : \Xi \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is a random lsc function, if the set-valued mapping $\xi \longmapsto \text{epi}\, f(\xi,\cdot) : \Xi \rightrightarrows \mathbb{R}^n \times \mathbb{R}$ is a random closed set. For more details on random sets see [11]. An equivalent definition is to say that $f : \Xi \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is a random lsc function, if both the following hold

(i) $\forall \xi \in \Xi$ the function $f(\xi,\cdot)$ is lsc;
(ii)$(\xi,x) \longmapsto f(\xi,x)$ is $(\mathcal{S} \otimes \mathcal{B})$ measurable.

A family of random lsc-functions is independent (identically distributed), if the associated random closed epi-graphs are independent (identically distributed). Again see [11] for details. We will also consider the space $\mathcal{L}_{CC}(\mathbb{R}^n \times \mathbb{R}^m)$ of closed convex-concave bi-functions which take values in $\overline{\mathbb{R}}$. A bi-function $L : \Xi \times \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \overline{\mathbb{R}}$ is a random lagrangian, if the following two conditions hold:

(i) $\forall \xi \in \Xi$, $L(\xi,\cdot,\cdot) \in \mathcal{L}_{CC}(\mathbb{R}^n \times \mathbb{R}^m)$
(ii) $F : \Xi \times \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \overline{\mathbb{R}}$ is a random lsc-function.

Two random lagrangians L_1 and L_2 are iid, if their convex parents F_1 and F_2 are iid. We also note that $L(\cdot,x,y)$ is measurable since it is the conjugate, with respect to the y variable, of a measurable function $f(\cdot,x,y^*)$ [1, Remark 3.6].

For $f : \Xi \times \mathbb{R}^n \to \overline{\mathbb{R}}$, the epi-integral is defined as

$$\text{e-}\int_\Xi f(\xi,\cdot)P(d\xi)(x) \quad := \quad \inf_{u\in U}\{\int_\Xi f(\xi,u(\xi))P(d\xi) | \int_\Xi u(\xi)P(d\xi) = x\}$$

where $U = \{u : \Xi \to \mathbb{R}^n | u \text{ is } \mathcal{S}- \text{ integrable } \}$.
The conjugate of an epi-integral [5, pg 23] is given by this formula

$$[\text{e-}\int_\Xi (\xi,\cdot)]^*(x) = \int_\Xi f^*(\xi,x)P(d\xi).$$

For $L : \Xi \times \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ the epi/hypo-integral is defined as

$$\text{e/h-}\textstyle\int_\Xi L(\xi,\cdot,\cdot)P(d\xi)(x,y)$$
$$:= \inf_{u\in U}\sup_{v\in V}\{\int_\Xi L(\xi,u(\xi),v(\xi))P(d\xi)|\int_\Xi u(\xi)P(d\xi)=x, \textstyle\int_\Xi v(\xi)P(d\xi)=y\}$$

where U is defined as above and $V=\{v:\Xi\to I\!\!R^n|v \text{ is } \mathcal{S}- \text{ integrable }\}$.
In this paper, when we epi-integrate or epi-sum a bivariate function, we always perform these operations with respect to the x-variable only. For example, if F is the convex parent of the random Lagrangian L then

$$\text{e-}\int_\Xi F(\xi,\cdot,y^*)P(d\xi)(x) \quad := \quad \inf_{u\in U}\{\int_\Xi F(\xi,u(\xi),y^*)P(d\xi)|\int_\Xi u(\xi)P(d\xi)=x\}$$

where $U=\{u:\Xi\to I\!\!R^n|u \text{ is } \mathcal{S}- \text{ integrable }\}$.

When considering a random lagrangian $L(\xi,x,y)$, we use a modification of condition H. Let ψ be a function that is bounded below such that $\lim_{t\to+\infty}\frac{\psi(t)}{t}=+\infty$.
We say L satisfies condition $(\hat{H})$, if there exists $y_0,y_1\in I\!\!R^m$ such that

(i) $x\longmapsto L(\xi,x,y_0)\geq\psi(||x||)$ a.s.
(ii) $x\longmapsto L(\xi,x,y_1)$ is proper a.s.,

and for all y,

(iii) $L(\xi,\cdot,y)$ is lsc a.s.

