A Strong Law of Large Numbers for Random Monotone Operators

Adil Salim*

Abstract

We provide a strong law of large numbers for random monotone operators. The expectation of a random monotone operator is defined through its Aumann integral. An application to empirical mean minimization is provided.

1 Introduction

Maximal monotone operators are set valued mappings which play an important role in various fields of convex analysis [7, 11], ranging from convex optimization to the analysis of Partial Differential Equations. Since the work of H. Attouch [3, Chap. I] (see also [4, Chap. III]), the set of maximal monotone operators $\mathscr{M}(H)$ over a separable Hilbert space H is a Polish space [3, Prop 1.1]. The Borelian sigma-field induced by this topology allows to study measurable maps with values in $\mathscr{M}(H)$ [3, Chap. II]. Following [8], a random variable with values in $\mathscr{M}(H)$ is called a random monotone operator. Random monotone operators were used to prove the convergence of the stochastic Forward Backward algorithm in [8, 9] where the expectation of a random monotone operator is defined through its Aumann integral [5] (generalization of Lebesgue integral to set valued mappings).

With the topology defined in [3] and the expectation used in [8, 9] one may ask if random monotone operators admit a law of large numbers.

Various laws of large numbers for random sets have already been proven in the literature. Different class of random sets were considered (compact, unbounded...), see *e.g.* [1, 2, 12, 21, 22, 23]. In particular, laws of large numbers for compact valued subdifferentials of random non convex functions were obtained in [12, 21, 23]. To our knowledge, these results don't cover the case of random monotone operators.

In this note, we prove a law of large numbers for random monotone operators and apply it to the convergence of solutions of the empirical mean minimization [10].

The next section provides some background knowledge on (random) monotone operators. Then, the main theorem is stated in section 3. Section 4 is devoted to the proof of the main result. An application to empirical mean minimization is provided in section 5. Finally, we conclude in section 6.

*King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.
adil.salim@kaust.edu.sa

2 Background

In this section we define maximal monotone operators, random monotone operators and their expectation.

2.1 Maximal monotone operators

We review some basic material regarding maximal monotone operators. The proof of these facts can be found in [7]. Let H be a separable Hilbert space and let I be the identity map over H. An operator A over H is a set valued mapping over H, *i.e* a function from H to the set of all subsets of H. An operator can be identified to its graph $G(\mathsf{A}) = \{(x,y) \in H \times H, y \in \mathsf{A}(x)\}$. The domain of A is $\operatorname{dom}(\mathsf{A}) = \{x \in H, \mathsf{A}(x) \neq \emptyset\}$. The inverse operator A^{-1} is defined by $G(\mathsf{A}^{-1}) = \{(y,x) \in H \times H, y \in \mathsf{A}(x)\}$, the resolvent operator is defined by $J_{\mathsf{A}} = (I + \mathsf{A})^{-1}$ and the set of zeros of A is $Z(\mathsf{A}) = \mathsf{A}^{-1}(0)$. Note that $\ell \in Z(\mathsf{A})$ if and only if $\ell \in J_{\mathsf{A}}(\ell)$. The operator A is said monotone if the following condition holds:

