Working Paper

STOCHASTIC OPTIMIZATION TECHNIQUES FOR FINDING OPTIMAL SUBMEASURES

Alexei Gaivoronski

May 1985
WP-85-28

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

STOCHASTIC OPTIMIZATION TECHNIQUES FOR FINDING OPTIMAL SUBMEASURES

Alexei Gaivoronski

May 1985
WP-85-28

Working Papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

FOREWORD

In this paper, the author look at some quite general optimization problems on the space of probabilistic measures. These problems originated in mathematical statistics but have applications in several other areas of mathematical analysis. The author extend previous work by considering a more general form of the constraints, and develop numerical methods (based on stochastic quasigradient techniques) and some duality relations for problems of this type.

This paper is a contribution to research on stochastic optimization currently underway within the Adaptation and Optimization Project.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences
Program

CONTENTS

1.	INTRODUCTION	1
2.	CHARACTERIZATION OF THE OPTIMAL SOLUTIONS	3
3.	STOCHASTIC OPTIMIZATION METHOD	14
4.	PARAMETRIC DEPENDENCE OF $g(y, H)$ ON MEASURE	41
5.	NUMERICAL EXPERIMENT	46

STOCHASTIC OPTIMIZATION TECHNIQUES FOR FINDING OPTIMAL SUBMEASURES

Alexei Gaivoronski

1. INTRODUCTION

Optimality conditions based on duality relations were studied in [1] for the following optimization problem.

Find the positive Borel measure H such that

$$\Psi^0(H) = \max \tag{1}$$

with respect to constraints

$$\Psi^i(H) \leq 0 \qquad i = 1:m \tag{2}$$

$$H^l(A) \leq H(A) \leq H^u(A) \tag{3}$$

for all Borel $A \subset Y \subset R^n$

$$H(Y) = 1 \tag{4}$$

where $Y-$ some subset of Euclidean space R^n, $\Psi^i(H)-$ function which depends on the measure H, usually some kind of directional differentiability and convexity is assumed. H^u and H^l are some positive Borel measures. Stochastic optimization methods for solving (1)-(4) in case when functions $\Psi^i(H)$ are linear with respect to H were developed in [1]. In this paper such methods are developed for nonlinear functions $\Psi^i(H)$ and for arbitrary finite measures. Interest for such a problem is originated from statistics where it appears in finite population sampling [2,3].

Suppose that we have collection D of N objects, each object is described by pair $(x_i, y_i), i = 1:N$. Variables y_i are known and variables x_i can be observed for each particular i in the following way:

$$z_i = x_i + \omega_i$$

where ω_i - random independent variables with zero mean, and z_i - observations. It is assumed usually that relationship between x_i and y_i is known up to the set of unknown parameters:

$$x_i = h^T(y_i)\vartheta$$

where $h(y) = (h_1(y), \ldots, h_l(y))$ are known functions and $\vartheta = (\vartheta_1, \ldots, \vartheta_l)$ are parameters to be determined.

The problem is to select subset $d \subset D$ consisting of n objects in order to get in some sense the best possible estimate of parameters ϑ. This estimate is based on observations z_i for objects belonging to d.

Applying the usual approach of optimal experimental design [4-6] one can substitute the collection $(y_1, \ldots, y_n)$ by measure $H^u(y)$ and subset of objects to be observed by measure $H(y)$. The variance matrix D of the best linear estimate in case all ω_i has the same variance becomes after such substitution proportional to matrix M, defined as follows:

$$M^{-1} = \int h(y) h^T(y) \mathrm{d}H(y)$$

and the problem becomes to minimize some function ψ of M, such as determinant, trace, the largest eigenvalue, etc.

$$\min_{H} \psi(M) \tag{5}$$

with respect to obvious constraint

$$\int_A \mathrm{d}H(y) \leq \int_A \mathrm{d}H^u(y) \tag{6}$$

for all Borel $A \subset Y$. Another possible application of the problem (1)-(4) are approximation schemes for stochastic optimization [7,8].

The purpose of this paper is to develop stochastic optimization methods dealing with such problems. In section 2 the characterization of solutions for quite general classes of measures is obtained. The conseptual algorithm for solving nonlinear problems is proposed in the section 3, which is applied in section 4 to particular problems of the kind (5)-(6). In section 5 results of some numerical experiments are presented.

2. CHARACTERIZATION OF THE OPTIMAL SOLUTIONS

We shall consider subset Y of Euclidean space R^n and some σ-field Ξ on it. We shall assume that all measures specified below are defined on this σ-field.

In this section, the representation of measures H, which are the solution of the following problem, will be developed:

$$\max \ \Psi(H) \tag{7}$$

subject to constraint

$$H^l \leq H \leq H^u \tag{8}$$

$$H(Y) = b \tag{9}$$

The constraint (8) means that $H^l(E) \leq H(E) \leq H^u(E)$ for any $E \subset \Xi$. Define $H^\Delta = H^u - H^l$. In what follows the spaces $L_1(Y,\Xi,H^\Delta)$ and $L_\infty(Y,\Xi,H^\Delta)$ play an important role, where $L_1(Y,\Xi,H^\Delta)$ is the space of all H^Δ-measurable functions $g(y)$ defined on Y and such that $\int_Y |g(y)| dH^\Delta < \infty$, $L_\infty(Y,\Xi,H^\Delta)$ is the space of all H^Δ-measurable and H^Δ-essentially bounded functions $g(y)$, defined on Y. In what follows we shall denote by $||\cdot||_\infty$ the norm in the space $L_\infty(Y,\Xi,H^\Delta)$, i.e.

$$||g(y)|| = H^\Delta - \text{ess}\sup_{y \in Y} |g(y)|$$

Let us denote by G the set of all measures, satisfying (8):

$$G = \{H: \quad H^u \leq H \leq H^l\}$$

and by G_b the set of all measures, satisfying in addition (9):

$$G_b = \{H: H \in G,\ H(Y) = b\}$$

Suppose that $f(y)$ is some function defined on Y, c-some number and define the following sets

$$Z^+(c, f) = \{y: y \in Y, f(y) > c\}$$
$$Z^-(c, f) = \{y: y \in Y, f(y) < c\}$$
$$Z^0(c, f) = \{y: y \in Y, f(y) = c\}$$

In notations below we shall substitute in this definition instead of f various particular functions. Take

$$c^* = \inf\{c : H^\Delta(Z^+(c, g)) \leq b - H^l(Y)\}$$

and define, as usual, by H^-, H^+ and $|H|$ positive, negative and total variation of the measure H.

We shall first consider the problem in which function $\Psi(H)$ is linear:

$$\max_{H \in G_b} \int g(y)dH \qquad (10)$$

and describe the set of all solutions of (10). The following result is generalization of Lemma 1 from [1].

THEOREM 1. Suppose that the following conditions are satisfied:

1. $H^l(Y) \leq b$, $|H^l|(Y) < \infty$, $H^u(Y) \geq b$

2. For any $E \in \Xi, H^\Delta(E) > 0$ exists $E_1 \in \Xi$, $E_1 \subset E$, such that either E_1 is H^Δ-atom or $0 < H^\Delta(E_1) < \infty$

3. $g(y) \in L_1(Y, \Xi, H^\Delta)$, $\int_Y |g(y)| d|H^l| < \infty$

4. If $c^* = 0$ then $H^\Delta(Y \setminus Z^-(0,g)) \geq b - H^l(Y)$

Then the solution of problem (10) exists and any such solution has the following representation:

(i) $H^*(A) = H^u(A)$ for any $A \in \Xi$, $A \subset Z^+(c^*,g)$

(ii) $H^*(A) = H^l(A)$ for any $A \in \Xi$, $A \subset Z^-(c^*,g)$

(iii) $H^u(A) \geq H^*(A) \geq H^l(A)$ for any $A \in \Xi$, $A \subset Z^0(c^*,g)$ and $H^*(Z^0(c^*,g)) = b - H^l(Y) - H^\Delta(Z^+(c^*,g))$

Conversely, any measure defined by (i)-(iii) is the solution of the problem (10).

PROOF. Let us first prove that the measure with properties (i)-(iii) exists. It is clear that any measure on (Y,Ξ) is defined by its values on subsets of $Z^+(c^*,g)$, $Z^0(c^*,g)$ and $Z^-(c^*,g)$ because these sets belong to Ξ due to $g(y) \in L_1(Y,\Xi,H^\Delta)$ and Y equals to union of these sets. Therefore it is sufficient to show that among measures satisfying (i)-(ii) exist measure which satisfies also (iii).

From the definition of $Z^+(c,g)$ we have:

$$Z^+(c^*,g) = \bigcup_{c>c^*} Z^+(c,g)$$

and

$$Z^+(c_1,g) \subset Z^+(c_2,g)$$

for all $c_1 > c_2$. This gives

$$\lim_{c \downarrow c^*} H^\Delta(Z^+(c,g)) = H^\Delta(Z^+(c^*,g))$$

and therefore

$$H^\Delta(Z^+(c^*,g)) \le b - H^l(Y)$$

According to the condition 4 we have $H^\Delta(Y \setminus Z^-(c^*,g)) \ge b - H^l(Y)$ in case if $c^* = 0$. In fact, it is true for arbitrary c^*. Suppose at first that $c^* > 0$. Note that for any $c > 0$ we have $H^\Delta(Z^+(c,g)) < \infty$ because $g(y) \in L_1(Y,\Xi,H^\Delta)$. Consider now the sequence c_s:

$$0 < c_s < c^*, \; c_{s+1} \ge c_s, \; c_s \to c^*.$$

We have

$$H^\Delta(Y \setminus Z^-(c^*,g)) = \lim_{s \to \infty} H^\Delta(Z^+(c_s,g))$$

because

$$H^\Delta(Z^+(c_s,g)) < \infty, \; Z^+(c_{s+1},g) \subset Z^+(c_s,g)$$

and

$$\bigcap_s Z^+(c_s,g) = Y \setminus Z^-(c^*,g)$$

From the definition of c^* and the fact that $c_s < c^*$ we have:

$$H^\Delta(Z^+(c_s,g)) > b - H^l(Y)$$

which gives

$$H^\Delta(Y \setminus Z^-(c^*,g)) \geq b - H^l(Y)$$

The case when $c^* < 0$ is treated in the same way taking into account the fact that $H^\Delta(Z^+(c,g)) < \infty$ for all c if $c^* < 0$. Thus, we obtain

$$H^\Delta(Z^+(c^*,g)) \leq b - H^l(Y),$$
$$H^\Delta(Z^0(c^*,g)) \geq b - H^l(Y) - H^\Delta(Z^+(c^*,g)) \geq 0$$

Now if $H^\Delta(Z^0(c^*,g)) < \infty$ the measure H^*:

$$H^*(E) = \begin{cases} H^u(E) & \text{if } E \subset Z^+(c^*,g) \\ H^l(E) & \text{if } E \subset Z^-(c^*,g) \\ & \text{or } E \subset Z^0(c^*,g) \text{ and } H^\Delta(Z^0(c^*,g)) = 0 \\ H^l(E) + \dfrac{b - H^l(Y) - H^\Delta(Z^+(c^*,g))}{H^\Delta(Z^0(c^*,g))} H^\Delta(E) & \\ & \text{otherwise} \end{cases} \tag{11}$$

satisfies all conditions (i)-(iii).