Lemma 3.1. *If a random, closed and convex-concave lagrangian L satisfies hypothesis $(\hat{H})$, then the convex parent of*

$$\text{e/h-}\int_\Xi L(\xi,x,y)P(d\xi)$$

is

$$\text{e-}\int_\Xi F(\xi,x,y^*)P(d\xi)$$

where F is the convex parent of L.

To simplify the notation in the proof, we will use L^1 to denote $L^1(\Xi)$ where Ξ is a probability space with measure P. We also use $\int u$ and $\int v$ to denote $\int_\Xi u(\xi)P(d\xi)$ and $\int_\Xi v(\xi)P(d\xi)$ respectively, where u and v are elements in L^1.

Proof of Lemma 3.1. Let $\Phi(x,y^*)$ be the convex parent of e/h-$\int_\Xi L(\xi,\cdot,\cdot)P(d\xi)$, then by definition

$$\Phi(x,y^*)=\sup_y\{<y,y^*>+\text{e/h-}\int_\Xi L(\xi,x,y)P(d\xi)\},$$

and hence

$$\Phi(x,y^*)=\sup_y\{<y,y^*>+\inf_{x=\int u}\sup_{y=\int v}\int_\Xi L(\xi,u(\xi),v(\xi))P(d\xi)\},$$

and

$$\Phi(x,y^*) = \sup_y \inf_{x=\int u} \{< y, y^* > + \sup_{y=\int v} \int_\Xi L(\xi, u(\xi), v(\xi)) P(d\xi)\}.$$

Now, for a fixed y^*, define $G : L^1 \times Y \to I\!\!R$:

$$G(u,y) = \{< y, y^* > + \sup_{y=\int v} \int_\Xi L(\xi, u(\xi), v(\xi)) P(d\xi)\}$$

By the assumptions of our lemma, G satisfies the conditions of Moreau's theorem [12]: Due to part (i) of condition $(\hat{H})$ and the fact that P is a probability measure, we have for y_0

$$\sup_{y=\int v} \int_\Xi L(\xi, u(\xi), v(\xi)) P(d\xi) \geq \int_\Xi L(\xi, u(\xi), y_0) P(d\xi).$$

Hence, by Theorem 1.3. in Chap VIII in [8], the set$\{u|G(u,y_0) \leq \alpha\}$ is weakly compact in L^1. Moreover, for any y, $G(\cdot, y)$ is weakly lsc over L^1 [8, Theorem 2.1, Chap. VIII], and thus [12]

$$\sup_y \inf_{x=\int u} G(u,y) = \inf_{x=\int u} \sup_y G(u,y)$$

Therefore,

$$\Phi(x,y^*) = \inf_{x=\int u} \sup_y \{< y, y^* > - \inf_{y=\int v} \int_\Xi -L(\xi, u(\xi), v(\xi)) P(d\xi)\}$$

$$= \inf_{x=\int u} [\text{e-}\int (-L) P(d\xi)]^y$$

$$= \inf_{x=\int u} \int (-L)^y P(d\xi)$$

$$= \inf_{x=\int u} \int F(\xi, u(\xi), y^*) P(d\xi).$$

□

The above proof is more general than the proof in [1]. Note that part (i) of our condition $(\hat{H})$ is required for a single y_0 whereas Abdulfattah required the lagrangians to be equi-coercive uniformly in all values of y (condition (i) in Theorem 5.6 in [1]), a condition that is difficult to verify in applications.

Theorem 3.2. *Consider the following sequence of functions* $\{F^\nu : \Xi \times I\!R^n \times I\!R^m \to \overline{I\!R}\}$. *Assume that almost surely,* $F^\nu(\xi, \cdot, \cdot)$ *is convex and lsc. Assume moreover,*