$$\forall (x,y), (x',y') \in G(\mathsf{A}), \langle x - x', y - y' \rangle \geq 0,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of H. In this case, $J_{\mathsf{A}}(x)$ is either the empty set or a singleton, *i.e* J_{A} can be identify with a classical function $\operatorname{dom}(J_{\mathsf{A}}) \to H$. Moreover, A is a maximal monotone operator, which we denote $\mathsf{A} \in \mathscr{M}(H)$, if A is a monotone operator such that $\operatorname{dom}(J_{\mathsf{A}}) = H$. In this case, $J_{\mathsf{A}} : H \to H$ is a 1-Lipschitz continuous function. The maximality of A is equivalent to the maximality (for the inclusion ordering) of $G(\mathsf{A})$ in the set of all graphs of monotone operators over H [18]. Moreover, if A is maximal, then for every $x \in H$, $\mathsf{A}(x)$ is a (possibly empty) closed convex set. If $x \in \operatorname{dom}(\mathsf{A})$, $\mathsf{A}_0(x)$ is defined as the projection of 0 onto $\mathsf{A}(x)$. In other words, $\mathsf{A}_0(x)$ is the least norm element in $\mathsf{A}(x)$. Given $\gamma > 0$, the Yosida approximation of A is the function $\mathsf{A}_\gamma(x) = \frac{x - J_{\gamma \mathsf{A}}(x)}{\gamma}$. The function A_γ is $1/\gamma$-Lipschitz continuous. Moreover, for every $x \in \operatorname{dom}(\mathsf{A})$,

$$\|\mathsf{A}_\gamma(x)\| \leq \|\mathsf{A}_0(x)\|. \tag{1}$$

Given two maximal monotone operators A and B, the sum $\mathsf{A} + \mathsf{B}$ is defined by $(\mathsf{A} + \mathsf{B})(x) := \mathsf{A}(x) + \mathsf{B}(x)$ where $\mathsf{A}(x) + \mathsf{B}(x)$ is the classical Minkowski sum of two sets. One can check that $\mathsf{A} + \mathsf{B}$ is a monotone operator, however, $\mathsf{A} + \mathsf{B}$ is not necessarily maximal [19, Page 54].

Consider the set $\Gamma_0(H)$ of convex lower semi-continuous and proper functions $F : H \to (-\infty, +\infty]$ (see [7]). Then, the subdifferential ∂F of F is a maximal monotone operator. In other words, $\mathscr{M}_s(H) = \{\partial F, F \in \Gamma_0(H)\}$ is a subset of $\mathscr{M}(H)$. Let C be a convex set and consider $F = \iota_C$ the convex indicator function of C, defined by $F(x) = 0$ if $x \in C$ and $F(x) = +\infty$ else. Then $F \in\in \Gamma_0(H)$ and ∂F is the normal cone N_C to C.

2.2 Random monotone operators

For every $x \in H$, consider the map p_x from $\mathscr{M}(H)$ to H defined by $p_x(\mathsf{A}) = J_{\mathsf{A}}(x)$. The topology of R-convergence is the initial topology on $\mathscr{M}(H)$ with respect to the family of functions $\{p_x, x \in H\}$. In other words, it is the coarsest topology on $\mathscr{M}(H)$ that makes the functions p_x continuous. Endowed with this topology, $\mathscr{M}(H)$ is a Polish space [3, Lemme 2.1] (metrizable, separable and complete).

In the sequel, we consider a probability space $(\Xi, \mathscr{G}, \mu)$ such that $\mathscr{G}$ is σ-finite and μ-complete, and a measurable map $A : (\Xi, \mathscr{G}, \mu) \to (\mathscr{M}(H), \mathscr{B}(\mathscr{M}(H)))$ (where $\mathscr{B}(\mathsf{X})$ denotes the Borelian sigma field over any topological space X). Such a measurable map is called a *random monotone operator*. A normal convex integrand is a measurable map $f : (\Xi \times H, \mathscr{G} \otimes \mathscr{B}(H)) \to ((-\infty, +\infty], \mathscr{B}((-\infty, +\infty]))$ such that for every $s \in \Xi$, $f(s, \cdot) \in \Gamma_0(H)$. Using [3, Theorem 2.3], $s \mapsto \partial f(s, \cdot)$ is a random monotone operator.

