Working Paper

A Branch and Bound Method for Stochastic Global Optimization

Vladimir I. Norkin
Georg Ch. Pflug
Andrzej Ruszczyński

WP-96-xxx
April 1996

IIASA
International Institute for Applied Systems Analysis ▫ A-2361 Laxenburg ▫ Austria
Telephone: +43 2236 807 ▫ Fax: +43 2236 71313 ▫ E-Mail: info@iiasa.ac.at

A Branch and Bound Method for Stochastic Global Optimization

Vladimir I. Norkin
Georg Ch. Pflug
Andrzej Ruszczyński

WP-96-xxx
April 1996

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

Abstract

A stochastic version of the branch and bound method is proposed for solving stochastic global optimization problems. The method, instead of deterministic bounds, uses stochastic upper and lower estimates of the optimal value of subproblems, to guide the partitioning process. Almost sure convergence of the method is proved and random accuracy estimates derived. Methods for constructing random bounds for stochastic global optimization problems are discussed. The theoretical considerations are illustrated with an example of a facility location problem.

Key words. Stochastic Programming, Global Optimization, Branch and Bound Method, Facility Location.

A Branch and Bound Method for Stochastic Global Optimization

Vladimir I. Norkin
Georg Ch. Pflug
Andrzej Ruszczyński

1 Introduction

Stochastic optimization problems belong to the most difficult problems of mathematical programming. Their solution requires either simulation-based methods, such as stochastic quasi-gradient methods [2], if one deals with a general distribution of the random parameters, or special decomposition methods for very large structured problems [1, 8, 14, 15], if the distribution is approximated by finitely many scenarios. Most of the existing computational methods (such as, e.g., all decomposition methods) are applicable only to convex problems. The methods that can be applied to multi-extremal problems, like some stochastic quasi-gradient methods of [7], usually converge to a local minimum.

There are, however, many important applied optimization problems which are, at the same time, stochastic and non-convex. One can mention here, for example, optimization of queueing networks, financial planning problems, stochastic optimization problems involving indivisible resources, etc. The main objective of this paper is to adapt the idea of the branch and bound method to such stochastic global optimization problems.

We consider problems of the following form:

$$\min_{x \in X \cap D} [F(x) = \mathbb{E} f(x, \theta(\omega))], \tag{1.1}$$

where X is a compact set in an n-dimensional Euclidean space $\mathbb{R}^n$, D is a closed subset of $\mathbb{R}^n$, $\theta(\omega)$ is an m-dimensional random variable defined on a probability space $(\Omega, \Sigma, \mathbb{P})$, $f : X \times \mathbb{R}^m \to \mathbb{R}$ is continuous in the first argument and measurable in the second argument, and $\mathbb{E}$ denotes the mathematical expectation operator. We assume that $f(x, \theta(\omega)) \le \bar{f}(\omega)$ for all $x \in X$, and $\mathbb{E}\bar{f} < \infty$, so that the expected value function in (1.1) is well-defined and continuous.

The reason for distinguishing the sets X and D is that we are going to treat X directly, as a 'simple' set (for example, a simplex, a parallelepiped, the integer lattice, etc.), while D can be defined in an implicit way, by some deterministic constraints.

Example 1.1. The customers are distributed in an area $X \subset \mathbb{R}^m$ according to some probability measure $\mathbb{P}$. The cost of serving a customer located at $\theta \in X$ from a facility at $x \in X$ equals $\varphi(x,\theta)$, where $\varphi : X \times X \to \mathbb{R}$ is quasi-convex in the first argument for $\mathbb{P}$-almost all θ. If there are n facilities located at $x^1, \ldots, x^n \in X$, the cost is

$$f(x^1, \ldots, x^n, \theta) = \min_{1 \le j \le n} \varphi(x^j, \theta).$$

The objective is to place n facilities in such a way that the expected cost

$$F(x^1, \ldots, x^n) = \mathbb{E}\left\{\min_{1 \le j \le n} \varphi(x^j, \theta)\right\}$$

is minimized. Since the objective function is invariant with respect to permutations of the locations $x^1, \ldots, x^n$, we can define D in a way excluding such permutations, e.g., $D = \{(x^1, \ldots, x^n) \in \mathbb{R}^{mn} : \ c^T x^j \le c^T x^{j+1}, \ j = 1, \ldots, n-1\}$, where $c \in \mathbb{R}^m$.

In the special case of $n = 1$ and

$$\varphi(x,\theta) = \begin{cases} 0 & \text{if } d(x,\theta) \le \delta, \\ 1 & \text{otherwise,} \end{cases} \tag{1.2}$$

with some convex distance function d, one obtains the problem of maximizing the probability of the event $A(x) = \{\omega \in \Omega : \ d(x, \theta(\omega)) \le \delta\}$.

Example 1.2. A portfolio consists of n asset categories. In time period t the value of assets in category j grows by a factor $\theta_j(t)$, where $\theta(t)$, $t = 1, 2, \ldots, T$, is a sequence of n-dimensional random variables. In the *fixed mix policy* one re-balances the portfolio after each time period to keep the proportions of the values of assets in various categories (the mix) constant. Each selling/buying assets induces transaction costs of a fraction α_j of the amount traded (for example). The problem is to find the mix that maximizes the expected wealth after T periods.

Denote the mix by $x \in X = \{x \in \mathbb{R}^n : \ x_j \ge 0, \ j = 1, \ldots, n, \ \sum_{j=1}^n x_j = 1\}$ and the wealth at the beginning of period t by $W(t)$. Then at the end of period t the wealth in category j equals $W(t) x_j \theta_j(t)$, while the transaction costs necessary to re-establish the proportion x_j are equal to $\alpha_j |W(t) x_j \theta_j(t) - x_j W(t) \sum_{\nu=1}^n x_\nu \theta_\nu(t)|$. Thus

$$W(t+1) = W(t) \sum_{j=1}^{n} \left(x_j \theta_j(t) - \alpha_j |x_j \theta_j(t) - x_j \textstyle\sum_{\nu=1}^n x_\nu \theta_\nu(t)|\right).$$

The objective (to be maximized) has therefore the form:

$$F(x) = \mathbb{E}\left\{\prod_{t=1}^{T} \sum_{j=1}^{n} \left(x_j \theta_j(t) - \alpha_j |x_j \theta_j(t) - x_j \textstyle\sum_{\nu=1}^n x_\nu \theta_\nu(t)|\right)\right\}.$$

Again, the set D may express additional requirements on the investments.

Other examples of stochastic global optimization problems, in which multi-extremal nature of the problem results from integrality constraints on decision variables, can be found in [10].

All these examples have common features:

- the objective function is multi-extremal;
- the calculation of the value of the objective requires evaluating a complicated multi-dimensional integral.

It is clear that we need special methods for stochastic global optimization to be able to solve problems of this sort. The approach proposed in this paper is a specialized branch and bound method for stochastic problems.

The main idea, as in the deterministic case, is to subdivide the set X into smaller subsets and to estimate from above and from below the optimal value of the objective within these subsets. In the stochastic case, however, deterministic lower and upper bounds are very difficult to obtain, so we confine ourselves to *stochastic lower and upper bounds*: some random variables whose expectations constitute valid bounds. They are used in the method to guide the partitioning and deletion process. Since it is far from being obvious that replacing deterministic bounds by their stochastic counterparts leads to meaningful results, the analysis of this question constitutes the main body of the paper.

In section 2 we describe the stochastic branch and bound method in detail and we make a number of assumptions about the stochastic bounds. Section 3 is devoted to the convergence analysis; we prove convergence with probability one, derive confidence intervals, and develop probabilistic deletion rules. In section 4 the problem of constructing stochastic bounds is discussed. We show that for many stochastic global optimization problems stochastic bounds can be easily obtained, while deterministic bounds and practically unavailable. Finally, in section 5 the theoretical considerations are illustrated with a computational example.

2 The method

2.1 Outline of the method

In the branch and bound method the original compact set X is sequentially subdivided into compact subsets $Z \subseteq X$ generating a partition $\mathcal{P}$ of X, such that $\bigcup_{Z\in\mathcal{P}} Z = X$. Consequently, the original problem is subdivided into subproblems

$$\min_{x\in Z\cap D}[F(x) = \mathbb{E}f(x,\theta(\omega))], \quad Z \in \mathcal{P}.$$

Let $F^*(Z\cap D)$ denote the optimal value of this subproblem. Clearly, the optimal value of the entire problem equals

$$F^* = F^*(X\cap D) = \min_{Z\in\mathcal{P}} F^*(Z\cap D).$$

The main idea of the stochastic branch and bound method is to iteratively execute three operations:

- partitioning into smaller subsets,
- stochastic estimation of the objective within the subsets,

- removal of some subsets.

In the basic algorithm the procedure continues infinitely, but at each iteration one has a probabilistic estimate of the accuracy of the current approximation to the solution.

Let us now describe in detail the concepts of partitioning and stochastic bounds.

