Kybernetika

Petr Lachout

Approximative solutions of stochastic optimization problems

Kybernetika, Vol. 46 (2010), No. 3, 513--523

Persistent URL: `http://dml.cz/dmlcz/140765`

KYBERNETIKA — VOLUME *46* (2010), NUMBER 3, PAGES 513–523

APPROXIMATIVE SOLUTIONS OF STOCHASTIC OPTIMIZATION PROBLEMS

Petr Lachout

The aim of this paper is to present some ideas how to relax the notion of the optimal solution of the stochastic optimization problem. In the deterministic case, ε-minimal solutions and level-minimal solutions are considered as desired relaxations. We call them approximative solutions and we introduce some possibilities how to combine them with randomness. Relations among random versions of approximative solutions and their consistency are presented in this paper. No measurability is assumed, therefore, treatment convenient for nonmeasurable objects is employed.

Keywords: the optimal solution, ε-minimal solutions, level-minimal solutions, randomness

Classification: 90C31, 62F12, 60F99

1. INTRODUCTION

We consider a general scheme of the stochastic optimization problem and we address the question what could mean "a solution" of it. We work on metric spaces and consider precise and approximate solutions of the stochastic optimization problem. We do not require measurability in this paper and uniqueness of the optimal solution is not assumed. Observed data, approximations and considered functions are maps from probability space to a metric space, only. Therefore, we become to be out of the standard theory based on the measurability assumption. The theory for nonmeasurable objects we employ is nicely explained in [8].

Working without the measurability assumption, there are several concepts for convergences almost surely and in probability. These definitions together with relations between them can be found in [8], chapter 1.9, pp. 52–56.

Our paper is closely related to the stochastic estimation theory because statistical estimators are usually defined as a solution of some particular stochastic optimization problem. For example, the concept of asymptotic minimizers is inspired by [7] where the authors introduced the Asymptotically Optimal Estimators (AOE).

2. CONSIDERED OPTIMIZATION PROBLEM

We consider an optimization problem written in the form

$$\inf \{ \mathsf{f}\,(x\,;\mu_0) \ \mid x \in \mathcal{X} \}\,, \tag{1}$$

where $\mu_0 \in \mathcal{P}$. We suppose $\mathcal{X}$ to be a metric space, $\mathcal{P}$ be a family of probability measures defined on a metric space $\mathcal{Y}$ and $f : \mathcal{X} \times \mathcal{P} \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} = [-\infty, +\infty]$ denotes the extended real line.

The objective function is known up to unknown probability measure μ_0. We assume a procedure producing an estimation of μ_0:

We suppose to observe $Z_t \in \mathcal{Z}_t$ at any time $t \in \mathbb{N}$. From observed data we construct probability measure $\mu_t(\bullet \,|\, Z_t)$ on $\mathcal{Y}$. This measure will play role of estimator for the "true" probability measure μ_0.

Typically, we observe a sequence of data $Y_1, Y_2, Y_3, \ldots$ belonging to a metric space $\mathcal{Y}$. Hence, we group observations available at time $t \in \mathbb{N}$ in a vector $Z_t = (Y_1, Y_2, \ldots, Y_{k_t})$, $\mathcal{Z}_t = \mathcal{Y}^{k_t}$ and $\mu_t(\bullet \,|\, Z_t) = \frac{1}{k_t} \sum_{i=1}^{k_t} \delta_{Y_i}$ is the empirical measure.

Let us introduce a denotation of objects of our interest. For a given function $f : \mathcal{X} \to \overline{\mathbb{R}}$ we are interested in its minimal value

$$\varphi\,[f] = \inf \{ f(x) \ | \ x \in \mathcal{X} \} \tag{2}$$

and in the set of all minimal solutions

$$\Phi \langle f \rangle = \{ x \in \mathcal{X} \ | \ f(x) = \varphi\,[f] \} . \tag{3}$$

Having $\varphi\,[f] \in \mathbb{R}$, we will consider the sets of all ε-minimal solutions

$$\Psi \langle f; \varepsilon \rangle = \{ x \in \mathcal{X} \ | \ f(x) \leq \varphi\,[f] + \varepsilon \} \quad \forall\, \varepsilon \in \mathbb{R}, \tag{4}$$

and the level sets

$$\mathsf{lev}_\Delta \langle f \rangle = \{ x \in \mathcal{X} \ | \ f(x) \leq \Delta \} \quad \forall\, \Delta \in \mathbb{R}. \tag{5}$$

Let us note that

$$\begin{aligned} &\Psi \langle f; \varepsilon \rangle = \emptyset \text{ if } \varepsilon < 0, \ \Psi \langle f; \varepsilon \rangle \neq \emptyset \text{ if } \varepsilon > 0, \\ &\mathsf{lev}_\Delta \langle f \rangle = \emptyset \text{ if } \Delta < \varphi\,[f]\,, \ \mathsf{lev}_\Delta \langle f \rangle \neq \emptyset \text{ if } \Delta > \varphi\,[f]\,. \end{aligned}$$

There is a direct relation between the level-optimal solutions and the ε-optimal solutions:

$$\Psi \langle f; \varepsilon \rangle = \mathsf{lev}_{\varphi[f]+\varepsilon} \langle f \rangle\,, \tag{6}$$

$$\mathsf{lev}_\Delta \langle f \rangle = \Psi \langle f; \Delta - \varphi\,[f] \rangle\,. \tag{7}$$

Therefore, results on the ε-optimal solutions can be easily translated to the level-optimal solutions, and vice versa.