(i) $F^0(\cdot, \cdot)$ is convex, lsc. and $\forall y$, $F^0(\cdot, y)$ is proper.
(ii) there exists a countable dense set $D \subset I\!R^m$ such that for all y, $\exists\, \Xi_y$ with measure one such that $\forall \xi \in \Xi_y$,

$$F^\nu(\xi, \cdot, y) \xrightarrow{e} F^0(\cdot, y).$$

Then $P-$almost surely,

$$F^\nu(\xi, \cdot, \cdot) \xrightarrow{e} F^0(\cdot, \cdot).$$

Proof. From our assumptions, and for any $y \in D$, let $\xi \in \Xi_y$, the assumption (ii) of Theorem 2.4 is satisfied. Take $\bar{\Xi} = \bigcap_{y \in D} \Xi_y$. Then, $\bar{\Xi}$ has measure 1, and for any $\xi \in \bar{\Xi}$ the assumptions of Theorem 2.4. are satisfied. Hence $P-$almost surely,

$$F^\nu(\xi, \cdot, \cdot) \xrightarrow{e} F^0(\cdot, \cdot).$$

□

Note that assumption (ii) is satisfied by taking $D = \mathcal{R}_2$.

4 Ergodic Theorems for Random Lower Semi-continuous Functions

Again, let $(\Xi, \mathcal{S}, P)$ be a complete probability space. A function $\varphi : \Xi \to \Xi$ is *measure preserving*, if for all $A \in \mathcal{S}, P(\varphi^{-1}(A)) = P(A)$. The event A is called *invariant*, if $\varphi^{-1}(A) = A$ almost surely, i.e. $P(\varphi^{-1}(A) \bigtriangleup A) = 0$ where $\bigtriangleup$ is the symmetric difference operator. Now we define ergodicity:

Definition 4.1 (ergodicity)**.** *Let $\mathcal{I}$ denote the σ-field of invariant events of the measure preserving map $\varphi : \Xi \to \Xi$. Then φ is ergodic if for all $A \in \mathcal{I}, P(A) \in \{0, 1\}$, i.e. $\mathcal{I}$ is trivial.*

Korf and Wets [11] proved this ergodic theorem for random lsc-functions using the scalarization technique:

Theorem 4.2 [11, Theorem 7.2]**.** *Let f be a random lsc function defined on $\Xi \times I\!R^n$, and let $\varphi : \Xi \to \Xi$ be an ergodic transformation. Then, whenever $\xi \to \inf_{\boldsymbol{R}^n} f(\xi, \cdot)$ is integrable,*

$$1/\nu \sum_{k=1}^{\nu} f(\varphi^{k-1}(\xi), \cdot) \xrightarrow{e} Ef \quad P - a.s.$$

In this paper, we will need an ergodic theorem for the *conjugate* or *average problem*. By the average problem (for random lsc functions) we mean the problem

$$\min_x \text{ e-}Ef(x) = \text{e-}\int_\Xi f(\xi, x)\ P(d\xi).$$

Theorem 4.3. *Let f be a random convex lsc function defined on $\Xi \times I\!R^n$, $\varphi : \Xi \to \Xi$ be an ergodic transformation. Further, let the following condition hold:*

$$\xi \to \inf_{\boldsymbol{R}^n} f^*(\xi, \cdot) \text{ is integrable.}$$

Then $P-$almost surely,

$$1/\nu \underset{e}{*} \text{ e-}\textstyle\sum_{k=1}^{\nu} f(\varphi^{k-1}(\xi), \cdot) \xrightarrow{e} \text{e-}Ef.$$

Proof. We have by the previous theorem

$$[1/\nu \underset{e}{*} \text{ e-}\textstyle\sum_{k=1}^{\nu} f(\varphi^{k-1}(\xi), \cdot)]^* = 1/\nu \sum_{k=1}^{\nu} f^*(\varphi^{k-1}(\xi), \cdot) \text{ and}$$
$$(\text{e-}Ef)^* = Ef^*.$$

The random lsc function f^* satisfies the hypothesis of the ergodic theorem. Hence,

$$1/\nu \textstyle\sum_{k=1}^{\nu} f^*(\varphi^{k-1}(\xi), \cdot) \xrightarrow{e} E(f^*) \quad P - a.s.$$

Then, the continuity of the Legendre-Fenchel transform yields

$$1/\nu \underset{e}{*} \text{ e-}\textstyle\sum_{k=1}^{\nu} f(\varphi^{k-1}(\xi), \cdot) \xrightarrow{e} \text{e-}Ef \quad P - a.s.$$

□

5 Ergodic Theorems for Random Lagrangians

This section contains our main results. The first is an ergodic theorem for the conjugate or average problem for random lagrangians. The average problem is defined as

$$\underset{\boldsymbol{R}^n, \boldsymbol{R}^m}{\text{argminimax}} \text{ e/h-}EL(x, y).$$