Assume that partitioning is done by some deterministic rule Ψ: for every closed subset $Z \subseteq X$, $\Psi(Z)$ is a finite collection of closed subsets Z^i of Z such that $\bigcup_i Z^i = Z$. We consider the countable set $\mathcal{S}$ of subsets obtained from X by sequential application of the rule Ψ to all subsets arising in this process. The family $\mathcal{S}$ can be organized in a tree $T(X)$. The set X is the root node with assigned level 1, at level 2 there are nodes corresponding to the subsets of $\Psi(X)$, etc. For each set $Z \in \mathcal{S}$, we denote by $\text{lev}(Z)$ the location depth of Z in $T(X)$.

We make the following assumptions **A1-A4**.

A1. *For every sequence of sets* $Z_i \in \mathcal{S}$, $i = 1, 2, \ldots,$ *if* $\text{lev}(Z_i) \to \infty$ *then* $\text{diam}(Z_i) \to 0$.

A2. *There exist real-valued functions* L *and* U *defined od the collection of compact* $Z \subseteq X$ *with* $Z \cap D \neq \emptyset$, *such that for every* Z

$$L(Z) \leq F^*(Z \cap D) \leq U(Z),$$

and for every singleton $z \in X \cap D$

$$L(\{z\}) = U(\{z\}) = F(z).$$

A3. *There are random variables* $\xi_k(Z, \omega)$, $\eta_k(Z, \omega)$, $k = 1, 2, \ldots,$ *defined on some probability space* $(\Omega, \Sigma, \mathbb{P})$ *such that for all compact* $Z \subseteq X$ *with* $Z \cap D \neq \emptyset$ *and for every* k

$$\mathbb{E}\xi_k(Z, \omega) = L(Z),$$

$$\mathbb{E}\eta_k(Z, \omega) = U(Z),$$

and for all compact $Z, Z' \subseteq X$ *with* $Z \cap D \neq \emptyset$ *and* $Z' \cap D \neq \emptyset$,

$$|\xi_k(Z, \omega) - \xi_k(Z', \omega)| \leq \bar{K}(\omega) \cdot g(\text{dist}(Z, Z')),$$

$$|\eta_k(Z, \omega) - \eta_k(Z', \omega)| \leq \bar{K}(\omega) \cdot g(\text{dist}(Z, Z')),$$

where $g : \mathbb{R} \to \mathbb{R}$ *is a monotone function such that* $g(\epsilon) \downarrow 0$ *as* $\epsilon \downarrow 0$, *and* $\mathbb{E}\bar{K}(\omega) < \infty$.

Here $\text{dist}(Z, Z')$ denotes the Hausdorff distance between sets Z and Z':

$$\text{dist}(Z, Z') = \max\left(\sup_{z \in Z} \inf_{z' \in Z'} \|z - z'\|, \sup_{z' \in Z'} \inf_{z \in Z} \|z - z'\| \right).$$

A4. *There exists a selection mapping s which assigns to each compact $Z \subseteq X$ with $Z \cap D \neq \emptyset$ a point*

$$s(Z) \in Z \cap D$$

such that

$$F(s(Z)) \leq U(Z).$$

Remark 2.1. Notice that assumptions **A2, A3** imply that the deterministic functions $F(z)$, $L(Z)$ and $U(Z)$ also satisfy the generalized Lipschitz property:

$$|F(z) - F(z')| \leq K \cdot g(\|z - z'\|),$$

$$|L(Z) - L(Z')| \leq K \cdot g(\mathrm{dist}(Z, Z')),$$

$$|U(Z) - U(Z')| \leq K \cdot g(\mathrm{dist}(Z, Z')),$$

for all compact $Z, Z' \subseteq X$ with $Z \cap D \neq \emptyset$ and $Z' \cap D \neq \emptyset$, and all $z, z' \in X$, where $K = \mathbb{E}\bar{K}(\omega)$.

In Section 4 we shall discuss some ways of constructing a lower bound function L. Random functions $\tilde{L}(Z, \theta(\omega))$ will be constructed, such that

$$\mathbb{E}\tilde{L}(Z, \theta(\omega)) = L(Z).$$

Then $\xi_k(Z)$ satisfying **A3** can be taken in the form

$$\xi_k(Z) = \frac{1}{n_k} \sum_{i=1}^{n_k} \tilde{L}(Z, \theta^i), \quad n_k \geq 1, \tag{2.1}$$

where θ^i, $i = 1 \ldots, n_k$, are i.i.d. observations of $\theta(\omega)$.

It is also worth noting that if the random function $f(\cdot, \theta(\omega))$ and the mapping $s(\cdot)$ satisfy the generalized Lipschitz condition

$$|f(y, \theta(\omega)) - f(y', \theta(\omega))| \leq K(\theta(\omega)) \cdot g(\|y - y'\|) \quad \forall y, y' \in X, \tag{2.2}$$

and for all compact $Z, Z' \subseteq X$ having nonempty intersection with D,

$$\|s(Z) - s(Z')\| \leq M \cdot \mathrm{dist}(Z, Z'), \tag{2.3}$$

with some monotone g, such that $g(\epsilon) \downarrow 0$ as $\epsilon \downarrow 0$, some constant $M \in R^1_+$ and some square integrable $K(\theta(\omega))$, then the function

$$U(Z) = F(s(Z))$$

and the estimates

$$\eta_k(Z) = \frac{1}{n_k} \sum_{i=1}^{n_k} f(s(Z), \theta^i), \quad n_k \geq 1, \tag{2.4}$$

where θ^i, $i = 1, \ldots, n_k$, are i.i.d. observations of $\theta(\omega)$, satisfy **A2-A4**.

2.2 The algorithm

Let us now describe the stochastic branch and bound algorithm in more detail. For brevity, we skip the argument ω from random partitions and random sets.

Initialization. Form initial partition $\mathcal{P}_1 = \{X\}$. Observe independent random variables $\xi_1(X)$, $\eta_1(X)$ and put $L_1(X) = \xi_1(X)$, $U_1(X) = \eta_1(X)$. Set $k = 1$.

Before iteration k we have partition $\mathcal{P}_k$ and bound estimates $L_k(Z)$, $U_k(Z)$, $Z \in \mathcal{P}_k$.

Partitioning. Select the record subset

$$Y_k \in \operatorname{argmin}\{L_k(Z) : \ Z \in \mathcal{P}_k\}$$

and approximate solution $x^k = s(X_k) \in X_k \cap D$,

$$X_k \in \operatorname{argmin}\{U_k(Z) : \ Z \in \mathcal{P}_k\}.$$

Construct a partition of the record set, $\Psi(Y_k) = \{Y_k^i, \ i = 1, 2, \ldots\}$ such that $Y_k = \cup_i Y_k^i$. Define a new full partition

$$\mathcal{P}'_k = (\mathcal{P}_k \backslash Y_k) \cup \Psi(Y_k).$$

Deletion. Clean partition $\mathcal{P}'_k$ of non-feasible subsets, defining

$$\mathcal{P}_{k+1} = \mathcal{P}'_k \backslash \{Z \in \mathcal{P}'_k : \ Z \cap D = \emptyset\}.$$

Bound estimation. For all $Z \in \mathcal{P}_{k+1}$ observe random variables $\xi_{k+1}(Z)$, independently observe $\eta_{k+1}(Z)$ and recalculate stochastic estimates

$$L_{k+1}(Z) = (1 - \frac{1}{k+1}) L_k(\overline{Z}) + \frac{1}{k+1} \xi_{k+1}(Z),$$

$$U_{k+1}(Z) = (1 - \frac{1}{k+1}) U_k(\overline{Z}) + \frac{1}{k+1} \eta_{k+1}(Z),$$

where $\overline{Z}$ is such that $Z \subseteq \overline{Z} \in \mathcal{P}_k$.

Set $k := k + 1$ and go to **Partitioning**.

3 Convergence

Convergence of the stochastic branch and bound method requires some validation because of the probabilistic character of bound estimates.

3.1 Convergence a.s.

Let us introduce some notation. Remind that the partition tree $T(X)$ consists of a countable set $\mathcal{S}$ of subsets $Z \subseteq X$. For a fixed level l in $T(X)$ we define

$$\mathcal{S}_l = \{Z \in \mathcal{S} \mid \mathrm{lev}(Z) = l\},$$

and

$$d_l = \max_{Z \in \mathcal{S}_l} \mathrm{diam}(Z). \tag{3.1}$$

By condition **A1**,

$$\lim_{l \to \infty} d_l = 0.$$

For a given l and all $Z \in \mathcal{S}$ denote

$$\Pi^l(Z) = \begin{cases} Z, & \text{if } \mathrm{lev}(Z) < l, \\ S \in \mathcal{S}_l, & \text{if } \mathrm{lev}(Z) \geq l \text{ and } Z \subseteq S \in \mathcal{S}_l. \end{cases}$$

Correspondingly, introduce the projected partition

$$\overline{\mathcal{P}}^l_k(\omega) = \{\Pi^l(Z) \mid Z \in \mathcal{P}_k(\omega)\}.$$

Let us observe that after a finite number of iterations, say $\tau^l(\omega)$, the projected partition $\overline{\mathcal{P}}^l_k(\omega)$ becomes stable (does not change for $k > \tau^l(\omega)$). Indeed, partitioning an element of $\mathcal{P}_k$ whose level is larger or equal than l does not change the projected partition, and the elements located above level l can be partitioned only finitely many times.