When $H^\Delta(Z^0(c^*,g)) = \infty$, condition 2 implies the existence $E_1 \subset Z^0(c^*,g)$ such that either

$$b - H^l(Y) - H^\Delta(Z^+(c^*,g)) \leq H^\Delta(E_1) < \infty$$

or $H^\Delta(E_1) = \infty$ and E_1 is H^Δ-atom. In the former case H^* is defined similarly to (11) using set E_1 instead of $Z^0(c^*,g)$ and taking $H^* = H^l$ on $Z^0(c^*,g) \setminus E_1$. In the latter case take

$$H^*(E_1) = b - H^l(Y) - H^\Delta(Z^+(c^*,g))$$

and again $H^* = H^l$ on $Z^0(c^*,g)\setminus E_1$. All this proves the existence of measure H^* which satisfies (i)-(iii). Let us prove that for any measure $\bar{H} \in G_b$ which violate (i)-(iii), we have

$$\int_Y g(y)d\bar{H} < \int_Y g(y)dH^*$$

Suppose that for measure $\bar{H}$ (i) does not hold, i.e., there is some set $E \subset Z^+(c^*,g)$ such that $\bar{H}(E) < H^u(E)$. This implies the existence of set $E_1 \subseteq E$ such that $g(y) > \bar{c} > c^*$ for $y \in E_1$ and $H^*(E_1) - \bar{H}(E_1) > \gamma > 0$. Notice that $H^*(A) \geq \bar{H}(A)$ for $A \subset Z^+(c^*,g)$ and $H^*(A) \leq \bar{H}(A)$ for $A \subset Z^-(c^*,g)$ due to definition of H^*. This gives

$$\int_{E_1} g(y)d(H^* - \bar{H}) \geq \bar{c}(H^* - \bar{H})(E_1)$$

$$\int_{Z^+(c^*,g)\setminus E_1} g(y)d(H^* - \bar{H}) \geq c^*(H^* - \bar{H})(Z^+(c^*,g)\setminus E_1)$$

$$\int_{Z^0(c^*,g)} g(y)d(H^* - \bar{H}) = c^*(H^* - \bar{H})(Z^0(c^*,g))$$

$$\int_{Z^-(c^*,g)} g(y)d(H^* - \bar{H}) \geq c^*(H^* - \bar{H})(Z^-(c^*,g))$$

This inequalities lead to the following estimate:

$$\int_Y g(y)dH^* - \int_Y g(y)d\bar{H} = \int_{E_1} g(y)d(H^* - \bar{H})$$

$$+ \int_{Z^+(c^*,g)\setminus E_1} g(y)d(H^* - \bar{H}) + \int_{Z^0(c^*,g)} g(y)d(H^* - \bar{H})$$

$$+ \int_{Z^-(c^*,g)} g(y)d(H^* - \bar{H}) \geq \bar{c}(H^* - \bar{H})(E_1)$$

$$+ c^*(H^* - \bar{H})(Y\setminus E_1)$$

$$= (\bar{c} - c^*)\gamma > 0$$

Thus, if $\bar{H} \in G_b$ violates (i) it cannot be the solution of (10). Other possibilities are considered in the same way. Therefore, any optimal measure has representation (i)-(iii). It follows from definition that all measures satisfying (i)-(iii) has the same value of

$$\int_Y g(y)dH$$

and therefore are optimal. Proof is complete.

REMARK. It is clear that in the characterization of optimal measures any $\tilde{c}$ can be taken instead of c^* such that

$$c^* \le \tilde{c} \le \sup \{c : H^\Delta(Z^+(c,g)) \ge b - H^l(Y)\}$$

Note that if measure H^Δ has bounded variation conditions 2 and 4 are satisfied automatically. For such measures the structure of solutions can be studied using general duality theory [9].

Let us now consider in more detail the set G_b. If the measure H^l has finite variation we have the following representation for arbitrary $H \in G_b$:

$$H = H^l + (H - H^l)$$

where measure $H - H^l$ is finite, positive and continuous with respect to measure H^Δ. If H^Δ is σ-finite we can use Radon-Nycodym theorem [10] and for arbitrary $H \in G_b$ obtain the following representation:

$$H(E) = H^l(E) + \int_E h_H(y)dH^\Delta \quad \forall E \in \Xi \tag{12}$$

where $h_H \in L_1(Y, \Xi, H^\Delta)$ and this representation is unique. For arbitrary $E \in \Xi$ we have:

$$0 \le \int_E h_H(y)dH^\Delta \le H^\Delta(E)$$

and therefore $0 \leq h_H(y) \leq 1$ H^Δ-everywhere. Consider now the set $K_b \subset L_1(Y, \Xi, H^\Delta)$:

$$K_b = \{h: \ 0 \leq h(y) \leq 1, \int_Y h(y) dH^\Delta = b - H^l(Y)\} \tag{13}$$

Each function from this set defines measure H_h from G_b:

$$H_h(E) = H^l(E) + \int_E h(y) dH^\Delta, \ E \in \Xi \tag{14}$$

Therefore (12), (14) defines isomorphism between sets G_b and K_b such that the problem (7)-(9) is equivalent to the following one:

$$\max \overline{\Psi}(h) \tag{15}$$

subject to constraints

$$0 \leq h(y) \leq 1 \tag{16}$$

$$\int_Y h(y) dH^\Delta = b - H^l(Y) \tag{17}$$

where the function $\overline{\Psi}(h) = \Psi(H_h)$. Optimal values of problems (15)-(17) and (7)-(9) are the same and each solution of (15)-(17) defines solution of (7)-(9) through (14) and vice versa. This equivalence together with certain convexity assumptions lead to solution representation for problems (7)-(9) similar to theorem 1:

THEOREM 2. Suppose that the following assumptions are satisfied:

1. Measures H^l and H^u have bounded variation,

$$H^l(Y) \leq b, \ \ H^u(Y) \geq b$$

2. $\Psi(H)$ is concave and finite for $H \in G_{b,\varepsilon} = G_b + G_\varepsilon$

where $G_\varepsilon = \{H_\varepsilon: |H_\varepsilon|(Y) \leq \varepsilon, \ H_\varepsilon$ is H^Δ-continuous $\}$ for some $\varepsilon > 0$. Then

1) For each $H_1 \in G_b$ exists $g(y, H_1) \in L_\infty(Y, \Xi, H^\Delta)$ such that

$$\Psi(H_2) - \Psi(H_1) \leq \int_Y g(y, H_1) d(H_2 - H_1) \tag{18}$$

for all $H_2 \in G_b$

2) The solution H^* of problem (7)-(9) exists.

3) For any $E \in \Xi$ and any optimal solution H^* of the problem (7)-(9) we have the following representation:

$$H^*(E) = \begin{cases} H^u(E) & \text{for} \quad E \subset Z^+(c^*, g(y, H^*)) \\ H^l(E) & \text{for} \quad E \subset Z^-(c^*, g(y, H^*)) \\ H^l(E) \leq H(E) \leq H^u(E) & \text{for} \quad E \subset Z^0(c^*, g(y, H^*)) \end{cases} \tag{19}$$

where

$$c^* = \inf\{c: \ H^\Delta(Z^+(c, g(y, H^*)) \leq b - H^l(Y)\}$$

and

$$g(y, H^*) \in L_\infty(Y, \Xi, H^\Delta), \quad \Psi(H) - \Psi(H^*) \leq \int_Y g(y, H^*) d(H - H^*)$$

for all $H \in G_b$. Conversely, if for some $H_1 \in G_b$ exists $g(y, H_1) \in L_\infty(Y, \Xi, H^\Delta)$ such that (18) is fulfilled and H_1 can be represented according to (19) then H_1 is the optimal solution of the problem (7)-(9).

PROOF. The previous argument shows that under assumptions of the theorem problem (7)-(9) is equivalent to the problem (15)-(17) and there is isomorphism between set $G_{\delta,\varepsilon}$ as defined in condition 2, and the following set $K_{\delta,\varepsilon} \in L_1(Y,\Xi,H^\Delta)$:

$$K_{\delta,\varepsilon} = K_\delta + K_\varepsilon,$$
$$K_\varepsilon = \{h:\ h \in L_1(Y,\Xi,H^\Delta), \int_Y |h(y)| dH^\Delta \le \varepsilon\}$$

Function $\bar{\Psi}(h)$ from (15) is concave on the set $K_{\delta,\varepsilon}$, which is ε-vicinity of K_δ in $L_1(Y,\Xi,H^\Delta)$. Therefore for each $L \in K_\delta$ exists subdifferential of concave function $\bar{\Psi}(h)$ [11, 12], which in this case is linear continuous functional $\bar{g} \in L_1^*(Y,\Xi,H^\Delta)$ such that

$$\bar{\Psi}(h_1) - \bar{\Psi}(h) \le \bar{g}(h_1 - h)$$

Taking into account representation of $L_1^*(Y,\Xi,H^\Delta)$ [10] we get:

$$\bar{\Psi}(h_1) - \bar{\Psi}(h) \le \int_Y \bar{g}(y,h)(h_1(y) - h(y))dH^\Delta \tag{20}$$

where

$$\bar{g}(y,h) \in L_\infty(Y,\Xi,H^\Delta)$$

which together with (12) implies

$$\Psi(H_1) - \Psi(H) \le \int_Y g(y,H)d(H_1 - H)$$

for all $H, H_1 \in G_\delta$ where $g(y,H) = \bar{g}(y,h_H)$. Thus, (18) is proved. Note that we may consider function $g(y,H_1)$ from (18) (possibly non-unique) as subdifferential of the function $\Psi(H)$ at point H_1.

Now observe that the set K_δ is weakly sequentially compact in $L_1(Y,\Xi,H^\Delta)$ because $H^\Delta(Y) < \infty$ and

$$\lim_{H^\Delta(E)\to 0} \int_E h(y)dH^\Delta = 0$$

uniformly for $h \in K_b$ (see [10, p.294]). Let us prove that it is also weakly closed. Consider the sequence $h^s(y)$, $h^s \in K_b$ and

$$\int_Y g(y)h^s(y)dH^\Delta \to \int_Y g(y)h(y)dH^\Delta$$

for some $h \in L_1(Y,\Xi,H^\Delta)$ and all $g \in L_\infty(Y,\Xi,H^\Delta)$. In particular, we have

$$\int_E h^s(y)dH^\Delta \to \int_E h(y)dH^\Delta$$

for all $E \in \Xi$ because the indicator function of the set $E \in \Xi$ clearly belongs to $L_\infty(Y,\Xi,H^\Delta)$. This gives $0 \le h(y) \le 1$ H^Δ-everywhere. Taking $g(y) \equiv 1$ we have also

$$\int_Y h^s(y)dH^\Delta \to \int_Y h(y)dH^\Delta$$

which gives

$$\int_Y h(y)dH^\Delta = b - H^l(Y)$$

Thus, $h \in K_b$ and K_b is weakly closed.

It follows from (20) that for any sequence $h^s \in K_b$, $h^s \to h$ weakly, $h \in K_b$ we have

$$\overline{\lim_{s\to\infty}}\ \overline{\Psi}(h^s) \le \overline{\Psi}(h)$$

This together with sequential compactness and closeness of K_b implies existence of h^* such that

$$\overline{\Psi}(h^*) = \sup_{h\in K_b} \overline{\Psi}(h)$$

Thus, solution of the problem (7)-(9) exists.