The second is an ergodic theorem for the standard problem:

$$\underset{\boldsymbol{R}^n,\boldsymbol{R}^m}{\text{argminimax}}\ EL(x,y).$$

Theorem 5.1. *Let $(\Xi, \mathcal{S}, P)$ be a probability space, $\varphi : \Xi \to \Xi$ be an ergodic transformation, and $L : \Xi \times I\!R^n \times I\!R^m \to \overline{I\!R}$ a random lagrangian with convex parent F. Assume :*
(i) almost surely, L satisfies condition $\hat{H}$ and for all ξ, $L(\xi,\cdot,\cdot)$ is closed.

(ii) for every y^*, $\Phi(x,y^*) = \text{e-}\int_\Xi F(\xi,\cdot,y^*)P(d\xi)(x)$ is proper.

(iii) for every y^*, there exists Ξ_{y^*} of measure one such that the function $\xi \to \inf_{x\in\boldsymbol{R}^n} F^x(\xi,\cdot,y^*) = -F(\xi,0,y^*)$ is summable.
Then $P-$almost surely,

$$1/\nu \underset{e/h}{*} \text{e/h-}\sum_{k=1}^{\nu} L(\varphi^{k-1}(\xi),\cdot,\cdot) \xrightarrow{e/h} \text{e/h-}EL.$$

Proof. By theorem 2.4, the convex parent of

$$1/\nu \underset{e/h}{*} \text{e/h-}\sum_{k=1}^{\nu} L(\varphi^{k-1}(\xi),\cdot,\cdot)$$

is

$$\Phi^\nu(\xi,x,y^*) = \{1/\nu \underset{e}{*} \text{e-}\sum_{k=1}^{\nu} F(\varphi^{k-1}(\xi),\cdot,y^*)\}(x).$$

Let Φ be the convex parent of e/h-EL = e/h-$\int_\Xi L(\xi,\cdot,\cdot)P(d\xi)$. Then by lemma 3.1, we have

$$\Phi(x,y^*) = \text{e-}\int_\Xi F(\xi,x,y^*)P(d\xi).$$

By assumption (iii) and theorem 4.3, we have

$$\Phi^\nu(\xi,\cdot,y^*) = 1/\nu \underset{e}{*} \text{e-}\textstyle\sum_{k=1}^{\nu} F(\varphi^{k-1}(\xi),\cdot,y^*) \xrightarrow{e} \Phi(\cdot,y^*) \quad P-a.s.$$

Moreover, using assumption (ii) and Theorem 3.2., we get, $P-$almost surely,

$$\Phi^\nu(\xi,\cdot,\cdot) \xrightarrow{e} \Phi(\cdot,\cdot)$$

Theorem 2.2 gives the conclusion:

$$1/\nu \underset{e/h}{*} \text{e/h-}\textstyle\sum_{k=1}^{\nu} L(\varphi^{k-1}(\xi),\cdot,\cdot) \xrightarrow{e/h} \text{e/h-}EL \quad P-a.s.$$

□

We prove a pivotal duality result that will allow us to use the previous theorem to prove our main result. Given a proper convex-concave and closed L, we define the dual lagrangian L^* as

$$L^* = -[[L^x]]^y.$$

In more detail,

$$-L^* = [\sup_x \{< x, x^* > -L(x,y)\}]^y.$$

Since $\{< x, x^* > -L(x, \cdot)\}$ is convex and proper, we have [13, Theorem 11.23]

$$-L^*(x^*, y^*) = \mathrm{cl}\,\mathrm{con}\inf_x \{[< x, x^* > -L(x,y)]^y\},$$

where cl is the lower closure with respect to y^*, and $\mathrm{con}\inf_x\{[< x, x^* > -L(x, \cdot]^y\}$ is the function whose epigraph is the convex hull of the epigraph of the function $\inf_x\{[< x, x^* > -L(x, \cdot)]^y\}$. Hence,

$$-L^*(x^*, y^*) = \mathrm{cl}\,\mathrm{con}\inf_x [\sup_y \{< y, y^* > - < x, x^* > +L(x,y)\}]$$

$$= \mathrm{cl}\,\mathrm{con}\inf_x [F(x, y^*) - < x, x^* >],$$

and hence,

$$-L^*(x^*, y^*) = \mathrm{cl}\ \mathrm{con}\ -F^x(x^*, y^*)$$

where F is the convex parent of L. In particular, note that $L^*(x^*, y^*) \geq F^x(x^*, y^*)$.

