RANDOMIZED ALGORITHMS FOR SEMI-INFINITE PROGRAMMING PROBLEMS

Vladislav B. Tadić[1], **Sean P. Meyn**[2], **Roberto Tempo**[3]

[1]Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD, UK
e-mail: v.tadic@sheffield.ac.uk
[2]Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, Illinois 61801, USA
e-mail: s-meyn@uiuc.edu
[3] IEIIT-CNR, Politecnico di Torino, 10129 Torino, Italy
e-mail: tempo@polito.it

Keywords: Randomized algorithms, semi-infinite programming, stochastic optimization, stochastic approximation, Monte Carlo methods.

Abstract

In this paper, we explore the possibility of applying Monte Carlo methods (i.e., randomization) to semi-infinite programming problems. Equivalent stochastic optimization problems are derived for a general class of semi-infinite programming problems. For the equivalent stochastic optimization problems, algorithms based on stochastic approximation and Monte Carlo sampling methods are proposed. The asymptotic behavior of the proposed algorithms is analyzed and sufficient conditions for their almost sure convergence are obtained.

1 Introduction

In this paper we consider semi-infinite programming problems in which there are a possibly uncountable number of constraints. As a special case, we also consider the determination of a feasible solution to an uncountable number of inequality constraints. Problems of this type naturally arise in optimal control, filter design, optimal experiment design, reliability and numerous other engineering problems where the underlying model contains at least one inequality for each value of a parameter and where the parameter (usually representing time, frequency or space) varies over an uncountable set (for more details see e.g., [2], [5], [8] and references cited therein).

Existing algorithms for semi-infinite programming problems are deterministic numerical procedures which use a deterministic grid to discretize the original inifinitely constrained model, i.e., to approximate the original problem by an optimization problem with finitely many constraints (for a recent survey see [9]; see also [3], [6]). Because of that, they suffer from the curse of dimensionality: their computational complexity is exponential in the problem dimension.

Randomized algorithms is currently an active area of research within the field of robust control (see e.g., [12]). In particular, in this paper, the possibility of applying Monte Carlo methods (i.e., randomization) to semi-infinite programming problems is explored. For a general class of semi-infinite programming problems, equivalent stochastic optimization problems are derived. Algorithms based on stochastic approximation and Monte Carlo sampling methods are proposed for the equivalent stochastic optimization problems. The asymptotic behavior of the proposed algorithms is analyzed and sufficient conditions for their almost sure convergence are obtained. The general results on Monte Carlo methods (see e.g., [10]), as well as the initial theoretical results reported in this paper suggest that the computational complexity of the proposed algorithms is considerably reduced in comparison with the existing deterministic methods.

The paper is organized as follows. In Section 2, the semi-infinite programming problems studied in this paper are formally defined. The problem of equivalent stochastic programming representation of semi-infinite programming problems are considered in Section 3, while algorithms for these equivalent problems are proposed and analyzed in Sections 4 and 5.

2 Semi-Infinite Programming Problems

Let $f : R^p \to R$ and $g : R^p \times R^q \to R$ be Borel-measurable functions, while

$$D = \{x \in R^p : g(x,y) \le 0, \forall y \in R^q\}. \quad (1)$$

In this paper, we consider the determination of a solution to the following system of uncountably many inequalities:

$$g(x,y) \le 0, \quad \forall y \in R^q \quad (2)$$

(i.e., we want to find a point from the set D). Besides a system of infinitely many inequalities, the following optimization problem with infinitely many constraints is also considered in this paper:

$$\begin{aligned} &\min f(x) \\ &\text{subject to:} \quad g(x,y) \le 0, \quad \forall y \in R^q \end{aligned} \quad (3)$$

(i.e., we want to find a (local) minimum of $f(\cdot)$ over the set D).

The problems (2) and (3) fall into the category of semi-infinite programming (for more details see [2], [5], [8] and references cited therein). The reason why (2) and (3) are called semi-infinite comes out of the fact that the number of constraints these problems require to be satisfied is infinite. Such requirements naturally arise in engineering design where it is necessary to maintain the response of a dynamical system within prescribed performance envelopes.