Let $\mathscr{L}^1(\Xi, \mathscr{G}, \mu)$ be the space of $\mathscr{G}$-measurable and μ-integrable H-valued functions defined on Ξ. For every $x \in H$, we define

$$\mathfrak{S}_x := \{\varphi \in \mathscr{L}^1(\Xi, \mathscr{G}, \mu) \,:\, \varphi(s) \in A(s)(x) \text{ for } \mu - \text{almost every (a.e.) } s \in \Xi\}\,.$$

We shall prefer the notation $A(s, x)$ for the set $A(s)(x)$. Note that the set $\mathfrak{S}_x$ might be empty. The mean operator $\mathcal{A}$ of A is defined by its Aumann integral [5],

$$\forall x \in H, \mathcal{A}(x) := \left\{ \int \varphi d\mu \,:\, \varphi \in \mathfrak{S}_x \right\}.$$

We shall refer to $\mathcal{A}$ as the expectation of A. One can check that $\mathcal{A}$ is a monotone operator.

Definition 1. The random monotone operator A is said **integrable** if $\mathcal{A}$ is a *maximal* monotone operator.

3 Main result

In this section we provide the main theorem and discuss our assumptions.

Theorem 1. Consider a family of i.i.d random variables $(\xi_n)_n$ from some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to $(\Xi, \mathscr{G})$ with distribution μ. Assume that the random monotone operator A is integrable and that for every $n \in \mathbb{N}$,

$$\overline{A_n} := \frac{1}{n} \sum_{k=1}^{n} A(\xi_k) \tag{2}$$

is $\mathbb{P}$-almost surely (a.s.) maximal.

Then, $\overline{A_n} : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathscr{M}(H), \mathscr{B}(\mathscr{M}(H)))$ is a random monotone operator and $\mathbb{P}$-a.s,

$$\overline{A_n} \underset{n\to\infty}{\longrightarrow} \mathcal{A}, \tag{3}$$

in the sense of R-convergence. Moreover, if $A(s) = \partial f(s, \cdot)$ where f is a normal convex integrand, then there exists $G \in \Gamma_0(H)$ such that $\mathcal{A} = \partial G$.

This theorem is a law of large numbers for the family of i.i.d random monotone operators $(A(\xi_n))_n$, where the limit is the expectation of A.

Let us now discuss the assumptions. The main assumption is the integrability of A (*i.e* the maximality of $\mathcal{A}$). The condition $\mathcal{A} \in \mathscr{M}(H)$ must hold to have convergence of $\overline{A_n}$ in the topological space $\mathscr{M}(H)$. Conditions under which $\mathcal{A}$ is maximal can be found in [8, Section 3.2]. The maximality of $\mathcal{A}$ is not a consequence of the other assumptions. More generally, there is no logical relationship between the maximality of $\mathcal{A}$ and the maximality of $\overline{A_n}$. To illustrate this, we shall use an example of two maximal monotone operators A and B provided in [19, Page 54], such that $\mathrm{dom}(\mathsf{A} + \mathsf{B}) = \{0\}$ but $\mathsf{A} + \mathsf{B} \neq N_{\{0\}}$ (and hence $\mathsf{A} + \mathsf{B}$ is not maximal). If A is uniformly distributed over $\{\mathsf{A}, \mathsf{B}, N_{\{0\}}\}$, then, $\mathcal{A}$ is maximal but with positive probability $\overline{A_2}$ is not maximal. If A is uniform over $\{\mathsf{A}, \mathsf{B}\}$, then, with positive probability, $\overline{A_2}$ is maximal although $\mathcal{A}$ is not maximal.

Remark 1. Regarding the subdifferential case where $A(s) = \partial f(s, \cdot)$, we say that the interchange property holds if $\mathcal{A}(x) = \partial F(x)$ where $F(x) = \mathbb{E}_\xi(f(\xi, x))$. The interchange property means that one can exchange the expectation $\mathbb{E}$ and the subdifferentiation ∂. If the interchange property holds, and if $F \in \Gamma_0(H)$, then A is integrable. Moreover, in this case, the function G appearing in the conclusion of the Theorem 1 can be set equal to F (see [3, Chap I, Section 3] or [4, 13] for the topology induced by R-convergence over $\mathscr{M}_s(H)$).