Theorem 5.2. *Let $(\Xi, \mathcal{S}, P)$ be a probability space, $\varphi : \Xi \to \Xi$ be an ergodic transformation and $L : \Xi \times I\!R^n \times I\!R^m \to \overline{I\!R}$ a random closed lagrangian. Suppose L^* and L satisfy condition $(\hat{H})$. Then the following are equivalent:*

(i)

$$1/\nu \textstyle\sum_{k=1}^{\nu} L(\varphi^{k-1}(\xi), \cdot, \cdot) \overset{e/h}{\to} \ EL \quad P - a.s.$$

(ii)

$$1/\nu \underset{e/h}{*} \text{e/h-}\textstyle\sum_{k=1}^{\nu} L^*(\varphi^{k-1}(\xi), \cdot, \cdot) \overset{e/h}{\to} \ \text{e/h-}E(L^*) \quad P - a.s.$$

Proof. We first calculate the convex parents of the terms in (ii). We have, by theorem 2.4 and the definition of L^*,

$$\Phi^\nu(x, y^*) = [-1/\nu \underset{e/h}{*} \text{e/h-}\sum_{k=1}^{\nu} L^*(\varphi^{k-1}(\xi), x, \cdot)]^y$$

$$= 1/\nu \underset{e}{*} \text{ e-}\sum_{k=1}^{\nu}(-L^*)^y(\varphi^{k-1}(\xi), x, y^*)$$

$$= 1/\nu \underset{e}{*} \text{ e-}\sum_{k=1}^{\nu}(L^x)(\varphi^{k-1}(\xi), x, y^*).$$

By definition of L^* and the the last equation of the proof of Lemma 3.1,

$$\Phi(x, y^*) = \text{e-}\int[-(L^*)]^y \ P(d\xi) = \text{e-}\int(L^x)(\xi, x, y^*) \ P(d\xi).$$

Now we calculate Ψ^ν and Ψ, the concave parents of the terms in (i). By [13, Proposition 1.2.1], we have

$$-\Psi^\nu(x^*, y) = \sup_x\{< x, x^* > -1/\nu\sum_{k=1}^{\nu}L(\varphi^{k-1}(\xi), x, y)\}$$

$$= 1/\nu \underset{e}{*} \text{ e-}\sum_{k=1}^{\nu}(L^x)(\varphi^{k-1}(\xi), x^*, y).$$

Similarly, since L satisfies part (iii) of $\hat{H}$, $\int L(\xi, \cdot, y)P(d\xi)$ is lsc, we have

$$-\Psi(x, y) = (EL)^x = \text{e-}E(L^x).$$

Of course Φ^ν and $-\Psi^\nu$ are the same and so are Φ and $-\Psi$. Hence, the conclusion of our theorem follows immediately from Theorem 2.2. □

We are now ready to state and prove our main result: the ergodic theorem for random lagrangians.

Ergodic Theorem 5.3. *Suppose L^* satisfies the conditions of Theorem 5.1. Suppose further that L satisfies $\hat{H}$, then*

$$1/\nu\sum_{k=1}^{\nu}L(\varphi^{k-1}(\xi), \cdot, \cdot) \xrightarrow{e/h} \ EL, \ P-a.s$$

and

$$\limsup_{\nu\to\infty} \operatorname*{argminimax}_{\boldsymbol{R}^n, \boldsymbol{R}^m}[1/\nu\sum_{k=1}^{\nu}L(\varphi^{k-1}(\xi), \cdot, \cdot)] \subseteq \operatorname*{argminimax}_{\boldsymbol{R}^n, \boldsymbol{R}^m} EL.$$