We will need to measure "distance" between sets. Convenient for our purposes is the excess from set $A \subset \mathcal{X}$ to set $B \subset \mathcal{X}$

$$\mathsf{excess}\,(A, B) = \sup_{a \in A} \inf_{b \in B} d(a, b), \tag{8}$$

where d is the metric of $\mathcal{X}$. Let us note that the excess itself is no metric. Nevertheles, it forms Hausdorff metric

$$\mathcal{H}\,(A, B) = \mathsf{excess}\,(A, B) + \mathsf{excess}\,(B, A)\,.$$

3. APPROXIMATIVE SOLUTIONS

3.1. Deterministic case

In deterministic case there are two possibilities how to approximate the optimal solutions of (2). We can consider ε-minimal solutions or level-solutions, i. e. we consider a sequence $\hat{\theta}_t \in \mathcal{X}$ such that either

$$\hat{\theta}_t \in \Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot|z_t)\right); \varepsilon_t \right\rangle \quad \forall\, t \in \mathbb{N} \tag{9}$$

or

$$\hat{\theta}_t \in \mathsf{lev}_{\Delta_t} \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot|z_t)\right) \right\rangle \quad \forall\, t \in \mathbb{N}. \tag{10}$$

Our aim is to derive consistency of this sequence.

Let us start with ε-minimal solutions. Let $\varepsilon_t \geq 0$, $z_t \in \mathcal{Z}_t$, $t \in \mathbb{N}$ be given and denote $\bar{\varepsilon} = \limsup\limits_{t \to +\infty} \varepsilon_t$. We introduce a set of properties

$$\varphi\left[\mathsf{f}\left(\bullet\,; \mu_t(\cdot|z_t)\right)\right] \in \mathbb{R} \quad \forall\, t \in \mathbb{N}, \tag{11}$$

$$\varphi\left[\mathsf{f}\left(\bullet\,; \mu_0\right)\right] \in \mathbb{R}, \tag{12}$$

$$\lim_{t \to +\infty} \varphi\left[\mathsf{f}\left(\bullet\,; \mu_t(\cdot|z_t)\right)\right] = \varphi\left[\mathsf{f}\left(\bullet\,; \mu_0\right)\right], \tag{13}$$

$$\Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot|z_t)\right); \varepsilon_t \right\rangle \neq \emptyset \quad \forall\, t \in \mathbb{N}, \tag{14}$$

$$\Phi \left\langle \mathsf{f}\left(\bullet\,; \mu_0\right) \right\rangle \neq \emptyset, \tag{15}$$

$$\text{there is a compact } \mathsf{K} \text{ such that} \tag{16}$$

$$\Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot|z_t)\right); \varepsilon_t \right\rangle \subset \mathsf{K} \quad \forall\, t \in \mathbb{N},$$

$$\mathsf{Ls}\left\{\Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot|z_t)\right); \varepsilon_t \right\rangle \,,\, t \in \mathbb{N}\right\} \subset \Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_0\right); \bar{\varepsilon} \right\rangle, \tag{17}$$

$$\lim_{t \to +\infty} \mathsf{excess}\left(\Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot|z_t)\right); \varepsilon_t \right\rangle, \Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_0\right); \bar{\varepsilon} \right\rangle\right) = 0, \tag{18}$$

where $\mathsf{Ls}\{A_t \,,\, t \in \mathbb{N}\}$ denotes the set of all cluster points of the sequence of sets A_t, $t \in \mathbb{N}$, i. e. set of all points reachable by a sequence $a_{t_n} \in A_{t_n}$, $n \in \mathbb{N}$.

These properties imply consistency of ε-minimal solutions.