Throughout the paper, the following notation is used. $\| \cdot \|$ denotes the Euclidean norm in R^p and R^q, while $B^p_\rho(x) = \{x' \in R^p : \|x - x'\| \leq \rho\}$, $B^q_\rho(y) = \{y' \in R^q : \|y - y'\| \leq \rho\}$, $B^p_\rho = B^p_\rho(0)$ and $B^q_\rho = B^q_\rho(0)$ for $x \in R^p$, $y \in R^q$, $\rho \in (0, \infty)$. Moreover, $d(\cdot, \cdot)$ is the distance in R^p and R^q induced by the Euclidean norm.

3 Equivalent Stochastic Programming Representation

In this section, the problem of the equivalent stochastic programming representation of the semi-infinite programming problems (2), (3) is considered.

Let $\mathcal{A}^q$ be the family of open sets from R^q, while $\mathcal{B}^q$ is the set of Borel-measurable sets from R^q. Moreover, let $\mu(\cdot)$ be a probability measure on $(R^q, \mathcal{B}^q)$, while $h : R \to [0, \infty)$ is a continuous function. The problem of equivalent stochastic programming representation of (2) and (3) is analyzed under the following assumptions:

A1 *$\mu(A) > 0$ for all $A \in \mathcal{A}^q$.*

A2 *$g(x, \cdot)$ is continuous for all $x \in R^p$.*

A3 *$h(t) = 0$ for all $t \in (-\infty, 0]$ and $h(t) > 0$ for all $t \in (0, \infty)$.*

Remark *It is straightforward to verify that A1 hold if*

$$\mu(B) = \int_B p(x) dx, \tag{4}$$

where $p(\cdot)$ is a strictly positive probability density function on (R^q, B^q) (i.e., $p : R^q \to (0, \infty)$).

The equivalent stochastic programming representation of the semi-infinite problems (2) and (3) is essentially based on the following theorem:

Theorem 1 *Let A1 – A3 hold. Then,*

$$D = \{x \in R^p : \psi(x) = 0\}, \tag{5}$$

where

$$\psi(x) = \int h(g(x, y))\mu(dy), \quad x \in R^p. \tag{6}$$

Proof Let x be an arbitrary element of D. Then,

$$h(g(x, y)) = 0$$

for all $y \in R^q$. Therefore, $\psi(x) = 0$, and consequently, $x \in D$. Hence,

$$D \subseteq \{x \in R^p : \psi(x) = 0\}. \tag{7}$$

Now, let x be an arbitrary element of $\{x \in R^p : \psi(x) = 0\}$. Suppose that $x \notin D$. Then, there exists $y \in R^q$ (depending on x) such that $g(x, y) > 0$. Since $g(x, \cdot)$ and $h(\cdot)$ are continuous, there exist constants $\delta, \varepsilon \in (0, \infty)$ such that

$$h(g(x, y')) \geq \varepsilon \tag{8}$$

for all $y' \in B^q_\delta(y)$. Due to A1 and (8)

$$\psi(x) \geq \int_{B^q_\delta(y)} h(g(x, y'))\mu(dy') \geq \varepsilon\mu(B^q_\delta(y)) > 0.$$

However, this is impossible since $\psi(x) = 0$. Consequently, $x \notin D$. Hence,

$$D \supseteq \{x \in R^p : \psi(x) = 0\}. \tag{9}$$

The theorem's assertion is a direct consequence of (7) and (9). ■

Let Y be an R^q-valued random variable defined on a probability space $(\Omega, \mathcal{F}, P)$ whose probability measure is $\mu(\cdot)$, i.e.,

$$P(Y \in B) = \mu(B) \tag{10}$$

for all $B \in \mathcal{B}^q$. Then,

$$\psi(x) = E(h(g(x, Y))) \tag{11}$$

for all $x \in R^p$, while the following corollaries are a direct consequence of Theorem 1:

Corollary 1 *Let A1 – A3 hold. Then, $x \in R^p$ solves the semi-infinite problem (2) if and only if it solves the equation*

$$E(h(g(x, Y))) = 0. \tag{12}$$

Corollary 2 *Let A1 – A3 hold. Then, the semi-infinite problem (3) is equivalent to the constrained stochastic optimization problem*

$$\begin{aligned} &\min f(x) \\ &\text{subject to:} \quad E(h(g(x, Y))) = 0. \end{aligned} \tag{13}$$