For a given element $Z \in \mathcal{P}_k(\omega)$ we denote by $\{S_i(Z)\}$ the sequence of parental subsets $S_i(Z) \in \mathcal{P}_i(\omega),\ \ Z \subseteq S_i(Z),\ \ i = 1, \ldots, k$. Analogously, we denote by $\{\overline{S}^l_i(Z)\}$ the sequence of sets $\overline{S}_i(Z) \in \overline{\mathcal{P}}^l_i(\omega)$ such that $Z \subseteq \overline{S}_i(Z),\ \ i = 1, \ldots, k$. Thus for $Z \in \mathcal{P}_k(\omega)$

$$L_k(Z) = \frac{1}{k} \sum_{i=1}^{k} \xi_i(S_i(Z)),$$

$$U_k(Z) = \frac{1}{k} \sum_{i=1}^{k} \eta_i(S_i(Z)).$$

In the analysis we shall exclude from Ω some pathological subsets of measure zero, namely

$$\begin{aligned} \Omega_Z &= \{\omega \in \Omega \mid \lim_{k\to\infty} \tfrac{1}{k} \textstyle\sum_{i=1}^k \xi_i(Z) \neq L(Z)\}, \\ \Omega_{\xi_i} &= \{\omega \in \Omega \mid \xi_i(\check{x}) \text{ is unbounded}\}, \\ \Omega_{\eta_i} &= \{\omega \in \Omega \mid \eta_i(\check{x}) \text{ is unbounded}\}, \\ \Omega_{\bar{K}} &= \{\omega \in \Omega \mid \bar{K}(\omega) \text{ is unbounded}\}, \end{aligned}$$

($\check{x}$ is some fixed point in X), i.e. define

$$\Omega' = \Omega \setminus \left[\bigcup_{Z \in \mathcal{S}} \Omega_Z \cup \bigcup_i \Omega_{\xi_i} \cup \bigcup_i \Omega_{\eta_i} \cup \Omega_{\bar{K}} \right].$$

By the Law of Large Numbers $\mathbb{P}(\Omega_Z)=0$. By the integrability of ξ_i, η_i and $\bar{K}(\omega)$ we have $\mathbb{P}(\Omega_{\xi_i})=\mathbb{P}(\Omega_{\eta_i})=\mathbb{P}(\Omega_{\bar{K}})=0$. Due to the countable number of exluded set of measure zero we have

$$\mathbb{P}(\Omega')=1.$$

Lemma 3.1. *Let $Z_k(\omega)\in\mathcal{P}_k(\omega)$, $k=1,2\dots$. Then for all$\omega\in\Omega'$*
(i) $\lim_{k\to\infty}|L(Z_k(\omega))-L_k(Z_k(\omega))|=0$;
(ii) $\lim_{k\to\infty}|U(Z_k(\omega))-U_k(Z_k(\omega))|=0$.

Proof. Let us fix some $l>0$. We have

$$\begin{aligned}|L(Z_k)-L_k(Z_k)| &= \left|L(Z_k)-\frac{1}{k}\sum_{i=1}^{k}\xi_i(S_i(Z_k))\right| \\ &\le \left|L(\Pi^l(Z_k))-\frac{1}{k}\sum_{i=1}^{k}\xi_i(\overline{S}_i^l(Z_k))\right|+\left|L(Z_k)-L(\Pi^l(Z_k))\right| \\ &\quad+\left|\frac{1}{k}\sum_{i=1}^{k}(\xi_i(\overline{S}_i^l(Z_k))-\xi_i(S_i(Z_k)))\right|. \end{aligned}\tag{3.2}$$

We shall estimate the components on the right hand side of (3.2). By Remark 2.1,

$$|L(Z_k)-L(\Pi^l(Z_k))|\le Kg(\mathrm{dist}(Z_k,\Pi^l(Z_k))\le Kg(d_l),\tag{3.3}$$

where d_l is given by (3.1). Similarly, by assumption **A3**, for the third component of (3.2) one has

$$\begin{aligned}\left|\frac{1}{k}\sum_{i=1}^{k}(\xi_i(\overline{S}_i^l(Z_k))-\xi_i(S_i(Z_k)))\right| &\le \frac{1}{k}\sum_{i=1}^{k}\bar{K}(\omega)g(\mathrm{dist}(\overline{S}_i^l(Z_k),S_i(Z_k))) \\ &\le g(d_l)\bar{K}(\omega).\end{aligned}\tag{3.4}$$

The first component at the right side of (3.2) for $k>\tau^l(\omega)$ can be estimated as follows:

$$\begin{aligned}&\left|L(\Pi^l(Z_k))-\frac{1}{k}\sum_{i=1}^{k}\xi_i(\overline{S}_i^l(Z_k))\right| \\ &\le \left|L(\Pi^l(Z_k))-\frac{1}{k}\sum_{i=\tau^l(\omega)}^{k}\xi_i(\overline{S}_i^l(Z_k))\right|+\left|\frac{1}{k}\sum_{i=1}^{\tau^l(\omega)-1}\xi_i(\overline{S}_i^l(Z_k))\right| \\ &\le \left|L(\Pi^l(Z_k))-\frac{1}{k}\sum_{i=\tau^l(\omega)}^{k}\xi_i(\overline{S}_i^l(Z_k))\right|+\left|\frac{1}{k}\sum_{i=1}^{\tau^l(\omega)-1}[\xi_i(\{\check{x}\})+\bar{K}(\omega)g(\mathrm{diam}(X))]\right|,\end{aligned}\tag{3.5}$$

where $\check{x}$ is the fixed point in X appearing in the definition of Ω'.

Since $Z_k\in\mathcal{P}_k$, by the definition of $\tau^l(\omega)$ the set $\Pi^l(Z_k)$ is an element of $\overline{\mathcal{P}}_i^l$ for all $i\in[\tau(\omega),k]$. Thus $S_i(Z_k)\subseteq\Pi^l(Z_k)$ for these i. Consequently,

$$\Pi^l(Z_k)=\overline{S}_i^l(Z_k),\ i\in[\tau^l(\omega),k].$$

For all $\omega \in \Omega'$ and all k the sets $\Pi^l(Z_k(\omega))$ can take values only from the finite family $\{Z \in \mathcal{S} : \operatorname{lev}(Z) \le l\}$. Therefore

$$\left| L(\Pi^l(Z_k)) - \frac{1}{k} \sum_{i=\tau^l(\omega)}^{k} \xi_i(\overline{S}_i^l(Z_k)) \right| \le \max_{\operatorname{lev}(Z)\le l} \left| L(Z) - \frac{1}{k} \sum_{i=\tau^l(\omega)}^{k} \xi_i(Z) \right|.$$

Substituting the above inequality into (3.5) and passing to the limit with $k \to \infty$ we obtain

$$\lim_{k\to\infty} \left| L(\Pi^l(Z_k(\omega))) - \frac{1}{k} \sum_{i=1}^{k} \xi_i(\overline{S}_i^l(Z_k(\omega))) \right| = 0, \ \omega \in \Omega'. \tag{3.6}$$

Using (3.3), (3.4) and (3.6) in (3.2) we conclude that with probability 1

$$\lim_{k\to\infty} |L(Z_k) - L_k(Z_k)| \le (K + \bar{K}) g(d_l).$$

Since l was arbitrary, by **A1** and **A3** we can make $g(d_l)$ arbitrarily small, which proves assertion (i). The proof of (ii) is identical. □

Theorem 3.1. *Assume* **A1**-**A4**. *Then*
(i) $\lim_{k\to\infty} L_k(Y_k(\omega)) = F^*$ *a.s.,*
(ii) *all cluster points of the sequence* $\{Y_k(\omega)\}$ *belong to the solution set* X^* *a.s., i.e.*

$$\limsup_{k\to\infty} Y_k(\omega) = \{y : \exists k_n \to \infty,\ y_{k_n} \in Y_{k_n},\ y_{k_n} \to y\} \subseteq X^* \quad a.s.,$$

(iii) $\lim_{k\to\infty} U_k(X_k(\omega)) = F^*$ *a.s.,*
(iv) *all accumulation points of the sequence of approximate solutions* $\{x_k(\omega)\}$ *belong to* X^* *a.s..*

Proof. Let $x^* \in X^*$. Let us fix $\omega \in \Omega'$ and choose a sequence of sets $Z_k(\omega) \in \mathcal{P}_k(\omega)$ in such a way that $x^* \in Z_k(\omega)$ for $k = 1, 2 \ldots$. By construction

$$L(Z_k) \le F^*,\ k = 1, 2, \ldots.$$

By Lemma 3.1,

$$\limsup_{k\to\infty} L_k(Z_k) = \limsup_{k\to\infty} L(Z_k) \le F^*.$$

By the definition of the record set, $L_k(Y_k) \le L_k(Z_k)$, so

$$\limsup_{k\to\infty} L_k(Y_k) \le F^*.$$

Using Lemma 3.1 again we see that

$$\limsup_{k\to\infty} L(Y_k) \le F^*. \tag{3.7}$$

Assume that there exists a subsequence $\{Y_{k_j}(\omega)\}$, where $k_j \to \infty$ as $j \to \infty$, such that $\lim L(Y_{k_j}(\omega)) < F^*$. With no loss of generality we can also assume that there is a subsequence $y^{k_j}(\omega) \in Y_{k_j}(\omega)$ convergent to some $y^\infty(\omega)$. Since $\operatorname{diam}(Y_k) \to 0$,

$$\lim_{j\to\infty} \operatorname{dist}(Y_{k_j}(\omega), \{y^\infty(\omega)\}) = 0. \tag{3.8}$$

Thus, with a view to Remark 2.1 and A2,

$$\lim_{j\to\infty} L(Y_{k_j}(\omega)) = L(\{y^\infty(\omega)\}) = F(y^\infty(\omega)) \geq F^*, \tag{3.9}$$

a contradiction. Therefore

$$\liminf_{k\to\infty} L(Y_k) \geq F^*. \tag{3.10}$$

Combining (3.7) with (3.10) yields

$$\lim_{k\to\infty} L(Y_k) = F^*. \tag{3.11}$$

Using Lemma 3.1(i) we obtain assertion (i).