Proposition 3.1. Let (11) – (18) be fulfilled and a sequence $\hat{\theta}_t$, $t \in \mathbb{N}$ be given by (9). Then,

$$\varphi\left[\mathsf{f}\left(\bullet\,; \mu_0\right)\right] \leq \liminf_{t \to +\infty} \mathsf{f}\left(\hat{\theta}_t\,; \mu_t(\cdot|z_t)\right) \tag{19}$$

$$\leq \limsup_{t \to +\infty} \mathsf{f}\left(\hat{\theta}_t\,; \mu_t(\cdot|z_t)\right) \leq \varphi\left[\mathsf{f}\left(\bullet\,; \mu_0\right)\right] + \bar{\varepsilon},$$

$$\text{the sequence } \hat{\theta}_t,\ t \in \mathbb{N} \text{ is compact}, \tag{20}$$

$$\mathsf{Ls}\left\{\{\hat{\theta}_t\} \,,\, t \in \mathbb{N}\right\} \subset \Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_0\right); \bar{\varepsilon} \right\rangle, \tag{21}$$

$$\lim_{t \to +\infty} \mathsf{excess}\left(\{\hat{\theta}_t\}, \Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_0\right); \bar{\varepsilon} \right\rangle\right) = 0. \tag{22}$$

P r o o f . (19) follows (13) and definition of $\bar{\varepsilon}$, (16) implies (20), (17) implies (21), and (18) implies (22). $\square$

A set of assumptions giving (11) – (18) is presented in [1].

Similar observation is true for the level-optimal solutions. Let $\Delta_t \in \mathbb{R}$, $z_t \in \mathcal{Z}_t$, $t \in \mathbb{N}$ be given and denote $\bar{\Delta} = \limsup_{t \to +\infty} \Delta_t$. We introduce a set of properties

$$\Delta_t \geq \varphi\left[\mathsf{f}\left(\bullet\,;\mu_t(\cdot|z_t)\right)\right] \in \mathbb{R} \quad \forall\, t \in \mathbb{N}, \tag{23}$$

$$\varphi\left[\mathsf{f}\left(\bullet\,;\mu_0\right)\right] \in \mathbb{R}, \tag{24}$$

$$\lim_{t \to +\infty} \varphi\left[\mathsf{f}\left(\bullet\,;\mu_t(\cdot|z_t)\right)\right] = \varphi\left[\mathsf{f}\left(\bullet\,;\mu_0\right)\right], \tag{25}$$

$$\mathsf{lev}_{\Delta_t}\left\langle \mathsf{f}\left(\bullet\,;\mu_t(\cdot|z_t)\right)\right\rangle \neq \emptyset \quad \forall\, t \in \mathbb{N}, \tag{26}$$

$$\Phi\left\langle \mathsf{f}\left(\bullet\,;\mu_0\right)\right\rangle \neq \emptyset, \tag{27}$$

$$\text{there is a compact } \mathsf{K} \text{ such that} \tag{28}$$

$$\mathsf{lev}_{\Delta_t}\left\langle \mathsf{f}\left(\bullet\,;\mu_t(\cdot|z_t)\right)\right\rangle \subset \mathsf{K} \quad \forall\, t \in \mathbb{N},$$

$$\mathsf{Ls}\left\{\mathsf{lev}_{\Delta_t}\left\langle \mathsf{f}\left(\bullet\,;\mu_t(\cdot|z_t)\right)\right\rangle\,,\, t \in \mathbb{N}\right\} \subset \mathsf{lev}_{\bar{\Delta}}\left\langle \mathsf{f}\left(\bullet\,;\mu_0\right)\right\rangle, \tag{29}$$

$$\lim_{t \to +\infty} \mathsf{excess}\left(\mathsf{lev}_{\Delta_t}\left\langle \mathsf{f}\left(\bullet\,;\mu_t(\cdot|z_t)\right)\right\rangle, \mathsf{lev}_{\bar{\Delta}}\left\langle \mathsf{f}\left(\bullet\,;\mu_0\right)\right\rangle\right) = 0. \tag{30}$$

Again, these properties imply consistency of level-minimal solutions.

Proposition 3.2. Let (23) – (30) be fulfilled and a sequence $\hat{\theta}_t$, $t \in \mathbb{N}$ be given by (10). Then,

$$\varphi\left[\mathsf{f}\left(\bullet\,;\mu_0\right)\right] \leq \liminf_{t \to +\infty} \mathsf{f}\left(\hat{\theta}_t\,;\mu_t(\cdot|z_t)\right) \tag{31}$$

$$\leq \limsup_{t \to +\infty} \mathsf{f}\left(\hat{\theta}_t\,;\mu_t(\cdot|z_t)\right) \leq \bar{\Delta},$$

$$\text{the sequence } \hat{\theta}_t,\ t \in \mathbb{N} \text{ is compact}, \tag{32}$$

$$\mathsf{Ls}\left\{\{\hat{\theta}_t\}\,,\, t \in \mathbb{N}\right\} \subset \mathsf{lev}_{\bar{\Delta}}\left\langle \mathsf{f}\left(\bullet\,;\mu_0\right)\right\rangle, \tag{33}$$

$$\lim_{t \to +\infty} \mathsf{excess}\left(\{\hat{\theta}_t\}, \mathsf{lev}_{\bar{\Delta}}\left\langle \mathsf{f}\left(\bullet\,;\mu_0\right)\right\rangle\right) = 0. \tag{34}$$

P r o o f . Using relations (6), (7) we can see that lemma 3.2 coincides with lemma 3.1. Therefore, any new proof is not necessary. $\square$

A set of assumptions presented in [1] is giving (23) – (30) because of (6), (7).