Remark *Corollaries 1 and 2 suggest that the semi-infinite problems (2) and (3) can be solved by Monte-Carlo methods, i.e., by sampling from the probability measure $\mu(\cdot)$ and using appropriate algorithms to solve the associated stochastic optimization problem.*

4 Algorithms for Systems of Infinitely Many Inequalities

In this section, algorithms for the semi-infinite programming problem (2) are proposed, and their asymptotic behavior is analyzed.

Suppose that $h(\cdot)$ is differentiable ($h'(\cdot)$ denotes the derivative of $h(\cdot)$) and $g(\cdot, y)$ is differentiable for all $y \in R^q$ (notice that $h(t) = (\max\{0, t\})^2$, $t \in R$, is differentiable and satisfies A3). Due to Theorem 1, any solution of (2) is a (global) minimum of $\psi(\cdot)$ (provided A1 – A3 hold). Therefore, the problem (2) can be considered as the minimization of $\psi(\cdot)$ and the following gradient algorithm can be used:

$$x_{n+1} = x_n - \gamma_{n+1} \nabla \psi(x_n), \quad n \geq 0,$$

where $\{\gamma_n\}_{n\geq 1}$ is a sequence of positive reals. Apart from few special cases (see e.g., [7]), it is hard (if possible at all) to determine analytically $\nabla\psi(\cdot)$, while the deterministic approximations of $\nabla\psi(\cdot)$ result in a computationally intractable algorithms. On the other hand, (11) implies that

$$\nabla \psi(x) = E(h'(g(x, Y)) \nabla_x g(x, Y)) \tag{14}$$

for all $x \in R^p$ (under mild regularity conditions; see B1 below), which itself provides a 'stochastic approximation' of $\nabla\psi(\cdot)$: $h'(g(x, Y))\nabla_x g(x, Y)$ can be used as an unbiased estimate of $\nabla\psi(\cdot)$. Then, it is quite natural to use the following stochastic gradient algorithm to search for the minima of $\psi(\cdot)$:

$$X_{n+1} = X_n - \gamma_{n+1} h'(g(X_n, Y_{n+1})) \cdot \nabla_x g(X_n, Y_{n+1}), \quad n \geq 0. \tag{15}$$

$\{\gamma_n\}_{n\geq 1}$ is a sequence of positive reals, while Y_n, $n \geq 1$, are independent random samples from the probability measure $\mu(\cdot)$, i.e., $\{Y_n\}_{n\geq 1}$ is a sequence of R^q-valued i.i.d. random variable satisfying

$$P(Y_n \in B) = \mu(B), \quad n \geq 1,$$

for all $B \in \mathcal{B}^q$.

The algorithm (15) falls into the category of stochastic approximation algorithms (for details see [1], [4] and references cited therein). Therefore, its analysis is based on asymptotic results for stochastic approximation.

The asymptotic behavior of the algorithm (15) is analyzed under the following assumptions:

B1 $\lim_{n\to\infty} \gamma_n = 0$, $\sum_{n=1}^{\infty} \gamma_n = \infty$ *and* $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$.

B2 *For all* $\rho \in [1, \infty)$, *there exists a Borel-measurable function* $\phi_\rho : R^q \to [1, \infty)$ *and such that*

$$\int \phi_\rho^4(y) \mu(dy) < \infty,$$

and

$$\max\{|h(g(x, y))|, |h'(g(x, y))|, \|\nabla_x g(x, y)\|\} \leq \phi_\rho(y),$$
$$|h'(g(x', y)) - h'(g(x'', y))| \leq \phi_\rho(y)\|x' - x''\|,$$
$$\|\nabla_x g(x', y) - \nabla_x g(x'', y)\| \leq \phi_\rho(y)\|x' - x''\|$$

for all $x, x', x'' \in B_\rho^p$, $y \in R^q$.

B3 $\nabla\psi(x) \neq 0$ *for all* $x \notin D$.