Proof. The proof follows from theorems 5.1, 5.2 and 2.1. □

Cautionary Note to the Reader: The reader may initially think that there is a much simpler proof of this theorem. He or she may observe that if one simply shows epi-convergence in the first argument and hypo-convergence in the second then this is sufficient to induce epi/hypo-convergence. In other words to show that

$$L^\nu(\cdot,\cdot) \quad \xrightarrow{e/h} \quad L(\cdot,\cdot)$$

first fix $y = \bar{y}$ (where $\bar{y}$ is arbitrary) and show

$$L^\nu(\cdot,\bar{y}) \quad \xrightarrow{e} \quad L(\cdot,\bar{y}) \tag{2}$$

and then fix $x = \bar{x}$ (where $\bar{x}$ is arbitrary) and show

$$-L^\nu(\bar{x},\cdot) \quad \xrightarrow{e} \quad -L(\bar{x},\cdot). \tag{3}$$

Such a theorem is true. However this new mode of convergence defined by (2) and (3) is much stronger than epi/hypo-convergence. In fact it is so strong, that it is quite useless. So any ergodic theorem which relies on this too-strong mode of convergence is also useless.

6 Application: Stochastic Programming

We apply Theorem 5.3 to the model described in the introduction and thus show that saddle points obtained from ergodic sampling converge almost surely to a saddle point of the original problem. In order to satisfy the assumptions of Theorem 5.3, we assume that f_0 is convex, lsc in x and measurable in ξ. Moreover, we assume that for every i, g_i is convex and lsc. We also need some conditions on the lagrangian

$$L(\xi,x,y) = f_0(\xi,x) + < g(x), y > -\theta^*(y)$$

and its convex parent

$$F(\xi,x,y^*) = f_0(\xi,x) + \theta(g(x+y^*)).$$

We will assume the following
A_1: $f_0(\xi,x) \geq 0$, and $\exists \bar{x}$ such that $\int_\Xi f_0(\xi,\bar{x})P(d\xi) < +\infty$.
A_2: For all i, g_i are level bounded over $I\!R^n$.
A_3: The function θ is convex, finite, and coercive. Hence, θ^* is also convex, finite, and coercive.
A_4: There exists a function $\psi_2 : I\!R \to \overline{I\!R}$ such that

$$f_0(\xi,x) + \theta(g(x)) \leq \psi_2(|x|),\ a.s.,$$

where ψ_2 satisfies the following conditions: $\psi_2(-r) = \psi_2(r)$, $\psi_2^*(\cdot)$ is bounded below, and $\lim_{t\to+\infty} \frac{\psi_2^*(t)}{t} = +\infty$.

Since f_0 is normally a cost function, A_1 is a natural assumption. The level-boundedness of g in A_2 is also a standard condition that is need for the existence of a solution for the problem. Clearly L satisfies parts (ii) and (iii) of

$\hat{H}$ (see Remark 1, page 7). Moreover, for $y_0 > 0$, A_3 implies that $L(\xi, x, y) = f_0(\xi, x) + < g(x), y_0 > -\theta^*(y_0)$ is coercive, which is part (i) of $\hat{H}$. Similarly, L^* satisfies parts (ii) and (iii) of condition $\hat{H}$. Condition A_4 implies that

$$\psi_2^*(|x^*|) \leq [f_0(\xi, x) + \theta(g(x))]^x \leq F^x(\xi, x^*, 0) \leq L^*(\xi, x^*, 0) \ a.s,$$

[13, theorem 11.21]. Hence, condition (i) of $(\hat{H})$ holds.
The convex parent of L^* is L^x. For any y, the function $\xi \to \inf_{x \in \boldsymbol{R}^n} [L^x]^{x^*}(\xi, x, y)$ is integrable since

$$\inf_{x \in \boldsymbol{R}^n} L(\xi, x, y) \leq f_0(\xi, \bar{x}) + < y, g(\bar{x}) > -\theta^*(y)$$

and the right side of the above inequality is integrable by A_1. Thus, condition (iii) of 5.1. is satisfied. Note also that $\Phi(\cdot, y) = \text{e-}\int_\Xi L^x(\xi, \cdot, y) P(\xi)$ is proper, which is condition (ii) of 5.1 : This true because for any y, $\int L^x(\xi, 0, y) P(d\xi) < +\infty$ since the function $\xi \to L(\xi, \bar{x}, y)$ is integrable. Moreover, $\forall x^*$, $\forall y$, and for $\bar{x}$ from A_1, we have