3.2. Randomness

Now, we start to combine approximative solutions with randomness.

We assume a probability space $(\Omega, \mathcal{A}, \mathsf{Prob})$, $\mathsf{f} : \mathcal{X} \times \mathcal{P} \times \Omega \to \mathbb{R}$ and $\varepsilon_t : \Omega \to \mathbb{R}$, $\varepsilon_t > 0$, $Z_t : \Omega \to \mathcal{Z}_t$ for each $t \in \mathbb{N}$. Also, "true" probability measure $\mu_0\left(\bullet|\,\omega\right)$ is assumed to be dependent on random event $\omega \in \Omega$.

Let us make a short comment to defend the suggested dependence on randomness. The random event $\omega \in \Omega$ corresponds to observed data. Therefore for $\omega \in \Omega$ with observations of higher computational complexity, the error control $\varepsilon_t(\omega)$ must be also higher. The theoretical probability measure μ_0 could naturally depend on random events. For example, if our observations forms a strictly stationary sequence then it is known that a limit of their relative frequencies always exists, coincides with the probability measure μ_0, and depends on random events. If the observed data possess ergodic property (e. g. i.i.d. sequence) then the limit is deterministic.

Considered objects are maps, only. We assume no measurability for them. Now, convergence can be considered in several senses. We will deal just with three of them: the convergence almost surely (as), almost uniformly (au), and, in outer probability (Prob*). These types of convergences are explained in [8]. For convenience, we placed basic definitions and properties in Appendix of our paper; see Definition 4.2.

Involving randomness, we can investigate several generalizations of ε-minimal solutions and level-minimal solutions. Let us mention some of them.

A map $\hat{\theta}_t : \Omega \to \mathcal{X}$ is called

random ε-minimal solution, whenever,

$$\hat{\theta}_t(\omega) \in \Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot \,|\, Z_t(\omega)), \omega\right); \varepsilon_t(\omega)\right\rangle \qquad \forall\ \omega \in \Omega\ \forall\ t \in \mathbb{N},$$

random level-minimal solution, whenever,

$$\hat{\theta}_t(\omega) \in \mathsf{lev}_{\Delta_t(\omega)} \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot \,|\, Z_t(\omega)), \omega\right)\right\rangle \qquad \forall\ \omega \in \Omega\ \forall\ t \in \mathbb{N}.$$

Slightly weaker notions are

almost sure ε-minimal solution, if there is $\Omega_O \subset \Omega$ with $\mathsf{Prob}\,(\Omega_O) = 1$ such that

$$\hat{\theta}_t(\omega) \in \Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot \,|\, Z_t(\omega)), \omega\right); \varepsilon_t(\omega)\right\rangle \qquad \forall\ \omega \in \Omega_O\ \forall\ t \in \mathbb{N},$$

almost sure level-minimal solution, if there is $\Omega_O \subset \Omega$ with $\mathsf{Prob}\,(\Omega_O) = 1$ such that

$$\hat{\theta}_t(\omega) \in \mathsf{lev}_{\Delta_t(\omega)} \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot \,|\, Z_t(\omega)), \omega\right)\right\rangle \qquad \forall\ \omega \in \Omega_O\ \forall\ t \in \mathbb{N}.$$

Following ideas of the Asymptotically Optimal Estimators (AOE) introduced in [7], we can investigate:

strict asymptotically ε-minimal solution, whenever, there is $\Omega_O \subset \Omega$ with $\mathsf{Prob}\,(\Omega_O) = 1$ such that for every $\omega \in \Omega_O$

$$\hat{\theta}_t(\omega) \in \Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot \,|\, Z_t(\omega)), \omega\right); \varepsilon_t(\omega)\right\rangle \qquad \forall\ t \in \mathbb{N} \quad \text{sufficiently large},$$

strict asymptotically level-minimal solution, whenever, there is $\Omega_O \subset \Omega$ with $\mathsf{Prob}\,(\Omega_O) = 1$ such that for every $\omega \in \Omega_O$

$$\hat{\theta}_t(\omega) \in \mathsf{lev}_{\Delta_t(\omega)} \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot \,|\, Z_t(\omega)), \omega\right)\right\rangle \qquad \forall\ t \in \mathbb{N} \quad \text{sufficiently large},$$

weak asymptotically ε-minimal solution, whenever,

$$\lim_{t\to+\infty} \mathsf{Prob}_* \left(\omega \in \Omega\ :\ \hat{\theta}_t(\omega) \in \Psi \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot \,|\, Z_t(\omega)), \omega\right); \varepsilon_t(\omega)\right\rangle\right) = 1,$$

weak asymptotically level-minimal solution, whenever,

$$\lim_{t\to+\infty} \mathsf{Prob}_* \left(\omega \in \Omega \ :\ \hat{\theta}_t(\omega) \in \mathsf{lev}_{\Delta_t(\omega)} \left\langle \mathsf{f}\left(\bullet\,;\mu_t(\cdot\mid Z_t(\omega)),\omega\right)\right\rangle\right) = 1.$$