The general results of convex analysis [11] now imply that under assumption 2 of the theorem for any solution H^* of the problem (7)-(9) exists subdifferential $g(y,H^*)$ of the function $\Psi(H)$ at point H^* such that

$$\int_Y g(y,H^*)d(H-H^*) \le 0, \ \Psi(H)-\Psi(H^*) \le \int g(y,H^*)d(H-H^*)$$

for all $H \in G_b$ or, in other words H^* is one of the solutions of the following problem:

$$\max_{H \in G_b} \int_Y g(y,H^*)dH \tag{21}$$

This problem is exactly of the type (10) and its solutions are characterized by the Theorem 1. Conversely, if for some $H^* \in G_b$ exists subdifferential $g(y,H^*)$ such that H^* is the solution of the problem (21) then H^* is the optimal solution of the original problem. Proof is now completed by using theorem 1. Some related results were obtained for a special kind of function $\Psi(H)$, atomless probability measure H^u and $H^l \equiv 0$ in [2].

Theorem 2 shows that solutions of the problem (7)-(9) can be viewed as indicator functions of some sets. Therefore many problems involving selection of optimal set [13] can be reformulated as problems of finding optimal measures.

3. STOCHASTIC OPTIMIZATION METHOD

Using the results of the previous section we can construct numerical methods for solving problem (7)-(9). From now on we shall assume that function $\Psi(H)$ is concave and finite on some vicinity of the set G and possess certain differentiability properties:

$$\Psi(H_1+\alpha(H_2-H_1)) = \Psi(H_1)+\alpha\int_Y g(y,H_1)d(H_2-H_1)+o(\alpha) \tag{22}$$

where $o(\alpha)/\alpha \rightarrow 0$ as $\alpha \rightarrow 0$ for all $H_1, H_2 \in G$. This means that subdifferential $g(y, H_1)$ from (18) is unique for all interior points of G and we can assume that $g(y, H^*)$ from (19) satisfies also (22).

Consider now the mapping $\Gamma(c, f)$ from $R \times L_\infty(Y, \Xi, H^\Delta)$ to G: if $H = \Gamma(C, f)$ then

$$H(E) = \begin{cases} H^u(E) & \text{for } E \subset Z^+(c, f) \\ H^l(E) & \text{for } E \subset Y \setminus Z^+(c, f) \end{cases} \tag{23}$$

for any $E \in \Xi$.

First of all we shall give an informal description of the algorithm. Suppose that some $H^s \in G$ is the current approximation to the solution of the problem (7)-(9). According to (22) local behavior of $\Psi(H)$ around H^s is approximated by linear form:

$$\Psi(H^s + \alpha(H - H^s)) = \Psi(H^s) + \alpha \int g(y, H^s) d(H - H^s) + o(\alpha)$$

and if H_*^s is the solution of the problem

$$\max_{H \in G_b} \int_Y g(y, H^s) dH \tag{24}$$

then direction $H_*^s - H^s$ will be the ascent direction at point H^s. Therefore we can take as the next approximation to the optimal solution

$$H^{s+1} = H^s + \alpha(H_*^s - H^s) \tag{25}$$

for some $\alpha > 0$. Consider now the problem of finding H_*^s or suitable approximation to it.

Suppose that we know the function $g(y, H^s)$ exactly. Then, according to theorem 1, all the possible H_*^s are fully described by pair $(c^*, g(y, H^s))$, where c^* is the solution of the problem

$$\inf_c c \tag{26}$$

$$H^u(Z^+(c,g(y,H^s))) + H^l(Y \setminus Z^+(c,g(y,H^s))) \le b \tag{27}$$

Observe now that function

$$W_c^s(c) = H^u(Z^+(c,g(y,H^s))) + H^l(Y \setminus Z^+(c,g(y,H^s))) - b \tag{28}$$

is nonincreasing and therefore solving (25)-(26) is equivalent to solving

$$\max_c W^s(c),\ W^s(c) = \int_T^c W_c^s(t)dt$$

for some T and $W_c^s(c)$ can be considered as subgradient of the function $W^s(c)$. Therefore we can use subgradient method for finding c^*:

$$c^{k+1} = c^k + \rho_k W_c^s(c^k) \tag{29}$$

However, computation of $W_c^s(c^k)$ according to (28) involves multidimensional integration over complex regions and this may be too complicated from the computational point of view. In this situation stochastic quasigradient methods [14] can be used. In such methods the statistical estimate ξ^k of W_c^k is implemented in (29) instead of W_c^k.

Once c^* is determined the measure $\Gamma(c^*,g(y,H^s))$ defined in (23), may be a reasonable approximation to the solution H_*^s of the problem (24) and can be used in algorithm (25). However, precise estimation of c^* from (29) requires infinite number of iterations and to make algorithm implementable, it is necessary to avoid this. It appears that under certain assumptions about stepsizes in (25) and (29), we may take in (29) $k = s$ and perform only one iteration in (29) per iteration in (25) using as approximation to H_*^s the measure $\bar{H}^s = \Gamma(c^s,g(y,H^s))$. Thus, along with sequence H^s we obtain also the sequence of numbers c^s. Note now that although $\bar{H}^s$ is quite simple, measure H^s would be excessively complex even for small s. However, H^s is only needed for getting gradient $g(y, H^s)$ and in particular cases some approximation $f(s, y)$ to $g(y, H^s)$ can be obtained using only $\bar{H}^s$ in the sort of updating formula similar to (25).

Once sequence $f(s, y)$ with property $|f(s,y) - g(y, H^s)| \to 0$ is obtained together with sequence c^s: $V^s(c^s) - \max_c V^s(c) \to 0$, the optimal solution of problem (7)-(9) is defined by Theorem 2 through accumulation points of these sequences. The structure of optimal solution is close to (23).

Now we shall define the algorithm for solving (7)-(9) formally.

1. At the beginning select initial approximation to solution H^0, function $f(0, y)$ and number c^0.

2. Suppose that at the step number s we get measure H^s, function $f(s, y)$ and number c^s. Then on the next step we do the following:

   2a. Pair $(c^s, f(s, y))$ defines measure $\bar{H}^s$ according to (23):

$$\bar{H}^s = \Gamma(c^s, f(s, y))$$

   New approximation to solution is obtained in the following way:

$$H^{s+1} = (1 - \alpha_s) H^s + \alpha_s \bar{H}^s \tag{30}$$

   2b. Now number c^{s+1} is obtained:

$$c^{s+1} = c^s + \rho_s \xi^s \tag{31}$$

   where

$$E(\xi^s / c^0, \cdots, c^s) = V_c^s(c^s)$$

$$V_c^s = H^u(Z^+(c, f(s, f(s, y))) + H^l(Y \setminus Z^+(c, f(s, y))) - b$$

$$V^s(c) = \int_{T^c} V_c^s(t) dt$$

i.e., the function $V^s(c)$ is defined similarly to $W^s(c)$ with the difference that $f(s, y)$ is used instead of $g(y, H^s)$.

2c. New function $f(s+1, y)$ is obtained in such a way as to approximate $g(y, H^{s+1})$. The precise way of achieving this can be specified only after considering particular ways of dependence $g(y.H)$ on H. One quite general case is considered in the next section. Here we shall only assume that

$$\| f(s, y) - g(y, H^s) \|_\infty \to 0 \tag{32}$$

as $s \to \infty$. The method of achieving this in particular situation will be described in the next section.

Before stating convergence results for algorithm (30)-(31), two examples of calculating ξ^s from (31) are presented.

(i) Measures H^u and H^l have piecewise-continuous densities $H_y^u(y)$ and $H_y^l(y)$ respectively with respect to Lebesque measure. Then we have

$$V_c^s(c) = \int_Y v^s(c, y) dy$$

where

$$v^s(c, y) = \begin{cases} H_y^u(y) - \dfrac{b}{\mu(Y)} & \text{if } y \in Z^+(c, f(s, y)) \\ H_y^l(y) - \dfrac{b}{\mu(Y)} & \text{otherwise} \end{cases}$$

and $\mu(Y)$ is Lebesque measure of Y. Therefore we can take

$$\xi^s = \mu(Y)\, v^s(c, \omega^s)$$

where ω^s is distributed uniformly over Y.

(ii) Measures H^u and H^l are defined by finite number of pairs

$$H^u = \{(y_i, p_i^u),\ i=1,\ldots,N,\ \sum_{i=1}^{N} p_i^u \geq b\}$$
$$H^l = \{(y_i, p_i^l),\ i=1,\ldots,N,\ \sum_{i=1}^{N} p_i^l \leq b\}$$

In this case

$$V_c^s(c) = \sum_{i=1}^{N} (\gamma_i^s(c) p_i^u + (1-\gamma_i^s(c)) p_i^l - b/N)$$

where

$$\gamma_i^s(c) = \begin{cases} 1 & \text{if } f(s, y_i) > c \\ 0 & \text{otherwise} \end{cases}$$

Therefore

$$\xi^S = N\gamma_{\omega^s}^s(c^s) p_{\omega^s}^u + N(1-\gamma_{\omega^s}^s(c^s)) p_{\omega^s}^l - b$$

where ω^s assumes value i, $1 \leq i \leq N$ with probability $1/N$.

Let us now investigate convergence of algorithm (30)-(31). In all statements concerning convergence of measures from the set G we shall use the weak-L_1 convergence, used already in the proof of Theorem 2:

$$H^k \to H \text{ iff } \int_Y g(y) dH^k \to \int_Y g(y) dH$$

for all $g \in L_\infty(Y, \Xi, H^\Delta)$, and topology, induced by this convergence will be used without further reference.

We shall assume that random variables $\xi^1, \cdots, \xi^s, \ldots$ are defined on some probability space, therefore $c^s, H^s, \bar{H}^s$ from (30)-(31) depend on event

ω of this space. For simplicity of notations this dependence will be omitted in formulas. Convergence, boundedness, etc. will be considered almost everywhere with respect to this probability space. It should be stressed that we are primarily interested in convergence properties of the sequences c^s and $f(s, y)$. The following theorem gives results in this direction.

THEOREM 3. Suppose that the following assumptions are satisfied:

1. Measures H^l and H^u have bounded variation,

   $H^l(Y) \le b, \ H^u(Y) \ge b$

2. $\Psi(H)$ is finite concave function for $H \in G + G_\varepsilon$ where

   $G_\varepsilon = \{H_\varepsilon: |H_\varepsilon|(Y) \le \varepsilon, \ H_\varepsilon \text{ is } H^\Delta\text{-continuous}\}$

   for some $\varepsilon > 0$, and satisfies (22) for $H_1, H_2 \in G$.

3. $||g(y, H^k) - g(y, H)||_\infty \to 0$ if $H^k \to H$;

   $||g(y, H^s) - g(y, H^{s+1})||_\infty \le \delta_s \to 0$ as $s \to \infty$.