The interchange property holds under quite general assumptions on f, see [20]. The condition $F \in \Gamma_0(H)$ holds if there exist $x_0 \in H$ and $C \in \mathbb{R}$ such that $f(\cdot, x_0)$ is μ-integrable and for every $x \in H$, $f(s, x) \geq C$ a.s. To see this, first note that F is convex and proper because $f(s, \cdot)$ is convex and proper. In the case where $C = 0$, the lower semicontinuity of F is a consequence of Fatou's lemma along with the sequential lower semicontinuity of $f(s, \cdot)$. The lower semicontinuity in the case where $C \neq 0$ is obtained from the case $C = 0$ by replacing $f(s, \cdot)$ by $f(s, \cdot) - C$. Finally note that $f(s, x) \geq C$ a.s. usually holds when F is the objective function of some minimization problem (otherwise there might be no minimizer).

4 Proof of the main result

Since $\overline{A_2}$ is maximal, it is a random monotone operator using [3, Theorem 2.4]. An alternative proof of the measurability of $\overline{A_2}$ is as follows: for every $y \in H$, $x = J_{\overline{A_2}}(y)$ is the solution to the monotone inclusion $0 \in (I - y)(x) + \frac{1}{2}A(\xi_1(\omega), x) + \frac{1}{2}A(\xi_2(\omega), x)$ for which the three operator splitting algorithm of [14] can be applied. It provides a sequence of iterates $(x_n(\omega))$ converging to x. It can be proven by induction that $\omega \mapsto x_n(\omega)$ is measurable. Therefore $J_{\overline{A_2}}(y)$ is also a random variable for every $y \in H$, which proves the measurability of $\overline{A_2}$ [3, Lemma 2.1]. Then, by induction, $\overline{A_n}$ is a random monotone operator for every n.

Lemma 2. Under the assumptions of Theorem 1, if $x_\star \in Z(\mathcal{A})$ then,

$$J_{\overline{A_n}}(x_\star) \longrightarrow x_\star,$$

as $n \to +\infty$, $\mathbb{P}$-a.s.

Proof. Since $0 \in \mathcal{A}(x_\star)$, there exists a measurable map $\varphi : (\Xi, \mathscr{G}, \mu) \to (H, \mathscr{B}(H))$ such that φ is μ-integrable, $\int \varphi d\mu = 0$ and $\varphi(s) \in A(s, x_\star)$ μ-a.s. Consider the random variables $\overline{\phi_n} = \frac{1}{n}\sum_{k=1}^{n} \varphi(\xi_k)$. Note that $\overline{\phi_n}$ is integrable, $\overline{\phi_n} \in \overline{A_n}(x_\star)$ $\mathbb{P}$-a.s. and $\mathbb{E}(\overline{\phi_n}) = 0$. Using inequality (1) with $\gamma = 1$ and $x = x_\star \in \operatorname{dom}(\overline{A_n})$ a.s.,

$$\|J_{\overline{A_n}}(x_\star) - x_\star\| \le \|(\overline{A_n})_0(x_\star)\| \le \|\overline{\phi_n}\|, \quad \mathbb{P}\text{-a.s.}$$

Using the Strong Law of Large Numbers in Hilbert spaces ([16, Corollary 7.10]) for $\overline{\phi_n}$ we have $\mathbb{P}$-a.s.,

$$\|\overline{\phi_n}\| \longrightarrow_{n\to+\infty} 0.$$

and hence $\mathbb{P}$-a.s.