Immediately from the definitions, we see that:

- Random ε-minimal solution $\Rightarrow$ almost sure ε-minimal solution.
- Almost sure ε-minimal solution $\Rightarrow$ strict asymptotically ε-minimal solution.
- Strict asymptotically ε-minimal solution $\Rightarrow$ weak asymptotically ε-minimal solution.
- Random level-minimal solution $\Rightarrow$ almost sure level-minimal solution.
- Almost sure level-minimal solution $\Rightarrow$ strict asymptotically level-minimal solution.
- Strict asymptotically level-minimal solution $\Rightarrow$ weak asymptotically level-minimal solution.

It is sufficient to investigate random ε-minimal solutions and random level-minimal solutions, only. It is because there is a natural construction converting the other types of approximative solutions to these two.

Construction:

Choose arbitrary sequence $\alpha_t : \Omega \to \mathbb{R}$, $\alpha_t > 0$. Then, $\varepsilon_t(\omega) + \alpha_t(\omega) > 0$ and one can select $\xi_t : \Omega \to \mathcal{X}$, $t \in \mathbb{N}$ such that

$$\xi_t(\omega) \in \Psi\left\langle \mathsf{f}\left(\bullet\,;\mu_t(\cdot\mid Z_t(\omega)),\omega\right);\varepsilon_t(\omega)+\alpha_t(\omega)\right\rangle \quad \forall\,\omega\in\Omega,\ \forall\, t\in\mathbb{N}.$$

For a map $\hat{\theta}_t : \Omega \to \mathcal{X}$ we set

$$\begin{aligned} \eta_t(\omega) &= \hat{\theta}_t(\omega) \quad \text{if } \hat{\theta}_t(\omega) \in \Psi\left\langle \mathsf{f}\left(\bullet\,;\mu_t(\cdot\mid Z_t(\omega)),\omega\right);\varepsilon_t(\omega)\right\rangle \\ &= \xi_t(\omega) \quad \text{otherwise.} \end{aligned}$$

The construction produces $\eta_t : \Omega \to \mathcal{X}$, $t \in \mathbb{N}$ fulfilling

$$\eta_t(\omega) \in \Psi\left\langle \mathsf{f}\left(\bullet\,;\mu_t(\cdot\mid Z_t(\omega)),\omega\right);\varepsilon_t(\omega)+\alpha_t(\omega)\right\rangle \quad \forall\,\omega\in\Omega,\ \forall\, t\in\mathbb{N}.$$

Thus, η_t is a random ε-minimal solution of course for enlarged epsilon $\varepsilon_t + \alpha_t$. If, moreover, $\lim_{t\to+\infty} \alpha_t(\omega) = 0$ then

$$\limsup_{t\to+\infty} \varepsilon_t(\omega) + \alpha_t(\omega) = \limsup_{t\to+\infty} \varepsilon_t(\omega) = \bar{\varepsilon}(\omega).$$

Now, we can compare asymptotic properties of $\hat{\theta}_t$, $t \in \mathbb{N}$ and η_t, $t \in \mathbb{N}$.

Theorem 3.3. Let $\hat{\theta}_t : \Omega \to \mathcal{X}$ be a strict asymptotically ε-minimal solution.
Then, there is $\Omega_O \subset \Omega$ with $\mathsf{Prob}\,(\Omega_O) = 1$ such that for every $\omega \in \Omega_O$

$$\hat{\theta}_t(\omega) = \eta_t(\omega) \quad \forall\, t \in \mathbb{N} \quad \text{sufficiently large.} \tag{35}$$