4. $f(s, y) \in L_\infty(Y, \Xi, H^\Delta)$,

   $||g(y, H^s) - f(s, y)||_\infty \le \Delta_s \to 0$ as $s \to \infty$

   $\sup_s ||f(s, y)||_\infty = \bar{f} < \infty$

5. $\rho_s \to 0, \ \sum_{s=0}^{\infty} \alpha_s = \infty, \ \sum_{s=0}^{\infty} \rho_s^2 < \infty,$

   $\alpha_s / \rho_s \to 0, \ \delta_s / \rho_s \to 0$

   $E((\xi^s - V_c^s(c^s))^2 / c^0, \cdots, c^s) \le M_1$

6. One of the following conditions is satisfied:

(i) $\frac{\alpha_{s+1}}{\rho_{s+1}} \le \frac{\alpha_s}{\rho_s}$

(ii) $\alpha_s > 0$ and $\frac{1}{\alpha_s}\left[\frac{\alpha_{s-1}}{\rho_{s-1}} - \frac{\alpha_s}{\rho_s}\right] \to 0$

(iii) $\sum_{s=0}^{\infty} \left| \frac{\alpha_s}{\rho_s} - \frac{\alpha_{s+1}}{\rho_{s+1}} \right| < \infty$

Then

1) $\Psi(H^s) \to \max_{H\in G} \Psi(H)$, $H^s(Y) \to b$ and all accumulation points of the sequence H^s belong to the set

$$\Phi = \{H:\ H\in G, \Psi(H) = \max_{H\in G} \Psi(H)\}$$

2) For any convergent subsequence $c^{s_k} \to c^*$ exists measure $H^* \in \Phi$ such that

$$H^*(A) = \begin{cases} H^+(A) & \text{for } A \subset Z^+(c^*, g(y, H^*)) \\ H^-(A) & \text{for } A \subset Z^-(c^*, g(y, H^*)) \\ H^-(A) \le H(A) \le H^+(A) & \text{for } A \subset Z^0(c^*, g(y, H^*)) \end{cases}$$

and

$$\| f(s_l, y) - g(y, H^*) \|_{\infty} \to 0$$

where s_l is some subsequence of the sequence s_k.

Condition 4 of the theorem means that it is possible to use approximations to gradient $g(y, H)$ and it is necessary that precision of these approximations increase as $s \to \infty$. Condition 6 is necessary to assure $H^s(Y) \to b$ although $\bar{H}^s(Y)$ from (30) may not be equal to b. In case if $\bar{H}^s(Y) = b$, i.e. $H^{\Delta}(Z^0(c^s, f(s, y))) = 0$ starting from some s, condition 6 is not necessary.

In order to prove this theorem we need some auxiliary results. Consider the following two iterative processes:

$$a^{s+1} = (1-\alpha_s)a^s + \alpha_s V_s \tag{33}$$

$$b^{s+1} = b^s + \rho_s \mu_s \tag{34}$$

$$E(\mu_s / b^0,\dots,b^s) = V_s,\ s=0,1,\dots$$

LEMMA 1. Suppose that

1. $|b^s| \le \bar{b}$ and $\bar{b} < \infty$ a.s., $|V_s| < \bar{V} < \infty$.

2. $\sum_{s=0}^{\infty} \alpha_s = \infty,\ \sum_{s=0}^{\infty} \rho_s^2 < \infty,\ \alpha_s / \rho_s \to 0$

   $E((\mu_s - V_s)^2 / b^0, \cdots, b^s) < M_1 < \infty$ a.s.

3. Condition 6 of the theorem 3 is satisfied.

Then

1) $a^s \to 0$ a.s.

2) $\sum_{t=s}^{k} \alpha_t b^t V_t = \sum_{t=s}^{k} \alpha_t \beta_t + \varepsilon(s)$

   where $\beta_s \to 0, \varepsilon(s) \to 0$ as $s \to \infty$ a.s.

PROOF.

1. Combining (33) and (34) we get:

$$a^{s+1} = (1-\alpha_s)a^s + \frac{\alpha_s}{\rho_s}(b^{s+1} - b^s) + \alpha_s(V_s - \mu_s)$$

   This equality gives

$$a^s = a_1^s + a_2^s + a_3^s + \frac{\alpha_{s-1}}{\rho_{s-1}} b^s,\ s=1,2,\dots \tag{35}$$

$$a_1^{s+1} = (1-\alpha_s)a_1^s + \alpha_s(V_s - \mu_s),\ a_1^1 = \alpha_0(V_0 - \mu_0)$$

$$a_2^{s+1} = (1-\alpha_s)a_2^s - \alpha_s \frac{\alpha_{s-1}}{\rho_{s-1}} b^s, \; a_2^1 = -\frac{\alpha_0}{\rho_0} b^0$$

$$a_3^{s+1} = (1-\alpha_s)a_3^s + \left[\frac{\alpha_{s-1}}{\rho_{s-1}} - \frac{\alpha_s}{\rho_s}\right] b^s, a_3^1 = (1-\alpha_0)a^0$$

Let us consider the sequence a_1^s. From its definition we have

$$(a_1^{s+1})^2 = (1-\alpha_s)^2(a_1^s)^2 + 2\alpha_s a_1^s(1-\alpha_s)(V_s - \mu_s) + \alpha_s^2(V_s - \mu_s)^2 \tag{36}$$

which implies that the $a_4^s = (a_1^s)^2 + \sum_{t=s}^{\infty} \alpha_t^2 (V_s - \mu_s)^2$ is supermartingale because $\sum_{t=0}^{\infty} \alpha_s^2 < \infty$, $E(V_s - \mu_s)^2 / b^0, \cdots, b^s) < M_1$. (36) also gives

$$E(a_1^{s+1})^2 \leq (1-\alpha_s)^2 E(a_1^s)^2 + M_1 \alpha_s^2$$

which gives $E(a_1^s)^2 \longrightarrow 0$ due to $\sum_{s=0}^{\infty} \alpha_s = \infty$. This implies $(a_1^2)^2 \longrightarrow 0$ a.s. because a_4^s is convergent nonnegative supermartingale.

From the definition of a_2^s we have:

$$|a_2^{s+1}| \leq (1-\alpha_s)|a_2^s| + \frac{\alpha_{s-1}}{\rho_{s-1}} |b^s| \leq |a_2^s| - \alpha_s \left[|a_2^s| - \frac{\alpha_{s-1}}{\rho_{s-1}} \bar{b}\right]$$

This inequality gives $|a_2^s| \longrightarrow 0$ a.s. because $\alpha_s / \rho_s \longrightarrow 0$ and $\sum_{s=0}^{\infty} \alpha_s = \infty$.

Consider now a_3^s and suppose that condition 6 (i) of the Theorem 3 is satisfied, i.e. $\frac{\alpha_{s+1}}{\rho_{s+1}} \leq \frac{\alpha_s}{\rho_s}$. In this case we have from the definition of a_3^s:

$$|a_3^{s+1}| \leq (1-\alpha_s)|a_3^s| + \left[\frac{\alpha_{s-1}}{\rho_{s-1}} - \frac{\alpha_s}{\rho_s}\right] |b^s|$$
$$\leq (1-\alpha_s)|a_3^s| + \bar{b}\left[\frac{\alpha_{s-1}}{\rho_{s-1}} - \frac{\alpha_s}{\rho_s}\right]$$

Summing up this inequality from s to k we get:

$$|a_3^k| \leq \prod_{j=s}^{k-1}(1-\alpha_j)|a_3^s| + \bar{b}\left[\frac{\alpha_{s-1}}{\rho_{s-1}} - \frac{\alpha_{k-1}}{\rho_{k-1}}\right] \tag{37}$$

Taking in (37), $s = 1$ we get $|a_3^k| < C_2$ for some $C_2 < \infty$. This gives for arbitrary s:

$$|a_3^k| \leq c_2 \prod_{j=s}^{k-1}(1-\alpha_j) + \bar{b}\left[\frac{\alpha_{s-1}}{\rho_{s-1}} - \frac{\alpha_{k-1}}{\rho_{k-1}}\right]$$

which implies $|a_3^k| \to 0$ for $k \to \infty$. Taking instead of 6 (i) conditions 6 (ii) and 6 (iii) of Theorem 3, we get $|a_3^k| \to 0$ similarly.

Thus, we obtain that the first three terms in (35) approach 0 as $s \to \infty$. $\frac{\alpha_{s-1}}{\rho_{s-1}} b^s \to 0$ due to assumptions 1 and 2. This gives $a^s \to 0$ a.s. and completes the proof of the first part.

2. Consider now the sum

$$\begin{aligned}
\sum_{t=s}^{k} \alpha_t b^t V_t &= \sum_{t=s}^{k} \alpha_t b^t \mu_t + \sum_{t=s}^{k} \alpha_t b^t (V_t - \mu_t) \qquad (38)\\
&= \frac{1}{2}\sum_{t=s}^{k} \alpha_t b^t \mu_t + \frac{1}{2}\sum_{t=s}^{k} \alpha_t (b^{t+1} - \rho_t \mu_t)\mu_t \\
&+ \sum_{t=s}^{k} \alpha_t b^t (V_t - \mu_t) = \frac{1}{2}\sum_{t=s}^{k} \frac{\alpha_t}{\rho_t}((b^{t+1})^2 - (b^t)^2) \\
&- \frac{1}{2}\sum_{t=s}^{k} \alpha_t \rho_t \mu_t^2 + \sum_{t=s}^{k} \alpha_t b^t (V_t - \mu_t) \\
&= \frac{1}{2}\sum_{t=s}^{k} \frac{\alpha_t}{\rho_t}((b^{t+1})^2 - (b^t)^2) + \\
&+ \varepsilon_1(s) - \frac{1}{2}\sum_{t=s}^{k} \alpha_t \rho_t V_t^2
\end{aligned}$$

We denoted here

$$\varepsilon_1(s) = \sum_{t=s}^{k} \alpha_t \left[b^t (V_t - \mu_t) - \frac{1}{2}\rho_t (V_t - \mu_t)^2 + \rho_t V_t (V_t - \mu_t) \right]$$

Similarly to the first part of the proof we obtain $\varepsilon_1(s) \to 0$ a.s. as $s \to \infty$ because

$$\sum_{t=0}^{\infty} \alpha_t^2 < \infty, \ |V_t| < \bar{V}, \ E((V_t - \mu_t)^2 / b^0, \cdots, b^t) < M_1 < \infty$$

Proceeding with the estimate we get:

$$\sum_{t=s}^{k} \frac{\alpha_t}{\rho_t}((b^{t+1})^2 - (b^t)^2) = -\frac{\alpha_s}{\rho_s}(b^s)^2 + \alpha_k \rho_k (b^{k+1})^2 + \sum_{t=s+1}^{k} (b^t)^2 \left[\frac{\alpha_{t-1}}{\rho i - 1} - \frac{\alpha_t}{\rho_t} \right]$$

The last term can be treated similar to the first part of the proof. After substitution of the previous inequality into (38) we get:

$$\sum_{t=s}^{k} \alpha_t b^t V_t = \sum_{t=s}^{k} \alpha_t \beta_t + \varepsilon(s)$$

where in the case if 3(i) of the Theorem 3 is true:

$$\beta_t = -\frac{1}{2}\rho_t V_t^2, \ |\varepsilon(s)| \leq |\varepsilon_1(s)| + \bar{b}^2 \left[\frac{\alpha_s}{\rho_s} + \frac{\alpha_k}{\rho_k} \right]$$

in the case if 3(ii) is true

$$\beta_t = -\frac{1}{2}\left[\rho_t V_t^2 + (b^t)^2 \frac{1}{\alpha_t} \left(\frac{\alpha_{t-1}}{\rho_{t-1}} - \frac{\alpha_t}{\rho_t} \right) \right]$$

$$\varepsilon(s) = \varepsilon_1(s) - \frac{1}{2}\frac{\alpha_s}{\rho_s}(b^s)^2 + \frac{1}{2}\frac{\alpha_k}{\rho_k}(b^{k+1})^2$$

and if 3(iii) is true

$$\beta_t = \frac{1}{2}\rho_t V_t^2, \quad |\varepsilon(s)| \le |\varepsilon_1(s)| + \frac{1}{2}\bar{b}^2\left[\frac{\alpha_s}{\rho_s} + \frac{\alpha_k}{\rho_k}\right] + \bar{b}^2 \sum_{t=s+1}^{k} \left|\frac{\alpha_{t-1}}{\rho_{t-1}} - \frac{\alpha_t}{\rho_t}\right|$$

In all three cases we obtain

$$\varepsilon(s) \to 0, \beta_s \to 0 \quad \text{a.s.}$$

which completes the proof.