$$\|J_{\overline{A_n}}(x_\star) - x_\star\| \longrightarrow_{n\to+\infty} 0.$$

□

Lemma 3. Consider $z \in H$. Then, $A - z : x \mapsto A(x) - z$ is a random monotone operator and

$$J_{A-z}(y) = J_A(y + z), \quad \forall y \in H. \tag{4}$$

Proof. Equation (4) is classical and can be found for example in [7], but we provide a full proof here for the sake of completeness. For any $y \in H$, the inclusion $y \in x + (A - z)(x)$ (where x is the unknown) is equivalent to $y + z \in x + A(x)$ and hence admits an unique solution $x = J_A(y + z)$. This implies that $A - z$ is μ-a.s a maximal monotone operator, and $J_{A-z}(y) = J_A(y + z)$. It is also seen that $s \mapsto J_{A(s)-z}(y)$ is measurable for every $y \in H$ and hence, $A - z$ is a random monotone operator (see [3, Lemme 2.1]). □

We now prove the Theorem 1. Consider $x \in H$. Since $\operatorname{dom}(J_{\mathcal{A}}) = H$, there exists an unique $(y, z) \in G(\mathcal{A})$ such that $x = y + z$. Therefore, $0 \in \mathcal{A}(y) - z$ *i.e* $y \in Z(\mathcal{A} - z)$. Using Lemma 3 and the maximality of $\overline{A_n}$, $A - z$ is a random monotone operator and $\frac{1}{n}\sum_{k=1}^{n}(A(\xi_k) - z) = \overline{A_n} - z$ is μ-a.s. maximal. Moreover, $A - z$ is μ-integrable with $\int (A - z)d\mu = \mathcal{A} - z$. Applying Lemma 2 to the random monotone operator $A - z$, we have $\mathbb{P}$-a.s,

$$J_{\overline{A_n}-z}(y) \longrightarrow y. \tag{5}$$

Using $y = J_{\mathcal{A}-z}(y)$, $x = y + z$ and Lemma 3, the convergence (5) can be rewritten as follows: for every $x \in H$, there exists a probability one event $\Omega_x \subset \Omega$ such that for every $\omega \in \Omega_x$,

$$J_{\overline{A_n(\omega)}}(x) \longrightarrow J_{\mathcal{A}}(x).$$

We now show that Ω_x can be taken independent of x. Consider a dense enumerable subset D of H, and the probability one event $\tilde{\Omega} = \bigcap_{x\in D} \Omega_x$. For every $\omega \in \tilde{\Omega}$, we have for every $x \in D$,

$$J_{\overline{A_n(\omega)}}(x) \longrightarrow J_{\mathcal{A}}(x).$$

Consider $x_0 \in H$. We shall prove that for every $\omega \in \tilde{\Omega}$, we also have

$$J_{\overline{A_n(\omega)}}(x_0) \longrightarrow J_{\mathcal{A}}(x_0).$$

Let $\varepsilon > 0$ and $x \in D$ such that $\|x - x_0\| < \varepsilon/3$. There exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$, $\|J_{\overline{A_n(\omega)}}(x) - J_{\mathcal{A}}(x)\| < \varepsilon/3$. Let us decompose

$$\begin{aligned}&\|J_{\overline{A_n(\omega)}}(x_0) - J_{\mathcal{A}}(x_0)\| \\ \leq &\|J_{\overline{A_n(\omega)}}(x) - J_{\mathcal{A}}(x)\| + \|J_{\overline{A_n(\omega)}}(x_0) - J_{\overline{A_n(\omega)}}(x)\| + \|J_{\mathcal{A}}(x_0) - J_{\mathcal{A}}(x)\|.\end{aligned}$$

Since resolvents are 1-Lipschitz continuous, $\|J_{\overline{A_n(\omega)}}(x_0) - J_{\mathcal{A}}(x_0)\| < \varepsilon$ for every $n \geq n_0$. We proved that for every $\omega \in \tilde{\Omega}$, $J_{\overline{A_n(\omega)}}(x) \longrightarrow J_{\mathcal{A}}(x)$, for every $x \in H$ *i.e.* $\overline{A_n}(\omega) \longrightarrow \mathcal{A}$.