Since for every $y^\infty(\omega) \in \limsup Y_k(\omega)$ one can find a subsequence $\{Y_{k_j}\}$ satisfying (3.8), from (3.9) and (3.11) we get

$$F(y^\infty(\omega)) = L(\{y^\infty(\omega)\}) = \lim_{j\to\infty} L(Y_{k_j}(\omega)) = F^*,$$

which proves assertion (ii).

Let us consider the sequence $\{X_k(\omega)\}$. By construction,

$$U_k(X_k(\omega)) \leq U_k(Y_k(\omega)).$$

Proceeding similarly to the proof of (3.11) one obtains

$$\lim_{k\to\infty} U(Y_k) = F^*.$$

Combining the last two relations and using Lemma 3.1(ii) we conclude that

$$\limsup_{k\to\infty} U_k(X_k) \leq F^*.$$

On the other hand, by Lemma 3.1

$$\begin{aligned}\liminf_{k\to\infty} U_k(X_k) = \liminf_{k\to\infty} U(X_k) &\geq \liminf_{k\to\infty} L(X_k) \\ &= \liminf_{k\to\infty} L_k(X_k) \geq \liminf_{k\to\infty} L_k(Y_k) = F^*,\end{aligned}$$

where in the last equality we used (i). Consequently,

$$\lim_{k\to\infty} U_k(X_k) = \lim_{k\to\infty} U(X_k) = F^*,$$

i.e. assertion (iii) holds. Finally,

$$F(x^k) \leq U(X_k) \to F^*,$$

which proves (iv). □

3.2 Accuracy estimates

Probabilistic accuracy estimates of current approximations $s(X_k)$ can be obtained under the following additional assumptions.

A5. *For each $Z \in \mathcal{S}$ the random variables $\xi_k(Z)$, $\eta_k(Z)$, $Z \in \mathcal{S}$, $k = 0, 1, \ldots,$ are independent and normally distributed with means $L(Z)$, $U(Z)$ and variances $\mu_k(Z)$, $\nu_k(Z)$ correspondingly.*

A6. *For the variances $\mu_k(Z)$, $\nu_k(Z)$, $Z \in \mathcal{S}$, some upper bounds are known:*

$$\mu_k(Z) \le \overline{\mu}_k(Z), \quad \nu_k(Z) \le \overline{\nu}_k(Z).$$

Remark 3.1. In section 4 we outline some methods to construct (in general not normal) random estimates $\tilde{L}(Z, \theta(\omega))$, $\tilde{U}(Z, \theta(\omega))$ (with expected values $L(Z)$, $U(Z)$) satisfying the Lipschitz condition. Suppose that the variances of $\tilde{L}(Z, \theta(\omega))$, $\tilde{U}(Z, \theta(\omega))$ are bounded by some quantities $\overline{\nu}(Z)$ and $\overline{\mu}(Z)$, respectively (in practice we can use empirical estimates for variances). Now (approximately) normal estimates $\xi_k(Z)$, $\eta_k(Z)$ of $L(Z)$, $U(Z)$ can be obtained by averaging several (say n_k) independent observations of $\tilde{L}(Z, \theta(\omega))$, $\tilde{U}(Z, \theta(\omega))$:

$$\xi_k(Z) = \frac{1}{n_k} \sum_{i=1}^{n_k} \tilde{L}(Z, \theta^i),$$

$$\eta_k(Z) = \frac{1}{n_k} \sum_{j=1}^{n_k} \tilde{U}(Z, \theta^j),$$

where θ^i and θ^j are i.i.d. observations of $\theta(\omega)$.

Additionally, we obtain tighter bounds for variances of $\xi_k(Z)$, $\eta_k(Z)$:

$$\mathbb{D}\xi_k(Z) \le \frac{1}{\sqrt{n_k}} \overline{\mu}(Z) = \overline{\mu}_k(Z),$$

$$\mathbb{D}\eta_k(Z) \le \frac{1}{\sqrt{n_k}} \overline{\nu}(Z) = \overline{\nu}_k(Z).$$

□

Let us take confidence bounds for $L(Z)$, $U(Z)$ in the form:

$$\underline{\xi}_k(Z) = \xi_k(Z) - c_k \overline{\mu}_k(Z),$$

and

$$\overline{\eta}_k(Z) = \eta_k(Z) + c_k \overline{\nu}_k(Z),$$

where constants c_k, $k = 0, 1, \ldots,$ are such that

$$\Phi(c_k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c_k} e^{-\tau^2/2} d\tau = 1 - \delta_k, \quad 0 < \delta_k < 1.$$

Lemma 3.2. *Under assumptions* **A5** *and* **A6** *for each k the following random accuracy estimate is true:*

$$\mathbb{P}\left\{F(s(X_k)) - F^* \le \overline{\eta}_k(X_k) - \min_{Z\in\mathcal{P}_k} \underline{\xi}_k(Z)\right\} \ge 1 - \delta_k.$$

Proof. Let

$$Z_k = \arg\min_{Z\in\mathcal{P}_k} L(Z),$$

$$\mathring{\xi}_k(Z_k) = \xi_k(Z_k) - L(Z_k),$$

$$\mathring{\eta}_k(X_k) = \eta_k(X_k) - U(X_k).$$

The following chain of estimates holds:

$$\begin{aligned}
&\mathbb{P}\left\{F(s(X_k)) - F^* \le \overline{\eta}_k(X_k) - \min_{Z\in\mathcal{P}_k} \underline{\xi}_k(Z)\right\}\\
&\quad \ge \mathbb{P}\left\{\overline{\eta}_k(X_k) - U(X_k) + F^* - \underline{\xi}_k(Z_k) \ge 0\right\}\\
&\quad = \mathbb{P}\left\{\eta_k(X_k) - U(X_k) + F^* - \xi_k(Z_k) + c_k(\overline{\mu}_k(Z_k) + \overline{\nu}_k(X_k)) \ge 0\right\}\\
&\quad \ge \mathbb{P}\left\{\eta_k(X_k) - U(X_k) + F^* - \xi_k(Z_k) + c_k\sqrt{\mu_k^2(Z_k) + \nu_k^2(X_k)} \ge 0\right\}\\
&\quad = \mathbb{E}\,\mathbb{P}\left\{\frac{\mathring{\xi}_k(Z_k) - \mathring{\eta}_k(X_k)}{\sqrt{\mu_k^2(Z_k) + \nu_k^2(X_k)}} \le c_k + \frac{F^* - L(Z_k)}{\sqrt{\mu_k^2(Z_k) + \nu_k^2(X_k)}} \;\middle|\; \mathcal{P}_k\right\}\\
&\quad = \mathbb{E}\,\Phi\left(c_k + \frac{F^* - L(Z_k)}{\sqrt{\mu_k^2(Z_k) + \nu_k^2(X_k)}}\right) \;\ge\; \Phi(c_k) \;=\; 1 - \delta_k,
\end{aligned}$$

where in the last inequality we used the fact that $F^* \ge L(Z_k)$. □

3.3 Deletion rule

An important feature of the branch and bound method is the possibility to delete non-prospective subsets from the partition by using current lower and upper bounds of the optimal value within the subsets. In the stochastic case, however, because of the randomness of the bounds, deletion may lead to the loss of the optimal solution. Particular caution is needed when deleting sets with 'poor' lower bounds. In the following deletion rule we do not delete subsets at each iteration, but only after carrying out a sufficiently large number N of iterations, and after deriving an independent estimate of the objective value at the current approximate solution.

We make the following additional assumption.