The next lemma deals with important property of sets of ε-solutions of convex optimization problems. For arbitrary function $p(y)$, $y \in X \subset R^n$ define ε-optimal set:

$$X(\varepsilon,p) = \{y : y \in X, p(y) \ge \sup_{z \in X} p(z) - \varepsilon\}$$

and for arbitrary sets $A_1, A_2 \subset R^n$ define the Hausdorf distance:

$$d(A_1,A_2) = \max\left\{\sup_{z_1 \in A_1}\inf_{z_2 \in A_2} \|z_1 - z_2\|, \sup_{z_1 \in A_2}\inf_{z_2 \in A_1} \|z_1 - z_2\|\right\}$$

LEMMA 2. Suppose that the following assumptions are satisfied:

1. The set $X \subset R^n$ is compact and

$$\max_{z_1, z_2 \in X} \|z_1 - z_2\| \le D$$

2. $p_1(y), p_2(y)$ are finite concave functions on some vicinity of X,

$$\max_{y \in X} |p_1(y) - p_2(y)| \le \eta$$

Then exists $\bar{\varepsilon} > 0$ such that

$$d(X(\varepsilon,p_1), X(\varepsilon,p_2)) \le \frac{2D\eta}{\varepsilon}$$

for $0 < \varepsilon \le \bar{\varepsilon}$

This lemma is straightforward reformulation of results, obtained in [15, 16]. It shows that ε-optimal sets exhibit continuous behavior.

PROOF OF THEOREM 3.

1. Let us prove at first that the sequence c^s is bounded with probability 1, i.e. exists random variable $\bar{c}$ such that $\bar{c} < \infty$ a.s. and $|c^s| < \bar{c}$. From (31) we have:

$$c^k = c^s + \sum_{i=s}^{k-1} \rho_i \xi^i = c^s + \sum_{i=s}^{k-1} \rho_i V_c^i(c^i) + \sum_{i=s}^{k-1} \rho_i (\xi^i - V_c^i(c^i)) \tag{39}$$

From assumption 5 we obtain

$$E((\xi^i - V_c^i(c^i))^2 / c^0, \ldots, c^i) < M_1$$

which gives

$$E \sum_{i=0}^{\infty} \left[\rho_i (\xi^i - V_c^i(c^i))\right]^2 \le M_1 \sum_{i=0}^{\infty} \rho_i^2 < \infty$$

and therefore

$$\left| \sum_{i=s}^{k-1} \rho_i (\xi^i - V_c^i(c^i)) \right| \le \tau(s)$$

where $\tau(s) \longrightarrow 0$ a.s.

According to the definition of $V_c^i(c^i)$ we have

$$V_c^i(c^i) = \begin{cases} H^l(Y) - b & \text{if } c^i > \bar{f} \\ H^u(Y) - b & \text{if } c^i < -\bar{f} \end{cases}$$

and $|V_c^i(c^i)| \le |H^u|(Y) + |H^l|(Y) + |b| = M_2 < \infty$ for arbitrary i due to finiteness of the measures H^u and H^l. Now suppose that $c^k > \bar{f}$. Take s such that $c^i \ge \bar{f}$ for $s < i \le k$ and $c^s < \bar{f}$. If such s does not exist then take $s = 0$. We obtain from (39):

$$c^k \le c^s + \rho_s V_c^s(c^s) + \tau(s) \le \max\{c^0, \bar{f}\} + M_2\rho_s + \tau(s)$$

because $V_c^i(c^i) \le 0$ for $s < i \le k$. This inequality proves that $c^k < \bar{c}_1$ for some random variable $\bar{c}_1 < \infty$ a.s. Considering the case when $c^k < -f$ we get $c^k > \bar{c}_2$, $\bar{c}_2 > \infty$ a.s. and finally $|c^k| < \max\{\bar{c}_1, \bar{c}_2\} = \bar{c} < \infty$ a.s.

2. Let us prove now that $H^s(Y) \to b$ as $s \to \infty$. Take $a^s = H^s(Y) - b$. Then

$$a^{s+1} = H^{s+1}(Y) - b = \int_Y d[(1-\alpha_s)H^s + \alpha_s \bar{H}^s] - b$$
$$= (1-\alpha_s)a^s + \alpha_s[H^u(Z^+(c^s, f(s,y))) + H^l(Y \setminus Z^+(c^s, f(s,y))) - b]$$

which gives

$$a^{s+1} = (1-\alpha_s)a^s + \alpha_s V_c^s(c^s)$$

Taking now in (33) $V_s = V_c^s(c^s)$ and in (34) $b^s = c^s$, $\mu_s = \xi^s$ we have all assumptions of the Lemma 1 satisfied due to conditions of the theorem and boundedness of the sequence c^s. This proves that $a^s \to 0$ i.e. $H^s(Y) \to b$.

3. Let us verify that $\Psi(H^s) \to \max_{H \in G} \Psi(H)$ Condition 2 of the theorem gives the following inequality:

$$\Psi(H^{s+1}) = \Psi((1-\alpha_s)H^s + \alpha_s \bar{H}^s) = \Psi(H^s) \tag{40}$$
$$+ \alpha_s \int_Y g(y,H^s)d(\bar{H}^s - H^s) + o(\alpha_s)$$
$$= \Psi(H^s) + \alpha_s \int_Y f(s,y)d(\bar{H}^s - H_f^s)$$
$$+ \alpha_s(\int_Y f(s,y)dH_f^s - \int_Y g(y,H^s)dH_g^s)$$
$$+ \alpha_s \int_Y (g(y,H^s) - f(s,y))d\bar{H}^s$$
$$+ \alpha_s \int_Y g(y,H^s)d(H_g^s - H^s) + o(\alpha_s)$$

We denoted here $H_f^s \in G$, $H_g^s \in G$,

$$\int_Y f(s,y)dH_f^s = \max_{H\in G} \int_Y f(s,y)dH$$
$$\int_Y g(y,H^s)dH_g^s = \max_{H\in G} \int_Y g(y,H^s)dH$$

Let us denote $M_3 = |H^u|(Y) + |H^l|(Y)$. For any $H \in G$ we have $|H|(Y) \le M_2$. Condition 4 of the theorem and definitions of measures H_f^s and H_g^s imply:

$$\left| \int_Y f(s,y)dH_f^s - \int_Y g(y,H^s)dH_g^s \right| \le M_3\Delta_s \tag{41}$$

And in the similar way

$$\left| \int_Y (g(y,H^s) - f(s,y))d\bar{H}^s \right| \le M_3\Delta_s \tag{42}$$

Consider now in more detail the term $\int_Y f(s,y)d(\bar{H}^s - H_f^s)$. According to the Theorem 1, the measures H_f^s has the following structure:

$$H_f^s(A) = \begin{cases} H^u(A) & \text{for } A \subset Z^+(c_*^s, f(s,y)) \\ H^l(A) & \text{for } A \subset Z^-(c_*^s, f(s,y)) \\ H^u(A) \ge H_f^s(A) \ge H^l(A) & \text{for } A \subset Z^0(c_*^s, f(s,y)) \end{cases}$$

where

$$c_*^s = \inf_c \{c : H^u(Z^+(c, f(s,y))) + H^l(Y \setminus Z^-(c, f(s,y))) \le b\}$$

The measure $\bar{H}^s$ has the similar structure for $c_*^s = c^s$ and $\bar{H}^s(A) = H^l(A)$ for $A \subset Z^0(c' f(s,y))$. Assuming $c^s > c_*^s$ we obtain:

$$\begin{aligned}
\int_Y f(s,y)\, d(\bar{H}^s - H_f^s) &= c^s \int_{Z^+(c^s, f(s,y))} d(H^u - H_f^s) \\
&+ c^s \int_{Z^-(c_*^s, f(s,y))} d(H^l - H_f^s) + \int_{Y \setminus Z^-(c_*^s, f(s,y)) \setminus Z^+(c^s, f(s,y))} d(H^l - H_f^s) \\
&= c^s (H^u(Z^+(c^s, f(s,y))) + H^l(Z^-(c^s, f(s,y))) - b) \\
&+ \int_{Y \setminus Z^-(c_*^s, f(s,y)) \setminus Z^+(c^s, f(s,y))} (c^s - f(s,y))\, d(H_f^s - H^l) \quad \ge c^s V_c^s(c^s)
\end{aligned}$$

Considering the case $c_*^s \ge c^s$ we get the same result, i.e.

$$\int_Y f(s,y)\, d(\bar{H} - H_f^s) \ge c^s V_c^s(c^s) \tag{43}$$

Substitution (41)-(43) into (40) and summation from s to k gives the following inequality:

$$\begin{aligned}
\Psi(H^k) &\ge \Psi(H^s) + \sum_{i=s}^{k-1} \alpha_i c^i V_c^i(c^i) \\
&+ \sum_{i=s}^{k-1} \alpha_i \left[\int_Y g(y, H^i)\, d(H_g^i - H^i) - \tau_1^i \right]
\end{aligned} \tag{44}$$

where

$$\tau_1^i = 2 M_3 \Delta_i + o(\alpha_i)/\alpha_i$$

Taking now in (34) $b^s = c^s, \mu_s = \xi^s, V_s = V_c^s(c^s)$, we obtain from the second part of Lemma 1 that

$$\sum_{i=s}^{k-1} \alpha_i c^i V_c^i(c^i) = \sum_{i=s}^{k-1} \alpha_i \beta_i + \varepsilon(s)$$

where $\beta_s \to 0$ and $\varepsilon(s) \to 0$ as $s \to \infty$. Substituting this into (44) we get the following basic inequality:

$$\Psi(H^k) \geq \Psi(H^s) + \sum_{i=s}^{k-1} \alpha_i \left[\int_Y g(y, H^i) d(H_g^i - H^i) - \tau_2^i \right] + \varepsilon(s) \tag{45}$$

where

$$\tau_2^i \to 0 \text{ as } i \to \infty, \quad \tau_2^i = \tau_1^i + \beta_i$$

Suppose now that there exists s^* and positive r such that

$$\int_Y g(y, H^s) d(H_g^s - H^s) > r > 0$$

for $s \geq s^*$. Then from (45) we obtain the following inequality for $k > s^*$:

$$\Psi(H^k) \geq \Psi(H^{s^*}) + \sum_{i=s}^{k-1} \alpha_i (r - \tau_2^i) + \varepsilon(s)$$

However, $\tau_2^i < r/2$ for sufficiently large s^* and $i > s^*$. Therefore last inequality contradicts with the fact that $\Psi(H^k)$ is bounded. This proves existence of the sequence s_k such that

$$\max \{ 0, \int_Y g(y, H^{s_k}) d(H_g^{s_k} - H^{s_k}) \} \to 0 \tag{46}$$

as $k \to \infty$.