$$L^x(\xi, x^*, y) \geq < x^*, \bar{x} > -L(\xi, \bar{x}, y > .$$

Hence, $\forall x(\cdot) \in L^1$ such that $\int x = x^*$, we have

$$\int_\Xi L(\xi, x(\xi), y) \geq \int_\Xi (< x(\xi), \bar{x} > -L(\xi, \bar{x}, y >) P(d\xi),$$

and thus

$$\Phi(x^*, y) = \text{e-}\int_\Xi L^x(\xi, x^*, y) P(d\xi) > -\infty.$$

We have showed that L and L^* satisfy condition $\hat{H}$, and that L^* satisfies the conditions of Theorem 5.1. Therefore, we can now apply theorem 5.3 to obtain the desired result about the convergence of saddle points generated through an ergodic sampling process. We summarize the above in our last theorem - the ergodic theorem for stochastic convex programming:

Theorem 6.1. *Consider the following stochastic convex program:*

$$\begin{aligned} &\textit{minimize } E\{f_0(\xi, x)\} = \textstyle\int_\Xi f_0(\xi, x) \ P(d\xi) \\ &\textit{subject to} \\ &\qquad\qquad g(x) \leq 0 \\ &\qquad\qquad x \in I\!R^n. \end{aligned}$$

Assume assumptions A_1 through A_4 are satisfied. Then

$$1/\nu \sum_{k=1}^{\nu} L(\varphi^{k-1}(\xi), \cdot, \cdot) \xrightarrow{e/h} \ EL, \ P - a.s.$$

and

$$\limsup_{\nu\to\infty} \operatorname*{argminimax}_{\boldsymbol{R}^n,\boldsymbol{R}^m} [1/\nu \sum_{k=1}^{\nu} L(\varphi^{k-1}(\xi),\cdot,\cdot)] \subseteq \operatorname*{argminimax}_{\boldsymbol{R}^n,\boldsymbol{R}^m} EL$$

where

$$L(\xi,x,y) = f_0(\xi,x) + < g(x), y > -\theta^*(y).$$

References

[1] S. Abdulfattah. *Loi des grandes nombres epi/hypographique*. PhD thesis, Université Montpellier II, 1991.

[2] Z. Artstein and S. Hart. Law of large numbers for random sets and allocation processes. *Mathematics of Operations Research*, 6(4):485–492, 1981.

[3] H. Attouch, D. Aze, and R. J-B Wets. On continuity properties of the partial legendre-fenchel transform: Convergence of sequences of augmented lagrangian functions, Moreau-Yosida approximates and subdifferential operators. In J.-B. Hiriat-Urruty, editor, *Fermat-Days 85: Mathematics for Optimization*, pages 1–42. North Holland, Amsterdam, 1986.

[4] H. Attouch and R. J-B Wets. A convergence theory for saddle functions. *transactions of the American Mathematical Society*, 280, 1983.

[5] H. Attouch and R. J-B Wets. Epigraphical analysis. In H. Attouch, J.-P. Aubin, F. Clark, and I. Ekeland, editors, *Analysis Non Linéaire*, pages 73–100. Gauthier-Villars, Paris, 1989.

[6] Hedy Attouch and Roger J-B Wets. Epigraphical processes: Laws of large numbers for random lsc functions. *Seminar d'Analyse Convexe*, pages 13.1 – 13.29, 1990.

[7] J.R. Birge and F.V. Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research. Springer-Verlag, 1997.

[8] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*, volume 1 of *studies in mathematics and its applications*. North-Holland, New York,U.S.A., 1976.

[9] P. Kall and S.W. Wallace. *Stochastic Programming*. Wiley interscience series in systems and optimization. J.Wiley, New York,U.S.A., 1994.

[10] A. King and R. J-B Wets. Epi-consistency of convex stochastic programs. *Stochastics and Stochastics Reports*, 34, 1990.

[11] Lisa A. Korf and Roger J-B Wets. An ergodic theorem for stochastic programming problems. *Mathematics of Operations Research (to appear)*, 1999.

[12] J.J. Moreau. Théorèmes ”inf-sup”. *C.R.A.S.*, 258:2720–2722, 1964.

[13] R.T. Rockafellar and R. J-B Wets. *Variational Analysis*, volume 317. Springer-Verlag, Heidelberg, Germany, 1998.