Consequently for every $\omega \in \Omega_O$,

$$\begin{aligned}
&\text{the sequence } \hat{\theta}_t(\omega),\ t \in \mathbb{N} \text{ is compact if and only if} && (36)\\
&\qquad \text{the sequence } \eta_t(\omega),\ t \in \mathbb{N} \text{ is compact,} \\
&\mathsf{Ls}\left\{\hat{\theta}_t(\omega)\,,\, t \in \mathbb{N}\right\} = \mathsf{Ls}\left\{\eta_t(\omega)\,,\, t \in \mathbb{N}\right\}, && (37)\\
&\mathsf{f}\left(\hat{\theta}_t(\omega)\,;\mu_t(\cdot \mid Z_t(\omega)), \omega\right) = \mathsf{f}\left(\eta_t(\omega)\,;\mu_t(\cdot \mid Z_t(\omega)), \omega\right) && (38)\\
&\qquad \text{for all } t \in \mathbb{N} \quad \text{sufficiently large,} \\
&\mathsf{excess}\left(\{\hat{\theta}_t(\omega)\}, \Psi\left\langle \mathsf{f}\left(\bullet\,;\mu_0(\omega), \omega\right);\bar{\varepsilon}(\omega)\right\rangle\right) && (39)\\
&\qquad = \mathsf{excess}\left(\{\eta_t(\omega)\}, \Psi\left\langle \mathsf{f}\left(\bullet\,;\mu_0(\omega), \omega\right);\bar{\varepsilon}(\omega)\right\rangle\right) \\
&\qquad \text{for all } t \in \mathbb{N} \quad \text{sufficiently large.}
\end{aligned}$$

P r o o f. The definition of the strict asymptotically ε-minimal solution assumes $\Omega_O \subset \Omega$ with $\mathsf{Prob}\,(\Omega_O) = 1$ such that for every $\omega \in \Omega_O$ and $t \in \mathbb{N}$ sufficiently large the formula (9) is fulfilled. Therefore, for every $\omega \in \Omega_O$

$$\hat{\theta}_t(\omega) = \eta_t(\omega) \quad \forall\, t \in \mathbb{N} \quad \text{sufficiently large.}$$

Since the sequences coincide for every $t \in \mathbb{N}$ sufficiently large the rest of Lemma 3.3 follows directly. □

Considering a weak asymptotically ε-minimal solution we are loosing direct relation. We are not able to compare compactness and cluster points of these sequences. Nevertheles, some relations remain valid in probability sense.

Theorem 3.4. Let $\hat{\theta}_t : \Omega \to \mathcal{X}$ be a weak asymptotically ε-minimal solution.
Then,

$$\lim_{t \to +\infty} \mathsf{Prob}_*\left(\hat{\theta}_t = \eta_t\right) = 1. \tag{40}$$

Consequently,

$$\lim_{t \to +\infty} \mathsf{Prob}_*\left(\mathsf{f}\left(\hat{\theta}_t\,;\mu_t(\cdot \mid Z_t), \cdot\right) = \mathsf{f}\left(\eta_t\,;\mu_t(\cdot \mid Z_t), \cdot\right)\right) = 1, \tag{41}$$

$$\begin{aligned}
&\lim_{t \to +\infty} \mathsf{Prob}_*\Big(\mathsf{excess}\left(\{\hat{\theta}_t\}, \Psi\left\langle \mathsf{f}\left(\bullet\,;\mu_0, \cdot\right);\bar{\varepsilon}\right\rangle\right) && (42)\\
&\qquad = \mathsf{excess}\left(\{\eta_t\}, \Psi\left\langle \mathsf{f}\left(\bullet\,;\mu_0, \cdot\right);\bar{\varepsilon}\right\rangle\right)\Big) = 1.
\end{aligned}$$

Proof. The definition of weak asymptotically ε-minimal solutions required that the formula (9) is fulfilled with inner probability tending to 1. Therefore,

$$\lim_{t\to+\infty} \mathsf{Prob}_* \left(\hat{\theta}_t = \eta_t\right) = 1. \tag{43}$$

The rest of Lemma 3.4 follows this observation. □

Similar construction can be done for level-minimal solutions.

Construction:

Choose arbitrary sequence $\alpha_t : \Omega \to \mathbb{R}$, $\alpha_t > 0$ such that

$$\Delta_t(\omega) + \alpha_t(\omega) > \varphi\left[\mathsf{f}\left(\bullet\,; \mu_t(\cdot \mid Z_t(\omega)), \omega\right)\right] \quad \forall\, t \in \mathbb{N}\ \forall\, \omega \in \Omega.$$

Now, one can select $\xi_t : \Omega \to \mathcal{X}$, $t \in \mathbb{N}$ such that

$$\xi_t(\omega) \in \mathsf{lev}_{\Delta_t(\omega)+\alpha_t(\omega)} \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot \mid Z_t(\omega)), \omega\right)\right\rangle \quad \forall\, \omega \in \Omega,\ \forall\, t \in \mathbb{N}.$$

For a map $\hat{\theta}_t : \Omega \to \mathcal{X}$ we set

$$\begin{aligned} \eta_t(\omega) &= \hat{\theta}_t(\omega) \quad \text{if } \hat{\theta}_t(\omega) \in \mathsf{lev}_{\Delta_t(\omega)} \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot \mid Z_t(\omega)), \omega\right)\right\rangle \\ &= \xi_t(\omega) \quad \text{otherwise.} \end{aligned}$$