A7. *A uniform bound σ^2 is known for the variances of all random variables $\xi_k(Z,\omega)$ and $\eta_k(Z,\omega)$, $Z \in \mathcal{S}$, $k = 1, 2, \ldots$.*

Deletion rule. After N steps we stop, take the subset $X_N(\omega)$ from the final partition $\mathcal{P}_N(\omega)$ and make N independent observations $\eta_{Ni}(X_N(\omega))$, $i = 1, \ldots, N$, calculating a new estimate for $U(X_N(\omega))$:

$$\overline{U}_N(X_N(\omega)) = \frac{1}{N}\sum_{i=1}^{N} \eta_{Ni}(X_N(\omega)).$$

Then, for some accuracy $\epsilon \in (0,1)$, we delete all sets $Z \in \mathcal{P}_N(\omega)$ such that

$$L_N(Z) > \overline{U}_N(X_N(\omega)) + 2c_N,$$

where $c_N^2 = \sigma^2/(N\epsilon)$.

Lemma 3.3. *Let x^* be a solution of (1.1). Then*

$$\mathbf{P}\{x^* \text{ is lost at the final deletion}\} \le 2\epsilon.$$

Proof. Let $Z_i(\omega) \in \mathcal{P}_i(\omega)$ be such that $Z_1(\omega) \supseteq Z_2(\omega) \supseteq \ldots \supseteq \{x^*\}$. Then

$$\begin{aligned}
&\mathbb{P}\{x^* \text{ is lost}\} \\
&\quad = \mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N}\xi_i(Z_i(\omega)) > \frac{1}{N}\sum_{i=1}^{N}\eta_{Ni}(X_N(\omega)) + 2c_N\right\} \\
&\quad = \mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N}\xi_i(Z_i(\omega)) - F^* - c_N > \frac{1}{N}\sum_{i=1}^{N}\eta_{Ni}(X_N(\omega)) - F^* + c_N\right\} \\
&\quad \le \mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N}\xi_i(Z_i(\omega)) - F^* - c_N > 0 \ \text{ or } \ \frac{1}{N}\sum_{i=1}^{N}\eta_{Ni}(X_N(\omega)) - F^* + c_N < 0\right\} \\
&\quad \le P_1 + P_2,
\end{aligned}$$

where

$$P_1 = \mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N}\xi_i(Z_i(\omega)) > F^* + c_N\right\},$$

$$P_2 = \mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N}\eta_{Ni}(X_N(\omega)) < F^* - c_N\right\}.$$

Denote by $\mathcal{F}_i$ the σ-algebra generated by the sequence of observations made up to iteration i and by $\mathbb{E}_i\xi_i(Z_i(\omega)) = \mathbb{E}\{\xi_i(Z_i(\omega)) \mid \mathcal{F}_i\}$ the conditional expectation of $\xi_i(Z_i(\omega))$ with respect to $\mathcal{F}_i$. Observe that

$$\mathbb{E}_i\xi_i(Z_i(\omega)) = L(Z_i(\omega)) \le F^*. \tag{3.12}$$

Denote by $\mathbb{D}\xi_i(Z_i(\omega))$ the variance of the random variable $\xi_i(Z_i(\omega))$. By Assumption **A7** for all $i \le N$

$$\mathbb{D}\xi_i(Z_i(\omega)) \le \sigma^2.$$

Then, by (3.12) and Chebyshev inequality,

$$
\begin{aligned}
P_1 &= \mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N}[\xi_i(Z_i(\omega)) - \mathbb{E}_i\xi_i(Z_i(\omega))] > c_N + F^* - \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_i\xi_i(Z_i(\omega))\right\} \\
&\le \mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N}[\xi_i(Z_i(\omega)) - \mathbb{E}_i\xi_i(Z_i(\omega))] > c_N\right\} \\
&\le \mathbb{P}\left\{\frac{1}{N}\left|\sum_{i=1}^{N}[\xi_i(Z_i(\omega)) - \mathbb{E}_i\xi_i(Z_i(\omega))]\right| > c_N\right\} \\
&\le \frac{1}{N^2 c_N^2}\sum_{i=1}^{N}\mathbb{D}\xi_i(Z_i(\omega)) \;\le\; \sigma^2/(N c_N^2) = \epsilon.
\end{aligned}
$$

The estimation of P_2 is similar. $\square$

4 Stochastic bounds

In the branch and bound method the original problem (1.1) is subdivided into subproblems of the form

$$\min_{x\in Z\cap D}[F(x) = \mathbb{E}f(x,\theta(\omega))], \tag{4.1}$$

where Z is some compact subset of X, and D represents additional deterministic constraints. We denote by $F^*(Z\cap D)$ the optimal value of this subproblem. In the method stochastic lower estimates $\xi_k(Z)$ and upper estimates $\eta_k(Z)$ of $F^*(Z\cap D)$ are used for branching, deleting non-prospective sets, and for estimating the accuracy of the current approximation.

As an upper deterministic estimate of $F^*(Z\cap D)$ one can always use the value of the objective function $F(s(Z)) = \mathbb{E}f(s(Z),\theta(\omega))$ at some feasible point $s(Z)\in Z\cap D$ (the function $f(\cdot,\theta)$ and the mapping $s(\cdot)$ are assumed Lipschitz continuous in the sense of (2.2), (2.3)). Then the function $U(Z) = F(s(Z))$ and its Monte-Carlo estimates (2.4) satisfy conditions **A2, A3**.

Construction of stochastic lower bounds is more difficult. We shall discuss here some ideas that exploit the stochastic nature of the problem. It should be stressed, however, that together with them deterministic bounding techniques known in deterministic global optimization can be used (such as relaxation of integrality constraints, dual estimates, tangent minorants, bounds using monotonicity, etc., see [5]).

4.1 Interchange relaxation

Interchanging the minimization and expectation operators in (4.1) we obtain the following estimate true:

$$F^*(Z\cap D) = \min_{x\in Z\cap D}\mathbb{E}f(x,\theta(\omega)) \ge \mathbb{E}\min_{x\in Z\cap D}f(x,\theta(\omega)).$$

Thus the quantity

$$L(Z) = \mathbb{E}\min_{x\in Z\cap D}f(x,\theta(\omega)) \tag{4.2}$$

is a valid deterministic lower bound for the optimal value $F^*(Z \cap D)$. A stochastic lower bound can be obtained by Monte-Carlo simulation: for i.i.d. observations θ^i of θ, $i = 1, \ldots, N$, one defines

$$\xi(Z) = \frac{1}{N} \sum_{i=1}^{N} \min_{x \in Z \cap D} f(x, \theta^i). \tag{4.3}$$

In many cases, for a fixed θ^i, it is easy to solve the minimization problems at the right hand side of (4.3). In particular, if the function $f(\cdot, \theta)$ is quasi-convex, stochastic lower bounds can be obtained by convex optimization methods.

Example 4.1. Consider the facility location problem of Example 1.1. The application of (4.3) yields the following stochastic lower bound:

$$\begin{aligned}\xi(Z) = \frac{1}{N} \sum_{i=1}^{N} \min_{x \in Z \cap D} f(x^1, \ldots, x^n, \theta^i) \; &= \; \frac{1}{N} \sum_{i=1}^{N} \min_{x \in Z \cap D} \min_{1 \le j \le n} \varphi(x^j, \theta^i) \\ = \frac{1}{N} \sum_{i=1}^{N} \min_{1 \le j \le n} \min_{x \in Z \cap D} \varphi(x^j, \theta^i). & \end{aligned} \tag{4.4}$$

If $\varphi(\cdot, \theta)$ is quasi-convex and $Z \cap D$ is convex, the inner minimization problem can be solved by convex programming methods, or even in a closed form (if $Z \cap D$ has a simple structure). The minimum over j can be calculated by enumeration, so the whole evaluation of the stochastic lower bound is relatively easy.

Example 4.2. For the fixed mix problem of Example 1.2 the application of (4.2) (with obvious changes reflecting the fact that we deal with a maximization problem) yields

$$\begin{aligned} F^*(Z \cap D) \;&\le\; \mathbb{E} \max_{x \in Z \cap D} \left\{ \prod_{t=1}^{T} \sum_{j=1}^{n} (x_j \theta_j(t) - \alpha_j |x_j \theta_j(t) - x_j \textstyle\sum_{\nu=1}^{n} x_\nu \theta_\nu(t)|) \right\} \\ &\le \mathbb{E} \left\{ \max_{\substack{x(t) \in Z \cap D, \\ t=1,\ldots,T+1}} \prod_{t=1}^{T} \sum_{j=1}^{n} (x_j(t) \theta_j(t) - \alpha_j |x_j(t) \theta_j(t) - x_j(t+1) \textstyle\sum_{\nu=1}^{n} x_\nu(t) \theta_\nu(t)|) \right\}, \end{aligned} \tag{4.5}$$

where in the last inequality we additionally split the decision vector x into $x(t)$, $t = 1, \ldots, T+1$. Let

$$Z \cap D \subseteq \{x \in \mathbb{R}^n : \; a_j \le x_j \le b_j, \; j = 1, \ldots, n\}.$$

Then the optimal value of the optimization problem inside (4.5) can be estimated as follows. Denote by $w_j(t)$ the wealth at the beginning of period t in assets j, and by $p_j(t)$ and $s_j(t)$ the amounts of money spent on purchases and obtained from sales in category j after period t. The optimal value of our problem can be calculated by solving the linear program:

$$\begin{aligned} &\max W(T+1) \\ &W(t) = \sum_{j=1}^{n} w_j(t), \; t = 1, \ldots, T+1, \end{aligned}$$

$$
\begin{aligned}
&a_j W(t) \le w_j(t) \le b_j W(t),\ j = 1,\dots,n,\ t = 1,\dots,T+1,\\
&w_j(t+1) = \theta_j(t) w_j(t) + \frac{p_j(t)}{1+\alpha_j} - \frac{s_j(t)}{1-\alpha_j},\ j = 1,\dots,n,\ t = 1,\dots,T,\\
&\sum_{j=1}^{n} p_j(t) = \sum_{j=1}^{n} s_j(t),\ t = 1,\dots,T,\\
&p_j(t) \ge 0,\ s_j(t) \ge 0,\ j = 1,\dots,n,\ t = 1,\dots,T,
\end{aligned}
$$

where $W(1) = 1$. In the above problem the last equation expresses the balance of cash involved in sales and purchases.