The concavity of function $\Psi(H)$ gives us for H^* such that

$$\Psi(H^*) = \max_{H \in G} \Psi(H)$$

the following inequality:

$$\Psi(H^*) - \Psi(H^s) \leq \int g(y, H^s) d(H^* - H^s) \leq \int g(y, H^s) d(H_g^s - H^s) \tag{47}$$

This inequality together with (46) implies

$$\max\{0, \Psi(H^*) - \Psi(H^{s_k})\} \to 0$$

Now suppose that for some sequence n_k and for some $r > 0$ we have

$$\Psi(H^*) - \Psi(H^{n_k}) > 3r$$

for $k > k^*$. We may assume without loss of generality that

$$\Psi(H^*) - \Psi(H^{s_k}) < r$$

and $s_k < n_k < s_{k+1} \cdots$ for $k > k^*$.

Consider the sequence m_k such that

$$s_k < m_k \le n_k, \quad \Psi(H^*) - \Psi(H^{m_k - 1}) < r$$

and $\Psi(H^*) - \Psi(H^i) \ge r$ for $m_k \le i \le n_k$. Boundedness and concavity of the function $\Psi(H)$ on the set $G + G_\varepsilon$ (assumption 2) implies that $\Psi(H)$ satisfies Lipschitz condition on G with some constant C [11] and we have particularly

$$\| g(y,H) \|_\infty < C \quad \text{for} \quad H \in G$$

and

$$| \Psi(H^{s+1}) - \Psi(H^s) | \le C \, \alpha_s \, |\overline{H}^s|(Y) \to 0 \text{ as } s \to \infty.$$

Therefore

$$\Psi(H^*) - \Psi(H^{m_k}) < 2r$$

for k sufficiently large. Observe now that (47) gives

$$\int g(y,H^i) \, d(H_g^i - H^i) \ge r$$

for $m_k \le i \le n_k$. Thus we obtain from (45):

$$\Psi(H^{n_k}) \ge \Psi(H^{m_k}) + \sum_{i=m_k}^{n_k - 1} \alpha_i (r - \tau_2^i) + \varepsilon(m_k)$$

Considering k is large enough for

$$r > \tau_2^t \quad \text{and} \quad |\varepsilon(m_k)| < r$$

we get

$$\Psi(H^{n_k}) > \Psi(H^{m_k}) - r$$

which finally gives

$$\Psi(H^*) - \Psi(H^{n_k}) < \Psi(H^*) - \Psi(H)^{m_k} + r < 3r$$

which contradicts assumption that

$$\Psi(H^*) - \Psi(H^{n_k}) > 3r$$

Therefore

$$\max \{ 0, \max_{H \in G} \Psi(H) - \Psi(H^s) \} \longrightarrow 0$$

Consider now the measure $\tilde{H}^s \in G$

$$\tilde{H}^s = (1 - \tilde{\alpha}_s) H^s + \tilde{\alpha}_s H_1^s$$

where

$$\tilde{\alpha}_s = \begin{cases} \dfrac{H^s(Y) - b}{H^s(Y) - H^l(Y)} & \text{if } H^s(Y) > b \\ \dfrac{b - H^s(Y)}{H^u(Y) - H^s(Y)} & \text{if } H^s(Y) < b \\ 0 & \text{if } H^s(Y) = b \end{cases}$$

and

$$H_1^s = \begin{cases} H^l & \text{if } H^s(Y) > b \\ H^u & \text{if } H^s(Y) < b \end{cases}$$

Assume that $H^u(Y) > b$, $H^l(Y) < b$, this gives $\tilde{\alpha}_s \longrightarrow 0$ because

$H^s(Y) \longrightarrow b$.

This gives

$$\Psi(\tilde{H}^s) = (1-\tilde{\alpha}^s)\Psi(H^s) + \tilde{\alpha}_s \int g(y, H^s) d(H_1^s - H^s) + o(\tilde{\alpha}_s) \le (1-\tilde{\alpha}^s)\Psi(H^s) + \tilde{\alpha}_s(2C+\varepsilon_s) + o(\tilde{\alpha}_s)$$

where $\varepsilon \longrightarrow 0$ as $s \longrightarrow \infty$ due to $H^s(Y) \longrightarrow b$. Finally we obtain

$$\max_{H \in G} \Psi(H) \ge \Psi(\tilde{H}^s) \ge \Psi(H^s) + \tilde{C}\tilde{\alpha}_s$$

for some $\tilde{C} < \infty$. Therefore in this case

$$\max\{0, \Psi(H^s) - \max_{H \in G} \Psi(H)\} \longrightarrow 0$$

Suppose now that, for instance, $H^l(Y)=b$. In this case $\int_Y d(H^s - H^l) \longrightarrow 0$ due to $H^s(Y) \longrightarrow b$ and $H^s(A) \ge H^l(A)$ for any $A \in \Xi$. Thus

$$\Psi(H^s) - \Psi(H^l) \le \int_Y g(y, H^l) d(H^s - H^l) \longrightarrow 0$$

due to

$$\| g(y, H^l) \|_\infty < C.$$

The case $H^u(Y)=b$ can be treated in the same way. Thus, we obtain

$$\Psi(H^s) \longrightarrow \max_{H \in G} \Psi(H)$$

Now for each convergent sequence $H^{s_k} \longrightarrow H^*$ wsp have

$$\Psi(H^*) = \max_{H \in G} \Psi(H)$$

and due to condition 3 of the theorem we obtain

$$\| g(y, H^{s_k}) - g(y, H^*) \|_\infty \longrightarrow 0.$$

This gives in turn

$$||f(s_k, y) - g(y, H^*)||_\infty \longrightarrow 0$$

for any convergent sequence $H^{s_k} \longrightarrow H^*$.

4. Let us investigate now the behavior of c^s. For arbitrary continuous function $p(y)$, let us denote

$$r(c,p) = H^u\left[Z^+(c,p(y))\right] + H^l\left[Y\setminus Z^+(c,p(y))\right] - b$$
$$R(c,p) = \int_T^c r(t,p)dt$$

for some fixed T. In this notation

$$V_c^s(c) = r(c, f(s,y)),\ V^s = R(c, f(s,y)),$$
$$W^s(c) = R(c, g(y, H^s))$$

Consider two arbitrary functions $p_1(y)$ and $p_2(y)$; $p_1, p_2 \in L_\infty(Y, \Xi, H^\Delta)$ and

$$||p_1(y) - p_2(y)||_\infty \le \eta \tag{48}$$

Take

$$Y^0 = \{y:\ y \in Y,\ |p_1(y) - p_2(y)| > \eta\}$$

We have $H^u(Y^0) = H^l(Y^0)$ and therefore

$$r(c,p) = H^u\left[Z^+(c,p(y))\setminus Y^0\right] + H^l\left[Y\setminus Z^+(c,p(y))\setminus Y^0\right] + H^u(Y^0) - b$$

for $p(y) \in L_\infty(Y,\Xi,H^\Delta)$ and

$$Z^+(c+\eta, p_1(y))\setminus Y^0 \subseteq Z^+(c, p_2(y))\setminus Y^0 \subseteq Z^+(c-\eta, p_1(y))\setminus Y^0$$

which gives

$$r(c+\eta, p_1(y)) \le r(c, p_2(y)) \le r(c-\eta, p_1(y))$$

Now we have the estimate:

$$R(c,p_2) = \int_T^c r(t,p_2(y))dt \le \int_T^c r(t-\eta, p_1(y))dt$$
$$= \int_{T-\eta}^{c-\eta} r(t,p_1(y))dt = R(c,p_1) + \int_T^{T-\eta} r(t,p_1(y))dt$$
$$+ \int_{c-\eta}^{c} r(t,p_1(y))dt \le R(c,p_1) + M_4\,\eta$$

where

$$M_4 = 2\max\{|H^u(Y)-b|, |H^l(Y)-b|\}$$

Finally this gives for arbitrary functions $p_1, p_2 \in L_\infty(Y, \Xi, H^\Delta)$ which satisfy (48):

$$\sup_c |R(c,p_1) - R(c,p_2)| \le M_4\,\eta \tag{49}$$

This inequality implies

$$\sup_c |W^s(c) - W^{s+1}(c)| \le M_4 \delta_s \tag{50}$$

$$\sup |W^s(c) - V^s(c)| \le M_4 \Delta_s \tag{51}$$

We shall consider the method (31)

$$c^{s+1} = c^s + \rho_s \xi^s$$
$$E(\xi^s / c^0, \cdots, c^s) = V_c^s(c^s)$$

as nonstationary maximization method [15,16] and prove that

$$\max_c W^s(c) - W^s(c^s) \longrightarrow 0$$

with probability 1.

In order to do this we need the definition of ε-optimal set $X_s(\varepsilon)$:

$$X_s(\varepsilon) = \{c_\varepsilon : \max_c W^s(c) - W^s(c_\varepsilon) \le \varepsilon\}$$

and optimal set

$$X_s^* = \{c^* : \max_c W^s(c) = W^s(c^*)\} = X_s(0)$$

It was proved that C exists such that $\|g(y,H)\|_\infty < C$ for all $H \in G$. Therefore according to the definition of W^s there are three opportunities:

(i) $X_s(\varepsilon) \subset \{c: |c| \le 2C\}$

(ii) $X_s(\varepsilon) \subset \{c: c \ge -2C\}$ and $\{c: c \ge C\} \subset X_s(\varepsilon)$

(iii) $X_s(\varepsilon) \subset \{c: c \le 2C\}$ and $\{c: c \le -C\} \subset X_s(\varepsilon)$

for all s, $0 \le \varepsilon \le \bar{\varepsilon}$ and some $\bar{\varepsilon} > 0$. This shows for $0 \le \varepsilon \le \bar{\varepsilon}$:

$$d(X_s(\varepsilon), X_{s+1}(\varepsilon)) = d(\tilde{X}_s(\varepsilon), \tilde{X}_{s+1}(\varepsilon))$$

where

$$\tilde{X}_s(\varepsilon) = \{c: |c| \le 2C, c \in X_s(\varepsilon)\}$$

and d denotes Hausdorf distance. Sets $\tilde{X}_s(\varepsilon)$ are compact and applying to them Lemma 2 and (50) we obtain

$$d(X_s(\varepsilon), X_{s+1}(\varepsilon)) \le M_5 \delta_s$$

where $M_5 = \dfrac{8CM_4}{\varepsilon}$ and we used also condition 3 of the theorem. This inequality enables us to get the following estimate for $w^s = \min_{z \in X_s(\varepsilon)} \|c^s - z\|^2$ and $\bar{w}^s = \min_{z \in X_s(\varepsilon)} (c^{s+1} - z)^2$

$$w^s \le (\sqrt{\bar{w}^s} + d(X_s(\varepsilon), X_{s+1}(\varepsilon))^2 \le \bar{w}^s + 2M_5 \bar{C} \delta_S + M_5^2 \delta_s^2 \qquad (52)$$

where $\bar{C} < \infty$ a.s. due to part 1 of the proof. Denoting now

This inequality is true for all $\varepsilon > 0$ which gives

$$\max_c W^s(c) - W^s(c^s) \longrightarrow 0$$

with probability 1. Consider now arbitrary convergent subsequence $c^{s_k} \longrightarrow c^*$ of the sequence c^s. Due to compactness of the sequence H^{s_k} we can assume without loss of generality that $H^{s_k} \longrightarrow H^* \in G$, $\Psi(H^*) = \max_{H \in G} \Psi(H)$. According to condition 3 this means

$$\| g(y,H^*) - g(y,H^{s_k} \|_\infty \longrightarrow 0$$

which gives, according to (49)

$$\sup_c | W^{s_k}(c) - W^*(c) | \longrightarrow 0$$

where $W^*(c) = R(c, g(y, H^*))$. Therefore $W^*(c^*) = \max_c W^*(c)$ according to the Remark to theorem 1, this means that all solutions of the problem

$$\max_{H \in G} \int_Y g(y, H^*) dH$$

have the representation

$$H(A) = \begin{cases} H^u(A) & \text{for } A \subset Z^+(c^*, g(y, H^*)) \\ H^l(A) & \text{for } A \subset Z^-(c^*, g(y, H^*)) \\ H^u(A) \geq H(A) \geq H^l(A) & \text{for } A \subset Z^0(c^*, g(y, H^*)) \end{cases}$$