The construction produces η_t, $t \in \mathbb{N}$ such that

$$\eta_t(\omega) \in \mathsf{lev}_{\Delta_t(\omega)+\alpha_t(\omega)} \left\langle \mathsf{f}\left(\bullet\,; \mu_t(\cdot \mid Z_t(\omega)), \omega\right)\right\rangle \quad \text{for all } t \in \mathbb{N},\ \omega \in \Omega.$$

Thus, η_t is a random level-minimal solution of course for enlarged level $\Delta_t + \alpha_t$. If, moreover, $\lim_{t\to+\infty} \alpha_t(\omega) = 0$ then

$$\limsup_{t\to+\infty} \Delta_t(\omega) + \alpha_t(\omega) = \limsup_{t\to+\infty} \Delta_t(\omega) = \bar{\Delta}(\omega).$$

Now, we can compare asymptotic properties of $\hat{\theta}_t$, $t \in \mathbb{N}$ and η_t, $t \in \mathbb{N}$.

Theorem 3.5. Let $\hat{\theta}_t : \Omega \to \mathcal{X}$ be a strict asymptotically level-minimal solution.

Then, there is $\Omega_O \subset \Omega$ with $\mathsf{Prob}\,(\Omega_O) = 1$ such that for every $\omega \in \Omega_O$

$$\hat{\theta}_t(\omega) = \eta_t(\omega) \quad \forall\ t \in \mathbb{N} \quad \text{sufficiently large.} \tag{44}$$

Consequently for every $\omega \in \Omega_O$,

$$\text{the sequence } \hat{\theta}_t(\omega),\ t \in \mathbb{N} \text{ is compact if and only if} \tag{45}$$
$$\text{the sequence } \eta_t(\omega),\ t \in \mathbb{N} \text{ is compact,}$$

$$\mathsf{Ls}\left\{\hat{\theta}_t(\omega)\,,\, t \in \mathbb{N}\right\} = \mathsf{Ls}\left\{\eta_t(\omega)\,,\, t \in \mathbb{N}\right\}, \tag{46}$$

$$\mathsf{f}\left(\hat{\theta}_t(\omega)\,; \mu_t(\cdot \mid Z_t(\omega)), \omega\right) = \mathsf{f}\left(\eta_t(\omega)\,; \mu_t(\cdot \mid Z_t(\omega)), \omega\right) \tag{47}$$
$$\text{for all } t \in \mathbb{N} \quad \text{sufficiently large,}$$

$$\mathsf{excess}\left(\{\hat{\theta}_t(\omega)\}, \mathsf{lev}_{\bar{\Delta}(\omega)} \left\langle \mathsf{f}\left(\bullet\,; \mu_0(\omega), \omega\right)\right\rangle\right) \tag{48}$$
$$= \mathsf{excess}\left(\{\eta_t(\omega)\}, \mathsf{lev}_{\bar{\Delta}(\omega)} \left\langle \mathsf{f}\left(\bullet\,; \mu_0(\omega), \omega\right)\right\rangle\right)$$
$$\text{for all } t \in \mathbb{N} \quad \text{sufficiently large.}$$

Theorem 3.6. Let $\hat{\theta}_t : \Omega \to \mathcal{X}$ be a weak asymptotically level-minimal solution.
Then,

$$\lim_{t\to+\infty} \mathsf{Prob}_* \left(\hat{\theta}_t = \eta_t\right) = 1. \tag{49}$$

Consequently,

$$\lim_{t\to+\infty} \mathsf{Prob}_* \left(\mathsf{f}\left(\hat{\theta}_t\,; \mu_t(\cdot \mid Z_t), \cdot\right) = \mathsf{f}\left(\eta_t\,; \mu_t(\cdot \mid Z_t), \cdot\right)\right) = 1, \tag{50}$$

$$\lim_{t\to+\infty} \mathsf{Prob}_* \Big(\mathsf{excess}\left(\{\hat{\theta}_t\}, \mathsf{lev}_{\bar{\Delta}}\left\langle \mathsf{f}\left(\cdot\,; \mu_0, \cdot\right)\right\rangle\right) \tag{51}$$

$$= \mathsf{excess}\left(\{\eta_t\}, \mathsf{lev}_{\bar{\Delta}}\left\langle \mathsf{f}\left(\cdot\,; \mu_0, \cdot\right)\right\rangle\right)\Big) = 1.$$

Theorems above showed to us that asymptotic of $\hat{\theta}_t(\omega),\ t \in \mathbb{N}$ is governed by asymptotic of its counterpart $\eta_t(\omega),\ t \in \mathbb{N}$. More precisely, to receive asymptotic almost surely (resp. almost uniformly) for strict asymptotically ε-minimal solutions or strict asymptotically level-minimal solutions we need to show the same convergence for its counterparts. Some results in this direction can be found in [1, 2, 5]. Let us note that in [5] the author calls convergence almost surely "sample-path optimization". To receive asymptotic in outer probability for weak asymptotically ε-minimal solutions or weak asymptotically level-minimal solutions we need to show the same convergence for its counterparts. Some results in this direction can be found in [7].