Denoting the optimal value of this problem by $f^*(Z,\theta)$ we arrive at the stochastic upper bound

$$
\zeta(Z) = \frac{1}{N} \sum_{i=1}^{N} f^*(Z, \theta^i),
$$

where θ^i are i.i.d. observations of the sequence $\theta(t)$, $t = 1,\dots,T$. As a result, stochastic bounds can be obtained by simulation and linear programming.

4.2 Using multiple observations

The simplest way to improve the lower bound (4.2) and its Monte-Carlo estimate would be to use M independent copies θ^l of θ to obtain

$$
F^*(Z \cap D) = \frac{1}{M} \min_{x \in Z \cap D} \sum_{l=1}^{M} \mathbb{E} f(x, \theta^l(\omega)) \ge \mathbb{E} \min_{x \in Z \cap D} \left[\frac{1}{M} \sum_{l=1}^{M} f(x, \theta^l(\omega)) \right] = L_M(Z). \quad (4.6)
$$

This leads to the following Monte-Carlo estimates

$$
\xi_{MN}(Z) = \frac{1}{N} \sum_{i=1}^{N} \min_{x \in Z \cap D} \left[\frac{1}{M} \sum_{l=1}^{M} f(x, \theta^{il}) \right], \quad (4.7)
$$

where θ^{il} are i.i.d. observations of θ, $i = 1,\dots,N$, $l = 1,\dots,M$. In other words, instead of solving (4.1) one minimizes the empirical estimates of the expected value function:

$$
F_M(x) = \frac{1}{M} \sum_{l=1}^{M} f(x, \theta^{il}). \quad (4.8)
$$

Obviously, $\mathbb{E}\xi_{MN}(Z) = L_M(Z)$ for all $M, N \ge 1$. Moreover, by increasing the number of observations M inside the minimization operation, the accuracy of the empirical estimate (4.7) can be made arbitrarily high.

Lemma 4.1. *For all Z such that $Z \cap D \ne \emptyset$, one has*
(i) *$L_M(Z) \uparrow F^*(Z \cap D)$, as $M \to \infty$,*
(ii) *for all $N \ge 1$ with probability one, $\xi_{MN}(Z) \to F^*(Z \cap D)$, as $M \to \infty$.*

Proof. To prove the monotonicity of the sequence $\{L_M(Z)\}$, note that

$$
\begin{aligned}
L_M(Z) &= \frac{1}{M+1}\sum_{j=1}^{M+1} \mathbb{E} \min_{x\in Z\cap D}\left[\frac{1}{M}\sum_{\substack{l=1\\ l\neq j}}^{M+1} f(x,\theta^l(\omega))\right] \\
&= \mathbb{E}\frac{1}{M+1}\sum_{j=1}^{M+1} \min_{x\in Z\cap D}\left[\frac{1}{M}\sum_{\substack{l=1\\ l\neq j}}^{M+1} f(x,\theta^l(\omega))\right] \\
&\leq \mathbb{E} \min_{x\in Z\cap D}\left[\frac{1}{M(M+1)}\sum_{j=1}^{M+1}\sum_{\substack{l=1\\ l\neq j}}^{M+1} f(x,\theta^l(\omega))\right] \\
&= \mathbb{E} \min_{x\in Z\cap D}\left[\frac{1}{M+1}\sum_{l=1}^{M+1} f(x,\theta^l(\omega))\right] \; = \; L_{M+1}(Z).
\end{aligned}
$$

Since $f(\cdot,\theta)$ is continuous and bounded by an integrable function, by Lemma A1 of Rubinstein-Shapiro, with probability 1,

$$
\lim_{M\to\infty}\frac{1}{M}\sum_{l=1}^{M} f(x,\theta^l) = \mathbb{E}f(x,\theta(\omega)),
$$

uniformly for $x\in X$. Taking the minimum of both sides with respect to $x \in Z\cap D$ one obtains (ii). Assertion (i) follows then from the Lebesque theorem. $\square$

However, the optimization problems appearing in (4.7) can be very difficult for non-convex $f(\cdot,\theta)$, and this is exactly the case of interest for us. Still, in some cases (like integer programming problems) these subproblems may prove tractable, or some deterministic lower bounds may be derived for them.

Using empirical estimates (4.8) is not the only possible way to use multiple observations. Let again θ^l, $l = 1,\dots,M$, be independent copies of θ and let $f(x,\theta)\geq 0$ for all x and θ. Let us assume (for reasons that will become clear soon) that the original problem is a maximization problem. Then the problem of maximizing $\mathbb{E}f(x,\theta(\omega))$ is equivalent to maximizing the function

$$
(F(x))^M = (\mathbb{E}f(x,\theta(\omega)))^M = \prod_{l=1}^{M}\mathbb{E}f(x,\theta^l(\omega)) = \mathbb{E}\left\{\prod_{l=1}^{M} f(x,\theta^l(\omega))\right\},
$$

where in the last equation we used the independence of θ^l, $l = 1,\dots,M$. Interchange of the maximization and expectation operators leads to the following stochastic bound

$$
\left(\max_{x\in Z\cap D}(F(x))\right)^M \; = \; \max_{x\in Z\cap D}(F(x))^M \leq \mathbb{E}\zeta_{MN}(Z),
$$

where

$$
\zeta_{MN}(Z) = \frac{1}{N}\sum_{i=1}^{N}\max_{x\in Z\cap D}\prod_{l=1}^{M} f(x,\theta^{il}), \tag{4.9}
$$

where θ^{il} are i.i.d. observations of θ, $i = 1, \ldots, N$, $l = 1, \ldots, M$. Products of some quasi-concave functions may be still quasi-concave, so the optimization problem inside (4.9) may be easier to solve than (4.7). In particular, if $\ln f(\cdot, \theta)$ is concave, the problem in (4.9) is equivalent to the convex optimization problem

$$\max_{x \in Z \cap D} \sum_{l=1}^{M} \ln f(x, \theta^{il}).$$

There is a broad class of so-called log-concave or α-concave functions for which similar transformations can be made (see [9, 13]).

Example 4.3. Consider the facility location problem of Example 1.1 again, but now with $n = 1$ and φ given by (1.2), and assume that the partition elements Z are convex. For $f(x, \theta) = 1 - \varphi(x, \theta)$ the problem is equivalent to maximizing the function

$$(F(x))^M = \mathbb{E}\left\{\prod_{l=1}^{M} f(x, \theta^l(\omega))\right\},$$

i.e. maximizing the probability of the event

$$A_M(x) = \left\{\omega \in \Omega : \ d(x, \theta^l(\omega)) \le \delta, \ l = 1, \ldots, M\right\}.$$

A stochastic upper bound for $\mathbb{P}\{A_M(x)\}$ over $x \in Z \cap D$ can be obtained as follows. We generate i.i.d. observations θ^{il}, $i = 1, \ldots, N$, $l = 1, \ldots, M$, and use (4.9) to get the stochastic bound

$$\zeta(Z) = \frac{1}{N} \sum_{i=1}^{N} \chi(Z, \theta^{i1}, \ldots, \theta^{iM}), \tag{4.10}$$

where

$$\chi(Z, \theta^{i1}, \ldots, \theta^{iM}) = \max_{x \in Z \cap D} \prod_{l=1}^{M} f(x, \theta^{il}).$$

Denoting

$$Y(Z, \theta^{i1}, \ldots, \theta^{iM}) = \{x \in Z \cap D : \ d(x, \theta^l) \le \delta, \ l = 1, \ldots, M\}.$$

we see that

$$\chi(Z, \theta^{i1}, \ldots, \theta^{iM}) = \begin{cases} 1 & \text{if } Y(Z, \theta^{i1}, \ldots, \theta^{iM}) \ne \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Since $Y(Z, \theta^{i1}, \ldots, \theta^{iM})$ is convex, calculation of the value of χ and of the corresponding upper bound (4.10) on the probability of M successes can be carried out by convex programming methods. In the special case of $M = 1$ we obtain the basic bound (4.3). Its calculation requires only veryfying for each observation θ^i whether $Z \cap D \cap \{x : \ d(x, \theta^i) \le \delta\} \ne \emptyset$.