In the case where $A(s) = \partial f(s, \cdot)$, $\overline{A_n}$ is a.s a sum of elements of $\mathscr{M}_s(H)$ and a maximal monotone operator. Therefore, $\overline{A_n} \in \mathscr{M}_s(H)$ a.s.[1] Since $\overline{A_n}$ converges a.s. to $\mathcal{A}$ and since $\mathscr{M}_s(H)$ is a sequentially closed subset of $\mathscr{M}(H)$ for the R-convergence ([3, Prop 1.3]), $\mathcal{A} \in \mathscr{M}_s(H)$.

5 Application to empirical risk minimization

Many machine learning and signal processing problems require to solve the so-called *theoretical risk minimization* problem

$$\min_{x \in H} F(x) := \mathbb{E}_\xi(f(\xi, x)) \tag{6}$$

where f is a convex normal integrand and ξ a random variable. In these contexts, ξ represents some random data with unknown distribution and hence evaluating F is prohibitive. In practice, a number n of i.i.d realizations (ξ_k) of the data ξ is given and the theoretical risk minimization is approximated by the *empirical mean minimization* problem

$$\min_{x \in H} \overline{f_n}(x) := \frac{1}{n} \sum_{k=1}^{n} f(\xi_k, x). \tag{7}$$

The empirical risk minimization is usually done using some optimization algorithm. The output of the algorithm can be a minimizer of $\overline{f_n}$ or a saddle point of a convex concave Lagrangian function associated to $\overline{f_n}$. Since $\overline{f_n}$ converges to F, one can expect the output to be close to its theoretical value if n is large. The convergence of the output as $n \to \infty$ can be studied using Epi-convergence tools [4, 6, 13]. The output is also a zero of a random monotone operator $\overline{A_n}$ taking the form of (2). In order to study the output as $n \to \infty$, we provide a simple consequence of the law of large numbers (Theorem 1) w.r.t. to $Z(\overline{A_n})$ as $n \to \infty$.

A random variable ℓ is an a.s. cluster point of the sequence (x_n) of random variables if there exists a probability one event $\tilde{\Omega}$ such that for every $\omega \in \tilde{\Omega}$ there exists a subsequence of $x_n(\omega)$ converging to $\ell(\omega)$. The subsequence of $x_n(\omega)$ is called a random subsequence of x_n.

[1] If $F_1, F_2 \in \Gamma_0(H)$, it is always true that $\partial F_1 + \partial F_2$ and $\partial(F_1 + F_2)$ are monotone operators, and that for every $x \in H$, $\partial F_1(x) + \partial F_2(x) \subset \partial(F_1 + F_2)(x)$. Therefore $G(\partial F_1 + \partial F_2) \subset G(\partial(F_1 + F_2))$. If it is known that $\partial F_1 + \partial F_2$ is maximal, then $G(\partial F_1 + \partial F_2) = G(\partial(F_1 + F_2))$ *i.e* $\partial F_1 + \partial F_2 = \partial(F_1 + F_2) \in \mathscr{M}_s(H)$.

Corollary 4. Let (x_n) be a sequence of H-valued random variables such that $x_n \in Z(\overline{A_n})$ a.s. Then, every a.s. cluster point ℓ of (x_n) is a.s. a zero of $\mathcal{A}$.

Proof. Consider a random subsequence of (x_n) converging a.s. to ℓ. This random subsequence is still denoted (x_n). Denote $\overline{A_{n,\gamma}}$ the Yosida approximation of $\overline{A_n}$ and set $\gamma = 1$. For every $n \geq 0$, $\overline{A_{n,\gamma}}(x_n) = 0$. Therefore,

$$\begin{aligned}\|\ell - J_{\mathcal{A}}(\ell)\| = \|\mathcal{A}_\gamma(\ell)\| &\leq \|\mathcal{A}_\gamma(\ell) - \overline{A_{n,\gamma}}(\ell)\| + \|\overline{A_{n,\gamma}}(\ell) - \overline{A_{n,\gamma}}(x_n)\| \\ &\leq \|J_{\mathcal{A}}(\ell) - J_{\overline{A_n}}(\ell)\| + \|\ell - x_n\|.\end{aligned}$$

Using the law of large numbers (Theorem 1) one can make $n \to +\infty$ and obtain $\ell = J_{\mathcal{A}}(\ell)$ a.s. □

The existence of cluster points is usually established independently using a coercivity assumption [9, 13].