4. APPENDIX – NONMEASURABLE MAPPINGS

This auxiliary section contains basic information on nonmeasurable mappings. Definitions and relations are taken from [8], chapter 1.9, pp. 52–56. All proofs can be found in [8], also.

Definition 4.1. Outer and inner probability of a set $B \subset \Omega$ are defined by

$$\mathsf{Prob}^*(B) = \inf\left\{\mathsf{Prob}(A)\ \mid B \subset A,\ A \in \mathcal{A}\right\},$$
$$\mathsf{Prob}_*(B) = \sup\left\{\mathsf{Prob}(A)\ \mid B \supset A,\ A \in \mathcal{A}\right\}.$$

Outer and inner probability are related with a simple formula

$$\mathsf{Prob}_*(B) + \mathsf{Prob}^*(\Omega \setminus B) = 1.$$

Definition 4.2. Let $\mathcal{X}$ be a metric space with metric d and $X_n, X : \Omega \to \mathcal{X}$, $n \in \mathbb{N}$ be arbitrary maps.

- X_n, $n \in \mathbb{N}$ converges almost surely to X if

$$\mathsf{Prob}_*\left(\lim_{n\to+\infty} d(X_n, X) = 0\right) = 1.$$

We use notation $X_n \xrightarrow[n\to+\infty]{\text{as}} X$.

- X_n, $n \in \mathbb{N}$ converges almost uniformly to X if for every $\varepsilon > 0$ there exists a measurable set $A_\varepsilon \subset \Omega$ with $\mathsf{Prob}\,(A_\varepsilon) \geq 1 - \varepsilon$ and $d\,(X_n, X) \longrightarrow 0$ uniformly on A_ε, i.e. $\sup_{\omega \in A_\varepsilon} d\,(X_n(\omega), X(\omega)) \longrightarrow 0$.

  We use notation $X_n \xrightarrow[n\to+\infty]{\mathrm{au}} X$.

- X_n, $n \in \mathbb{N}$ converges in outer probability to X if

$$\text{for every } \varepsilon > 0 \quad \mathsf{Prob}^*\,(d\,(X_n, X) > \varepsilon) \to 0.$$

  We use notation $X_n \xrightarrow[n\to+\infty]{\mathsf{Prob}^*} X$.

There are known relations among these convergences:

- $X_n \xrightarrow[n\to+\infty]{\mathrm{au}} X$ implies $X_n \xrightarrow[n\to+\infty]{\mathrm{as}} X$;

- $X_n \xrightarrow[n\to+\infty]{\mathrm{au}} X$ implies $X_n \xrightarrow[n\to+\infty]{\mathsf{Prob}^*} X$.

The implication $X_n \xrightarrow[n\to+\infty]{\mathrm{as}} X$ implies $X_n \xrightarrow[n\to+\infty]{\mathrm{au}} X$ is not valid in general. For this result we have to add measurability or certain kind of "asymptotic measurability"; see [8] for details.

ACKNOWLEDGEMENT

This work was partially supported by the Czech Science Foundation under Grant 201/08/0486 and by the project MSM 0021620839 financed by Ministry of Education, Youth and Sport of the Czech Republic.

(Received April 14, 2010)

REFERENCES

[1] P. Lachout: Stability of stochastic optimization problem – nonmeasurable case. Kybernetika *44* (2008), 2, 259–276.

[2] P. Lachout: Stochastic optimization sensitivity without measurability.In: Proc. 15th MMEI held in Herľany, Slovakia (K. Cechlárová, M. Halická, V. Borbeľová, and V. Lacko, eds.)2007, pp. 131–136.

[3] P. Lachout, E. Liebscher, and S. Vogel: Strong convergence of estimators as ε_n-minimizers of optimization problems. Ann. Inst. Statist. Math. *57* (2005), 2, 291–313.

[4] P. Lachout and S. Vogel: On continuous convergence and epi-convergence of random functions. Part I: Theory and relations. Kybernetika *39* (2003), 1, 75–98.

[5] S. M. Robinson: Analysis of sample-path optimization. Math. Oper. Res. *21* (1996), 3, 513–528.

[6] T. Rockafellar and R. J.-B. Wets: Variational Analysis. Springer-Verlag, Berlin 1998.

[7] I. Vajda and M. Janžura: On asymptotically optimal estimates for general observations. Stoch. Process. Appl. *72* (1997), 1, 27–45.

[8] A. W. van der Vaart and J. A. Wellner: Weak Convergence and Empirical Processes. Springer, New York 1996.

Petr Lachout, Department of Probability and Statistics, Charles University of Prague, Sokolovská 83, 186 75 Praha 8 and Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.

e-mail: lachout@karlin.mff.cuni.cz