4.3 Stochastic correction terms

Let $\lambda(x,\theta)$ be a real-valued function, which is continuous in x and measurable in θ. If $\mathbb{E}\lambda(x,\theta(\omega)) = 0$ for all $x \in Z \cap D$, the following relations hold:

$$F^*(Z \cap D) = \min_{x \in Z \cap D} \mathbb{E} f(x,\theta(\omega)) = \min_{x \in Z \cap D} \mathbb{E}[f(x,\theta(\omega)) + \lambda(x,\theta(\omega))]$$
$$\geq \mathbb{E} \min_{x \in Z \cap D} [f(x,\theta(\omega)) + \lambda(x,\theta(\omega))] = L_\lambda(Z). \quad (4.11)$$

This can be used to construct stochastic lower bounds by Monte-Carlo estimates of the right hand side of (4.11), exactly as in (4.3), (4.7) and (4.9), just the objective function is modified by the correction term $\lambda(x,\theta)$.

Of course, the quality of the resulting bounds depends on the choice of the function $\lambda(x,\theta)$, which may be interpreted as the value of information. It is theoretically possible to select the correction term in such a way that $L_\lambda(Z) = F^*(Z \cap D)$. Indeed, setting

$$\lambda(x,\theta(\omega)) = \mathbb{E} f(x,\theta(\omega)) - f(x,\theta(\omega))$$

one obtains equality in (4.11), but it has no practical meaning. Still, in some applications it may be possible to guess a good correction term.

5 Application to facility location

5.1 The problem

As an illlustrative example we use a particular version of the facility location problem (Examples 1.1 and 4.1) on the line. We assume that the cost of serving a customer located at $\theta \in [0,1]^k$ from facility at $x \in [0,1]^k$ equals to

$$\varphi(x,\theta) = \frac{\|x-\theta\|^\alpha}{\gamma + \|x-\theta\|^\beta}, \quad (5.1)$$

where $\alpha \geq \beta \geq 0$ and $\gamma > 0$. If there are n facilities located at $x = (x^1, \ldots, x^n)$, the cost to serve customer θ is

$$f(x,\theta) = \min_{1 \leq j \leq n} \varphi(x^j,\theta).$$

The objective is to place n facilities in $[0,1]$ in such a way that the expected cost

$$F(x) = \mathbb{E} f(x,\theta)$$

is minimized.

If the customer distribution $\mathbb{P}$ sits on finitely many points, the problem is called in [3] the *multisource Weber problem*. In [3] a special branch and bound method is considered and in the case of $\alpha = 1$ and $\beta = 0$, the polyhedral annexation algorithm is discussed.

It is also worth mentioning that the methods of so-called d.-c. optimization are applicable to our case, too, because ieach φ can be expressed as a difference of two convex functions:

$$\varphi(x,\theta) = \frac{1}{\gamma}\|x-\theta\|^\alpha - \frac{\|x-\theta\|^{\alpha+\beta}}{\gamma(\gamma + \|x-\theta\|^\beta)},$$

We shall not exploit this property. Instead, we shall use the quasi-concavity of each distance function φ.

In the case of finitely many customers $\theta_1, \ldots, \theta_m$ located on the line, and with $0 \le \beta \le \alpha \le 1$, it is sufficient to consider facilities located at customers' points θ_s, $s = 1, \ldots, m$. For one facility ($x \in \mathbb{R}^1$) it is obvious, since the function $F(x)$ is concave between two neighboring customers and hence achieves its minimum at one of the ends. The same can be easily shown for many facilities, because the minimum of concave functions is concave. In these cases many efficient algorithms exist. There are also interesting extensions to network domains (see [6]).

Our objective function is more general, non-concave (if $\alpha > 1$), and allows for a continuous distribution of the customers. The entire domain has to be considered as potential locations of facilities. The objective function may have many local and global minima; to reduce their number we can safely restrict the search region $X = [0,1]^{kn}$ by the additional requirement

$$x \in D = \{x \in \mathbb{R}^n \ : \ x_1^1 \le x_1^2 \le \ldots \le x_1^n\}.$$

In the following subsections we shall show how to specialize the general ideas of the stochastic branch and bound method to the problem under considerations. In particular, we shall show that even for deterministic facility location problems, if we interpret them as stochastic problems, new classes of lower bounds can be derived.

5.2 Branching

The original hypercube $X = [0,1]^{kn}$ is sequentially subdivided into hyperrectangles of the form $X' = X^1 \times \ldots \times X^n$, where

$$X^i = \{x^i \in X \ : \ c_j^i \le x_j^i \le d_j^i, \ \ j = 1, \ldots, k\}, \quad i = 1, \ldots, n.$$

At each step of the branch and bound method we first select a subset X' with the smallest lower bound, then we find the variable x_j^i with the longest feasible interval $|d_j^i - c_j^i|$ and divide this interval by half to produce two new subsets.

5.3 Upper bound estimation

To obtain an upper bound for $F^*(X' \cap D)$ we have to point out a reasonable heuristic solution $\tilde{x} \in X' \cap D$ for the subproblem

$$\min_{x \in X' \cap D} F(x).$$

Suppose $X' \cap D \neq \emptyset$. Let

$$\tilde{c}_1^i = \max_{1 \le l \le i} c_1^l, \quad \tilde{d}_1^i = \min_{i \le l \le n} d_1^l, \quad i = 1, \ldots, n;$$

$$\tilde{c}_j^i = c_j^i, \quad \tilde{c}_j^i = c_j^i, \text{ for all } i, j, \ j \neq 1;$$

and define

$$\tilde{X}' = \{x \in X \ : \ \tilde{c}_j^i \le x_j^i \le \tilde{d}_j^i, \ \ j = 1, \ldots, k; \ \ i = 1, \ldots, n\}.$$

Clearly, $X' \cap D = \tilde{X}' \cap D$. We assume that the customers lie in a finite set $\Theta = \{\theta_1, \ldots, \theta_m\}$. To find a heuristic solution we associate each customer with at least one of the sets $[\tilde{c}^i, \tilde{d}^i]$, $i = 1, \ldots, n$, in the following way. Denote

$$\mathrm{dist}(\theta, [c, d]) = \min_{x \in [c,d]} \varphi(x, \theta)$$

and define for each index i the subset

$$\Theta_i = \left\{ \theta \in \Theta \ : \ \mathrm{dist}(\theta, [\tilde{c}^i, \tilde{d}^i]) = \min_{1 \le j \le n} \mathrm{dist}(\theta, [\tilde{c}^j, \tilde{d}^j]) \right\}.$$

Let us try to locate the facility $x^i \in [\tilde{c}^i, \tilde{d}^i]$ to satisfy customers $\theta \in \Theta_i \neq \emptyset$ in the best way. This is equvalent to solving the problem

$$\min_{x^i \in [\tilde{c}^i, \tilde{d}^i]} [F_i(x^i) = \sum_{\theta \in \Theta_i} \varphi(x^i, \theta)], \tag{5.2}$$

but except for the discrete case, when the solution is known to lie either at one of the ends $\tilde{c}^i$, $\tilde{d}^i$ or in $\theta \in \Theta_i \cap [\tilde{c}^i, \tilde{d}^i]$, one has to resort to a heuristics. The simplest and relatively effective one is to take the point

$$\tilde{x}^i = \begin{cases} \mathbb{E}\{x^i(\theta) |\, \Theta^i\}, & \Theta^i \neq \emptyset, \\ (\tilde{c}^i + \tilde{d}^i)/2, & \text{otherwise.} \end{cases}$$

5.4 Lower bound estimation

We shall use the interchange relaxation of section 4.1. Notice that $X' \cap D \subseteq \tilde{X}'$. Then, as in Example 4.1,

$$F^*(X' \cap D) \ge \mathbb{E} \min_{x \in \tilde{X}'} \min_{1 \le i \le n} \varphi(x^i, \theta) = \mathbb{E} \min_{1 \le i \le n} \min_{x \in \tilde{X}'} \varphi(x^i, \theta) = L_1(X').$$

We shall also use lower bounds with double sampling (see (4.6)) which take on the form:

$$L_2(X') = \frac{1}{2} \mathbb{E}_{\theta_1} \mathbb{E}_{\theta_2} \min_{x \in \tilde{X}'} (\min_{1 \le i \le n} \varphi(x^i, \theta_1) + \min_{1 \le j \le n} \varphi(x^j, \theta_2)).$$

In the case of finitely many locations of customers $\theta_1, \ldots, \theta_m$ taken with probabilities $p_1, \ldots, p_m$ the above lower bound can be written as follows:

$$L_2(X') = \frac{1}{2} \sum_{s,t=1}^{m} p_s p_t \min \left\{ \min_{1 \le i \le m} \min_{x^i \in \tilde{X}'_i} (\varphi(x^i, \theta_1^s) + \varphi(x^i, \theta_2^t)), \right.$$

$$\left. \min_{\substack{1 \le i,j \le m \\ i \ne j}} (\min_{x^i \in X'_i} \varphi(x^i, \theta_1^s) + \min_{x^j \in X'_j} \varphi(x^j, \theta_2^t)) \right\}.$$

To calculate it one has to solve (quasi-convex) problems of the form

$$\min_{x^i \in X'_i} \varphi(x^i, \theta),$$

which ere easy (simply project θ on the set X'_i), and non-convex problems of the form

$$\min_{x^i \in \tilde{X}'_i} (\varphi(x^i, \theta_1) + \varphi(x^i, \theta_2)),$$

which are more difficult, but for which simple lower bounds can be calculated.