But according to theorem 2, H^* is among these solutions which together with

$$\| f(s_k, y) - g(y, H^*) \|_\infty \longrightarrow 0$$

completes the proof.

$$E(\tau_1^{s+1}/\tau_1^0, \cdots, \tau_1^s) = 0; \sum_{i=0}^{\infty} E(\tau_1^i)^2 < \infty,$$
$$\sum_{i=0}^{\infty} E\,\tau_2^i < \infty, \quad \tau_2^i > 0.$$

According to the martingale theorems [17] this gives

$$\sum_{i=0}^{\infty} \tau_1^i < \infty, \quad \sum_{i=0}^{\infty} \tau_2^i < \infty \tag{55}$$

with probability 1. Making now summation from k to s, substituting (53) into (52) and applying (49)-(51), (54) we obtain:

$$w^s \le w^k - M_6 \sum_{i=k}^{s-1} (w^i - \tau_3^i) + \tau_4^s \tag{56}$$

where $M_6 = 2\varepsilon/3C$,

$$\tau_3^s = \frac{2M_5\bar{C}}{M_6}\,\frac{\delta_s}{\rho_s} + \frac{M_5^2}{M_6}\,\frac{\delta_s^2}{\rho_s} + \frac{4M_4}{M_6}\,\Delta_s$$
$$\tau_4^s = \sum_{i=k}^{s-1} \tau_1^i + \sum_{i=k}^{s-1} \tau_2^i$$

According to the assumptions 3,4,5 we have $\delta_s/\rho_s \to 0$, $\Delta_s \to 0$ which gives $\tau_3^s \to 0$ as $s \to \infty$. We obtain from (55) that $\tau_4^s \to 0$ with probability 1.

Inequality (56) is the basic inequality which gives after fairly standard argument (see [14-16]) $w^s = \min_{z \in X_s(\varepsilon)} (c^s - z)^2 \to 0$ with probability 1 for $0 < \varepsilon < \bar{\varepsilon}$. Definition of function $W^s(c)$ implies

$$|W^s(c_1) - W^s(c_2)| \le M_3 |c_1 - c_2|$$

and therefore

$$\max_c W^s(c) - W^s(c^s) \le \varepsilon + M_3 \sqrt{w^s}$$

$$\bar{c}^s : \bar{c}^s \in X_s(\varepsilon), \ |c^s - \bar{c}^s| = \bar{w}^s$$

we obtain:

$$\begin{aligned} \bar{w}^s &\leq (c^{s+1} - \bar{c}^s)^2 = (c^s + \rho_s \xi^s - \bar{c}^s)^2 \\ &= w^s + 2\rho_s V_c^s(c^s)(c^s - \bar{c}^s) + \rho_s^2(\xi^s)^2 \\ &+ 2\rho_s(\xi^s - V_c^s(c^s))(c^s - \bar{c}^s) \leq w^s \\ &+ 2\rho_s(V^s(c^s) - V^s(\bar{c}^s)) + \rho_s^2(\xi^s)^2 \\ &+ 2\rho_s(\xi^s - V_c^s(c^s))(c^s - \bar{c}^s) \leq w^s \\ &+ 2\rho_s(W_c^s(\bar{c}^s), c^s - \bar{c}^s) + 4M_4\rho_s\Delta_s \\ &+ \rho_s^2(\xi^s)^2 + 2\rho_s(\xi^s - V_c^s(c^s))(c^s - \bar{c}^s) \end{aligned} \tag{53}$$

Let us estimate the term $W_c^s(\bar{c}^s)(c^s - \bar{c}^s)$ and assume at first that $c^s \bar{\in} X_s(\varepsilon)$, therefore $c^s \neq \bar{c}^s$. Assuming $|\bar{c}^s| < 2C$ for sufficiently small ε and taking $W^s(c^*) = \max W^s(c)$ we obtain:

$$\begin{aligned} \varepsilon &= W^s(c^*) - W(\bar{c}^s) \leq W_c^s(\bar{c}^s)(c^* - \bar{c}^s) \\ &\leq |W_c^s(\bar{c}^s)| \, |c^* - \bar{c}^s| \leq 3C \, |W_c^s(\bar{c}^s)| \end{aligned}$$

which gives

$$|W_c^s(\bar{c}^s)| \geq \frac{\varepsilon}{3C}$$

for $c^s \bar{\in} X_s(\varepsilon)$. This implies

$$W_c^s(\bar{c}^s)(c^s - \bar{c}^s) \leq -\frac{\varepsilon}{3C} w^s \tag{54}$$

which is also true for $c^s \in X_s(\varepsilon)$ because in this case $w^s = 0$.

Let us denote $\tau_1^2 = 2\rho_s(\xi^s - V_c^s(c^s))(c^s - \bar{c}^s)$, $\tau_2^2 = \rho_s^2(\xi^s)^2$ and consider sums $\sum_{t=0}^{\infty} \tau_1^t$ and $\sum_{t=0}^{\infty} \tau_2^t$. Due to assumption 5 we have

Theorem 3 means that if $H^{\Delta}(Z^0(c f(s,y))) = 0$ starting from some s then the measure $\bar{H}^s = \Gamma(c^s f(s,y))$ defined in (23) is good approximation for the optimal solution if s is large. If this is not the case then $f(s,y)$ can still be used for constructing optimal solution, but more careful choice of c is needed.

4. PARAMETRIC DEPENDENCE OF $g(y,H)$ ON MEASURE

The method described in the previous section is actually general framework for solving problem (7)-(9). In order to implement it, we have to specify how to compute function $f(s,y)$ satisfying

$$\| f(s,y) - g(y,H^s) \|_{\infty}$$

without calculating distribution H^s. This can be done for particular ways of dependence between gradient $g(y,H^s)$ and measure H^s. One such important case will be studied in this section. It will be assumed that

$$g(y,H) = Q(y,\alpha) \tag{57}$$

where

$$\alpha = \int_Y q(y)dH$$

$$q(y) = (q_1(y), \cdots, q_t(y))$$

and functions $q(y)$ and $Q(y,\alpha)$ are known. This case includes, for instance, finite population sampling example described in the introduction. Function $\Psi(M)$ from (7) usually has the form: $\Psi(M)$ = trace (M^{-p}) for p in $[-1,\infty]$ with $p=0$ interpreted as det (M^{-1}) (D- optimality) [2-6]. Therefore in this case

$$g(y,H) = ph^T(y)M^{-(p+1)}h(y) \tag{58}$$

and in particular for $p=0$ it is possible to take

$$g(y,H) = h^T(y)M^{-1}h(y) \tag{59}$$

where

$$M = M(H) = \int_Y h(y)h^T(y)dH,$$

i.e., finite population sampling is covered by (57).

If (57) is true, step 2c of the algorithm can be performed in the following way:

$$u^{s+1} = (1-\alpha_s)u^s + \alpha_s \eta^s \tag{60}$$

$$E(\eta^s / u^0, \cdots, u^s) = \int_Y g(y)d\bar{H}^s \tag{61}$$

$$f(s+1,y) = Q(y, u^{s+1})$$

The process depends on the random element ω from probability space in which sequences η^s from (60) and ξ^s from (31) are defined. All subsequent statements about convergence, boundedness and other properties are fulfilled a.s. with respect to probability measure in this space.

In these notations we shall get $g(y, H^s) = Q(y, a^s)$ where

$$a^s = \int_Y q(y)dH^s$$

Note that algorithm reduces to (31) and (60)-(61), formula (30) is not needed as long as we can compute u^s. Vector η^s from (61) can be easily computed in many cases. For instance, if H^u and H^l have densities H_y^u and H_u^l with respect to Lebesque measure, η^s can be chosen as follows:

$$\eta^s = \mu(Y)q(\omega^s)\bar{H}_y^s(\omega^s)$$

where

$$\bar{H}_y^s(\omega^s) = \begin{cases} H_y^u(\omega^s) & \text{if } \omega^s \in Z^+(c^s, Q(y, u^s)) \\ H_y^l(\omega^s) & \text{otherwise} \end{cases}$$

ω^s is distributed uniformly over Y and $\mu(Y)$ denotes Lebesque measure of Y.

The conditions for convergence of algorithm (30), (59)-(60) are summarized in the following theorem.

THEOREM 4. Suppose that assumptions 1,2,5,6 of the theorem 3 are satisfied and the following additional conditions are fulfilled:

1. $q_i(y) \in L_\infty(Y, \Xi, H^\Delta)$, $i = 1:t$
2. $||Q(y, \alpha') - Q(y, \alpha)||_\infty \to 0$ as $\alpha' \to \alpha$,

   $||Q(y, u^{s+1}) - Q(y, u^s)||_\infty \le \delta_s \to 0$ as $s \to \infty$.
3. $E(||\eta^s - \int_Y q(y) d\bar{H}^s||^2 / u^0, \cdots, u^s) \le M_1$

Then the sequences u^s, c^s are bounded a.s. and for any converging subsequence $u^{s_k} \to u^*$, $c^{s_k} \to c^*$ exists measure $H^* \in G_b$ such that $\Psi(H^*) = \max_{H \in G_b} \Psi(H)$ and

$$H^*(A) = \begin{cases} H^u(A) & \text{for } A \subset Z^+(c^*, Q(y, u^*)) \\ H^l(A) & \text{for } A \subset Z^-(c^*, Q(y, u^*)) \end{cases}$$

PROOF. It is sufficient to prove that conditions 3 and 4 of theorem 3 are satisfied. Condition 3 is obviously satisfied due to assumptions 1 and 2 of the theorem. Let us prove that $||u^s - \alpha^s|| \to 0$ a.s. This will give $||g(y, H^s) - f(s, y)||_\infty \to 0$ a.s. Denoting $z^s = \int_Y g(y) d\bar{H}^s$ we obtain the following estimate:

$$
\begin{aligned}
\| u^{s+1} - a^{s+1} \|^2 &= \| (1-\alpha_s) u^s + \alpha_s \eta^s - (1-\alpha_s) a^s - \alpha_s z^s \|^2 \\
&= (1-\alpha_s^2) \| u^s - a^s \|^2 + 2\alpha_s (1-\alpha_s)(u^s - a^s, \eta^s - z^s) \\
&+ \alpha_s^2 \| \eta^s - z^s \|^2
\end{aligned}
$$

which gives

$$E \| u^{s+1} - a^{s+1} \|^2 \le (1-\alpha_s)^2 E \| u^s - a^s \|^2 + M_1 \alpha_s^2$$

This inequality proves that $\| u^s - a^s \| \to 0$ a.s. due to $\sum_{i=1}^{\infty} \alpha_s = \infty$ and $\sum_{i=1}^{\infty} \alpha_2^2 < \infty$. Properties of the measures H^u, H^l and functions $q_i(y)$ yield boundedness of the set

$$A_G = \{a : \ a = \int_Y g(y) dH, H \in G\}$$

which together with continuity give

$$\| Q(y, a^s) - Q(y, u^s) \|_\infty \to 0 \text{ a.s.}$$

We have also $\sup_s \| Q(y, u^s) \|_\infty < \infty$ a.s. due to boundedness a.s. of the sequence u^s. All conditions of the Theorem 3 are satisfied and proof is completed.