6 Conclusion

We proved a law of large numbers for random monotone operators. This work opens the door to the study of random monotone operators as random elements. An interesting question is whether their exists an universal distribution for random monotone operators, as the gaussian distribution for real random variables, or other probabilistic objects, see *e.g* [15, 16, 17].

References

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

[2] Zvi Artstein and Richard A Vitale. A strong law of large numbers for random compact sets. *The Annals of Probability*, pages 879–882, 1975.

[3] Hédy Attouch. Familles d'opérateurs maximaux monotones et mesurabilité. *Ann. Mat. Pura Appl.*, 120(1):35–111, 1979.

[4] Hédy Attouch. *Variational convergence for functions and operators*, volume 1. Pitman Advanced Publishing Program, 1984.

[5] Robert J Aumann. Integrals of set-valued functions. *J. Math. Anal. Appl.*, 12:1–12, 1965.

[6] Dominique Azé, Hédy Attouch, and Roger J-B Wets. Convergence of convex-concave saddle functions: applications to convex programming and mechanics. In *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*, volume 5, pages 537–572. Elsevier, 1988.

[7] Heinz H Bauschke and Patrick L Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.

[8] Pascal Bianchi and Walid Hachem. Dynamical behavior of a stochastic Forward-Backward algorithm using random monotone operators. *J. Optim. Theory Appl.*, 171(1):90–120, 2016.

[9] Pascal Bianchi, Walid Hachem, and Adil Salim. A constant step Forward-Backward algorithm involving random maximal monotone operators. *J. Convex Anal.*, 26(2):397–436, 2019.

[10] Léon Bottou and Olivier Bousquet. The tradeoffs of large scale learning. In *NIPS*, pages 161–168, 2008.

[11] Haïm Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland mathematics studies. Elsevier Science, Burlington, MA, 1973.

[12] Charles Castaing and Paul Raynaud de Fitte. Law of large numbers and ergodic theorem for convex weak star compact valued gelfand-integrable mappings. In *Advances in Mathematical Economics Volume 17*, pages 1–37. Springer, 2013.

[13] Gianni Dal Maso. *An introduction to* Γ*-convergence*, volume 8. Springer Science & Business Media, 2012.

[14] Damek Davis and Wotao Yin. A three-operator splitting scheme and its optimization applications. *Set-valued and variational analysis*, 25(4):829–858, 2017.

[15] Jean-François Le Gall. Uniqueness and universality of the brownian map. *The Annals of Probability*, 41(4):2880–2960, 2013.

[16] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013.

[17] Vladimir A Marčenko and Leonid A Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.

[18] George J Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29(3):341–346, 1962.

[19] Robert R Phelps. *Convex functions, monotone operators and differentiability*, volume 1364. Springer, 2009.

[20] Ralph T Rockafellar and Roger J-B Wets. On the interchange of subdifferentiation and conditional expectations for convex functionals. *Stochastics*, 7(3):173–182, 1982.

[21] Alexander Shapiro and Huifu Xu. Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions. *Journal of mathematical analysis and applications*, 325(2):1390–1399, 2007.

[22] Robert L Taylor and Hiroshi Inoue. *Laws of Large Numbers for Random Sets*, pages 347–360. Springer New York, New York, NY, 1997.

[23] Pedro Terán. On a uniform law of large numbers for random sets and subdifferentials of random functions. *Statistics & Probability letters*, 78(1):42–49, 2008.