5.5 Deletion rule

There are two reasons to delete a subset X' from the current partition $\mathcal{P}_k$: 'infeasibility', i.e. $X' \cap D = \emptyset$, and 'nonprospectiveness', i.e. $L(X') > F(\overline{x})$ for some point $\overline{x} \in X \cap D$.

The feasibility test for our set D amounts to checking the inequalities $\tilde{c}_1^i \leq \tilde{d}_1^i$, $i = 1, \ldots, n$.

If the lower bounds $L(X')$, $X' \in \mathcal{P}_k$, are calculated exactly as in the case of discrete distribution of customers, then 'nonprospective' subsets can be safely deleted from the current partition. If we have only stochastic estimates $\xi_k(X')$ of lower bounds $L(X')$ then deletion of sets X' such that $\xi_k(X') > F(\overline{x})$ leads to a heuristic branch and bound algorithm, which, nevertheless, can be helpful, if we apply it many times to the same problem.

5.6 Numerical results

To illustrate the properties of the method let us consider the facility location problem on the line with the distance function (5.1) with $\alpha = \beta = 2$ and $\gamma = 0.1$. The locations of the customers $\theta_1, \ldots, \theta_m$ were drawn from the uniform distribution in $[0, 1]$. For each customer θ_s, its relative weight w_s was drawn from the uniform distribution in $[0, \theta_s(1 - \theta_s)^2]$, so that the weight had the Beta distribution in $[0, 1]$ with parameters (2,3). To allow an easy probabilistic interpretation the weights were normalized to get $p_s = w_s / \sum_{l=1}^m w_l$. All data are collected in Table 1.

The performance of the method is summarized in Table 2. The method terminated when the size of the record set was below 10^{-3} for each facility, or when the maximum number of iterations (10^5) was reached.

We see that the method is capable of solving within a reasonable time problems with highly non-convex functions and with a large number of local minima.

References

[1] Birge J.R. (1985), Decomposition and partitioning methods for multistage stochastic linear programs, *Operations Research* 33(1985) 989-1007.

[2] Ermoliev Yu.M. (1983), Stochastic quasi-gradient methods and their application to systems optimization, *Stochastics* 4, 1-37.

[3] Hansen P., B. Jaumard and H. Tuy (1995), Global optimization in location,1995), in: *Facility Location - A Survey of Applications and Methods*, Z. Drezner (ed.), Springer Serier in Operations Research, Springer-Verlag, Berlin, pp. 43-68.

[4] Horst R. and H. Tuy (1990), *Global Optimization*, Springer-Verlag, Berlin.

[5] Horst R. and P.M. Pardalos (eds.) (1994), *Handbook of Global Optimization*, Kluwer Academic Publishers.

[6] Labbé M., D. Peeters and J.-F. Thisse (1995), Location on Networks, Ch. 7 of *Handbooks in OR & MS, Vol. 8*, M.O. Ball (ed.), Elsevier, Amsterdam, pp. 551-624.

[7] Mikhalevich V.S., A.M. Gupal and V.I. Norkin (1987), *Methods of Non-Convex Optimization*, Nauka, Moscow (in Russian).

[8] Mulvey J.M. and A. Ruszczyński (1995), A new scenario decomposition method for large-scale stochastic optimization, *Operations Research* 43(1995) 477-490.

[9] Norkin V.I. (1993), *The analysis and optimization of probability functions*, Working Paper WP-93-6, International Institute for Applied System Analysis, Laxenburg, Austria.

[10] Norkin V.I., Yu.M. Ermoliev and A. Ruszczyński (1994), *On Optimal Allocation of Indivisibles Under Uncertainty* , Working Paper WP-94-21, International Institute for Applied System Analysis, Laxenburg, Austria.

[11] Pflug G. Gh. (1994), *Adaptive Designs in Discrete Stochastic Optimization*, Working Paper WP-94-59, International Institute for Applied System Analysis, Laxenburg, Austria.

[12] Pijavskii S.A. (1972), An algorithm for searching for the absolute extremum of a function, *Zh. Vychisl. Mat. Mat. Fiz.*, Vol. 12, No. 4, pp. 888-896. (In Russian).

[13] Prekopa A. (1980), Logarithmic concave measures and related topics, in: M.A.H. Dempster (ed.), *Stochastic Programming*, Academic Press, London, pp. 63-82.

[14] Rockafellar R.T. and R.J.-B. Wets (1991), Scenarios and policy aggregation in optimization under uncertainty, *Mathematics of Operations Research* 16(1991) 1-23.

[15] R.J.-B. Wets (1988), Large scale linear programming techniques, in: Yu. Ermoliev and R.J.-B. Wets (eds.), *Numerical Methods in Stochastic Programming* (Springer-Verlag, Berlin, 1988) pp. 65-94.

$m = 10$					
Value	0.193362	0.260176	0.326991	0.453538	0.606194
Probability	2.99106E-02	2.51909E-01	2.97208E-01	1.90780E-01	9.85524E-02
Value	0.646899	0.720796	0.835398	0.885398	0.933185
Probability	7.25310E-03	6.39571E-02	2.77984E-02	2.75176E-02	5.11388E-03
$m = 20$					
Value	0.03799l	0.100437	0.193362	0.260176	0.326991
Probability	2.15658E-02	3.55347E-03	8.26707E-02	9.97986E-02	2.58563E-02
Value	0.442882	0.453538	0.480874	0.518605	0.518865
Probability	1.20332E-01	1.42203E-01	9.61612E-02	7.31521E-02	2.01054E-02
Value	0.595109	0.606194	0.646899	0.720796	0.747336
Probability	7.57130E-02	8.86943E-02	4.32473E-02	1.66506E-02	3.97192E-02
Value	0.835398	0.847773	0.885398	0.933185	0.999739
Probability	2.55506E-02	1.52854E-02	8.41436E-03	1.32620E-03	7.08675E-08
$m = 50$					
Value	0.030084	0.036942	0.037991	0.100437	0.141456
Probability	9.75020E-03	1.44166E-02	1.05256E-02	2.06615E-02	9.14801E-03
Value	0.147532	0.149556	0.163543	0.179640	0.193362
Probability	3.66614E-02	4.64783E-02	3.49908E-02	3.25987E-02	1.37885E-02
Value	0.209725	0.250422	0.260176	0.269027	0.326991
Probability	4.96726E-02	3.23409E-03	5.72782E-02	6.11222E-02	5.35531E-02
Value	0.389365	0.442882	0.453130	0.453538	0.480874
Probability	4.65960E-02	2.97852E-02	9.72700E-03	3.90675E-02	4.66834E-02
Value	0.514380	0.518344	0.518605	0.518865	0.555292
Probability	2.21801E-02	9.28766E-03	3.12571E-02	4.05158E-02	1.44466E-02
Value	0.575999	0.587544	0.592240	0.595109	0.599090
Probability	3.12875E-04	1.34509E-02	1.35490E-02	2.65526E-02	3.94564E-02
Value	0.606194	0.611501	0.627076	0.646899	0.655837
Probability	2.03139E-02	1.47669E-02	3.27974E-02	5.66449E-03	3.46706E-02
Value	0.657526	0.703552	0.720796	0.739772	0.747336
Probability	3.90958E-03	1.92653E-03	4.59931E-03	5.66151E-03	9.29691E-03
Value	0.797293	0.835398	0.847773	0.880528	0.885398
Probability	1.00864E-02	8.35174E-04	6.77332E-03	4.79698E-03	3.03047E-03
Value	0.887304	0.933185	0.953974	0.988455	0.999739
Probability	1.98831E-03	1.82071E-03	2.98361E-04	1.56585E-05	1.81248E-08

Table 5.1: Customer distribution.

m	n	Nodes	Time	Solution
10	1	51	3.33	0.351974
	2	916	62.27	(0.327390, 0.691934)
	3	9419	194.47	(0.291554, 0.453538, 0.693887)
	4	42333	4701.66	(0.292259, 0.453538, 0.649881, 0.865814)
	5	23347	2168.23	(0.253749, 0.326991, 0.453538, 0.649556, 0.865890)
20	1	60	4.06	0.486726
	2	813	59.70	(0.227088, 0.529698)
	3	8427	723.99	(0.226360, 0.470136, 0.658535)
	4	84471	14352.06	(0.227246, 0.469872, 0.610555, 0.790888)
	5	100000	19392.85	(0.045469, 0.242011, 0.469479, 0.609782, 0.790392)
50	1	106	8.86	0.391113
	2	943	106.15	(0.203609, 0.548336)
	3	14901	2409.89	(0.202405, 0.470259, 0.635109)
	4	100000	28924.74	(0.151286, 0.284003, 0.471469, 0.634133)

Table 5.2: Performance of the method.