The second part of assumption 2 is actually satisfied automatically and is only needed together with condition on stepsize $\delta_s / \rho_s \to 0$. If $Q(y, a)$ is locally Lipschitz with respect to a, i.e.

$$\| Q(y, a') - Q(y, a) \|_\infty \le L \| a' - a \|$$

assumption 2 is fulfilled automatically because in this case $\delta_s \sim \alpha_s$.

In some problems $Q(y, a)$ is not continuous with respect to a everywhere. In fact in optimal experiment design and finite population sampling $Q(y, a)$ is not defined for some a, because matrix M^{-1} from (59) may not exist. In this case theorem 4 remains correct if we assume that condition 2 is satisfied for $\alpha \in \bar{A}$ where $\bar{A}$ is closed, $A_G \subset \bar{A}$ and $u^s \in \bar{A}$ for all s.

In the remaining part of this section we shall consider in more detail the finite population sampling with D-optimality criterion and derive from (31), (60)-(61) specific algorithm for this case. We shall assume that matrix

$$M(H) = \int_Y h(y)h^T(y)dH$$

is invertible for all $H \in G_b$ and $h_i(y) \in L_\infty(Y,\Xi,H^\Delta)$. Method (60)-(61) will be transformed to

$$M_{s+1} = (1-\alpha_s)M_s + \alpha_s \eta^s \tag{62}$$

$$E(\eta^s / M_0, \cdots, M_s) = M(\bar{H}^s) \tag{63}$$

$$f(s,y) = h^T(y)M_s^{-1}h(y)$$

The method (62)-(63) requires matrix inversion on each iteration. This can be avoided by using the following equality:

$$(I + de^T)^{-1} = I \frac{de^T}{1+d^Te} \tag{64}$$

much in the same way as it is done in recursive least squares. Suppose, for instance, that measures H^u and H^l has densities H_y^u and H_y^l with respect to Lebesque measure. In this case we can take

$$M_{s+1} = (1-\alpha_s)M_s + \alpha_s \mu(Y)h(\omega^s)h^T(\omega^s)\bar{H}_y^s(\omega^s)$$

where ω^s is distributed uniformly over y. If $B_s = M_s^{-1}$ then (64) gives the following algorithm:

$$B_{s+1} = \frac{1}{1-\alpha_s}\left[B_s - \frac{\alpha_s \mu(Y)\bar{H}_y^s(\omega^s)B_s h(\omega^s)h^T(\omega^s)B_s}{1-\alpha_s+\alpha_s \mu(Y)\bar{H}_y^s(\omega^s)h^T(\omega^s)B_s h(\omega^s)}\right] \tag{65}$$

$$c^{s+1} = c^s + \rho_s(\mu(Y)\bar{H}_y^s(\omega^s) - b) \tag{66}$$

$$\bar{H}_y^s(\omega^s) = \begin{cases} H_y^u(\omega^s) & \text{if } h^T(\omega^s)B_s h(\omega^s) > c^s \\ H_y^l(\omega^s) & \text{otherwise} \end{cases}$$

and $\mu(Y)$ is the Lebesque measure of Y. If (c^*, B^*) is some accumulating point of the sequence (c^s, B_s) then the measure H^* with density H_y^*:

$$H_y^*(y) = \begin{cases} H_y^u(y) & \text{if } h^T(y)B^* h(y) < c^* \\ H_y^l(y) & \text{if } h^T(y)B^* h(y) > c^* \end{cases}$$

and such that $H^*(Y) = b$ will be optimal measure.

5. NUMERICAL EXPERIMENT

We used method (65)-(66) for solving the following problem:

$$\min_H \det [(\int_{-1}^{1} h(y)h^T(y)H_y(y)dy)^{-1}] \tag{67}$$

subject to

$$0 \le H_y(y) \le 0.5 \tag{68}$$

$$\int_{-1}^{1} H_y(y)dy = 0.5 \tag{69}$$

where

$$h(y) = \begin{bmatrix} 1 \\ y \\ y^2 \end{bmatrix}$$

The problem can be solved analytically and the solution is

$$H_y^*(y) = \begin{cases} 0.5 & \text{if } -1 \le y \le -0.71,\ -0.21 \le y \le 0.21,\ 0.71 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The optimal value of objective function $\Psi^* = 132.13$.

The method (65)-(66) was applied with initial value of level $c^0 = 1$, initial matrix B_0 diagonal with diagonal element 100. This choice corresponds to the density of initial measure $\bar{H}_y^0 \equiv 0.5$ on the whole interval $[-1,1]$ which is very poor initial approximation. The stepsizes ρ_s and α_s were taken piecewise-constant:

$$\rho_s = \begin{cases} 0.1 & 1 \le s \le 150 \\ 0.02 & 151 \le s \le 300 \\ 0.01 & 301 \le s \le 450 \\ 0.002 & 451 \le s \le 650 \\ 0.0008 & 651 \le s \le 900 \end{cases}$$

$$\alpha_s = \begin{cases} 0.008 & 1 \le s \le 150 \\ 0.001 & 151 \le s \le 450 \\ 0.0002 & 451 \le s \le 900 \end{cases}$$

150 runs were performed with the same initial values for c^0 and B_0, same stepsizes but with different sequences of random numbers, chosen independently. Thus we obtained 150 sequences

$$\{c^{s,j},\ B_{s,j}\}_{s=1}^{900}, \quad j = 1:150$$

where j is used for indexing different runs. For each iteration we calculated averaged violation Δ_1^s of constraint (69) by measure $H^{s,j}$ with density $H_u^{s,j}$

$$H_y^{s,j} = \begin{cases} 0.5 & \text{if } h^t(y)B_{s,j}h(y) > c^{s,j} \\ 0 & \text{otherwise} \end{cases}$$

$$\Delta_1^s = \frac{\sum_{j=1}^{150} \left| \int_{-1}^{1} H_y^{s,j}(y)\,dy - 0.5 \right|}{150}$$

The averaged difference Δ_2^s between value $\Psi(H^{s,j})$ of objective function (67) and optimal value Ψ^* was also computed:

$$\Delta_2^s = \frac{\sum_{j=1}^{150} \left| \Psi(H^{s,j}) - \Psi^* \right|}{150}$$

Evolution of these two quantities gives impression about average behavior of algorithm. The results are summarized in the following table:

Iteration number s	Stepsize ρ_s	Stepsize α_s	$\hat{c}^s$	Δ_1^s	Δ_2^s
10	0.1	0.008	1.450	0.5000	98.3
50	0.1	0.008	3.450	0.5000	98.3
100	0.1	0.008	4.534	0.0738	94.1
150	0.1	0.008	4.614	0.0640	87.1
240	0.02	0.001	4.650	0.0317	16.0
300	0.02	0.001	4.695	0.0138	8.4
350	0.01	0.001	4.697	0.0142	9.5
400	0.01	0.001	4.685	0.0121	7.7
450	0.01	0.001	4.669	0.0116	7.1
500	0.002	0.0002	4.665	0.0093	5.6
650	0.002	0.0002	4.661	0.0083	4.5
800	0.0008	0.0002	4.664	0.0067	3.7
900	0.0008	0.0002	4.662	0.0056	3.2

Here $\hat{c}^s = \frac{\sum_{j=1}^{150} c^{s,j}}{150}$ They show that algorithm behaves reasonably well, especially taking into account that the intitial point was very far from optimal solution. The first two rows of the table differ only in value of c^s because during the first approximately 80 iterations $H_y^{s,j} \equiv 0.5$. After this constraint violation Δ_1^s and the difference between the current and optimal value of the objective function Δ_2^S rapidly decrease until iteration 300 and then continue to decrease, but more slowly. In 900 iterations we obtained in average 0.012 relative accuracy in constraint violation and 0.035 relative accuracy of determining optimal value. Note that each iteration requires computation of $h(y)$ only at one point.

REFERENCES

1. Yu. Ermoliev, and A. Gaivoronski (1984), Duality relations and numerical methods for optimization problems on the space of probability measures with constraints on probability densities. Working Paper WP-84-46, Laxenburg, Austria: International Institute for Applied Systems Analysis.

2. H. P Wynn. Optimum submeasures with application to finite population sampling. Private communication.

3. H.P. Wynn (1977), Optimum designs for finite population sampling. S.S. Gupta and D.S. Moore, eds., in: Statistical Decision and Related Topics II. Academic Press, New York.

4. J. Kiefer, and J. Wolfowitz (1959), Optimum designs in regression problems. Annals of *Mathematicsl Statistics* 30:271-294.

5. P. Whittle (1973), Some general points in the theory of optimal experimental design. Journal of the *Royal Statistical Society*, Series B 35:123-150.

6. V. Fedorov (1972), Theory of Optimal Experiments. Academic Press, New York, 1972.

7. J. Birge, and R. Wets (1983), Designing approximation schemes for stochastic optimization problems, in particular for stochastic problems with recourse. Working paper WP-83-114, Laxenburg, Austria: International Institute for Applied Systems Analysis.

8. P. Kall, Karl Frauendorfer, and A. Ruszczyński (1984), Approximation techniques in stochastic programming. Working paper, Institute of Operations Research, University of Zurich.

9. W.K. Klein Haneveld (1984), Abstract LP duality and bounds on variables. Discussion paper 84-13-OR, University of Groningen.

10. N. Dunford, and J.T. Schwartz (1957), Linear Operators. Part I: General Theory Interscience Publ. Inc. New York.

11. R.T. Rockafellar (1970), Convex Analysis. Princeton University Press, Princeton.

12. F.H. Clarke (1983), Optimization and nonsmooth analysis, John Wiley & Sons, New York.

13. R.J. T. Morris (1979), Optimal constrained selection of a measurable set, *J. Math. Anal. Appl.* 70:546-562.

14. Yu. Ermoliev (1976), Methods of stochastic programming (in Russian). Nauka, Moscow.

15. A. Gaivoronski (1978), Nonstationary problem of stochastic programming with varying constraints, in: Yr. Ermoliev, I. Kovalenko, eds., Mathematical Methods of Operations Research and Reliability Theory. Institute of Cybernetics Press, Kiev, 1978.

16. Yu. Ermoliev, and A. Gaivoronski, Simultaneous nonstationary optimization estimation and approximation procedures. *Stochastics*, to appear.

17. M. Loeve (1955), Probability theory. Van Nostrand, New York.