B1 holds if $\gamma_n = n^{-c}$, $n \geq 1$, where $c \in (1/2, 1]$ is a constant (which is a typical choice for the step-sizes in stochastic approximation algorithms).

In the case where $h(t) = (\max\{0, t\})^2$, $t \in R$ (which is probably the most convenient choice for $h(\cdot)$), B2 is satisfied if for all $\rho \in [1, \infty)$ there exists a Borel-measurable function $\varphi_\rho : R^q \to [1, \infty)$ such that

$$\int \varphi_\rho^4(y) \mu(dy) < \infty,$$

and

$$\max\{g^2(x, y), \|\nabla_x g(x, y)\|\} \leq \varphi_\rho(y),$$
$$|g(x', y) - g(x'', y)| \leq \varphi_\rho(y)\|x' - x''\|,$$
$$\|\nabla_x g(x', y) - \nabla_x g(x'', y)\| \leq \varphi_\rho(y)\|x' - x''\|$$

for all $x, x', x'' \in B_\rho^p$, $y \in R^q$. On the other hand, in the case where $g(x, y) = a^T(y)x + b(y)$, $x \in R^p$, $y \in R^q$ and $a : R^q \to R^p$, $b : R^q \to R$ are Borel-measurable functions (the case of linear constraints in (3)), such a Borel-measurable function $\varphi_\rho(\cdot)$ exists for all $\rho \in [1, \infty)$ if

$$\int \|a(y)\|^4 \mu(dy) < \infty,$$

$$\int |b(y)|^4 \mu(dy) < \infty.$$

B3 is satisfied if $g(\cdot, y)$ is convex for all $y \in R^q$ (the case of convex constraints in (3), which could be considered as one of the most important special cases of (3)) and $h(\cdot)$ is non-decreasing and convex. In that case, $\psi(\cdot)$ itself is convex, and consequently, $\nabla\psi(x) \neq 0$ for all $x \notin D$, since $\psi(x) = 0$ for all $x \in D$ and $\psi(x) > 0$ for all $x \notin D$.

Theorem 2 *Let B1 and B2 hold. Suppose that* $D \neq \emptyset$. *Then,* $\lim_{n\to\infty} d(X_n, C) = 0$ $w.p.1$ *on the event* $\{\sup_{0\leq n} \|X_n\| < \infty\}$, *where* $C = \{x \in R^p : \nabla\psi(x) = 0\}$. *Moreover, if A1 – A3 and B3 hold, then* $\lim_{n\to\infty} d(X_n, D) = 0$ $w.p.1$ *on the event* $\{\sup_{0\leq n} \|X_n\| < \infty\}$.

For a proof see [11].

In most practical situations, we want to find a solution of the semi-infinite problem (2) which lies in a predetermined bounded set. In that case, instead of (15), projected stochastic gradient algorithms can be used. These algorithms are defined by the following difference equation:

$$X_{n+1} = \Pi_Q(X_n - \gamma_{n+1} h'(g(X_n, Y_{n+1})) \cdot \nabla_x g(X_n, Y_{n+1})), \quad n \geq 0. \tag{16}$$

$\{\gamma_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ have the same meaning as in the case of the algorithm (15), while $Q \subset R^p$ is a compact convex set and $\Pi_Q(\cdot)$ is the projection on Q defined as

$$\Pi_Q(x) = \arg\min_{x' \in Q} \|x - x'\|, \quad x \in R^p.$$

5 Algorithms for Semi-Infinite Optimization Problems

In this section, algorithms for the semi-infinite programming problem (3) are proposed, and their asymptotic behavior is analyzed.

Suppose that $h(\cdot)$ is differentiable and $g(\cdot, y)$ is differentiable for all $y \in R^q$. Due to the Theorem 1, the semi-infinite problem (3) is equivalent to the following constrained optimization problem:

$$\begin{aligned} &\min f(x) \\ &\text{subject to:} \quad \psi(x) = 0 \end{aligned} \tag{17}$$

(provided A1 – A3 hold). Let $\{\delta_n\}_{n\geq 1}$ be an increasing sequence of positive reals satisfying $\lim_{n\to\infty} \delta_n = \infty$. Since $\psi(x) \geq 0$ for all $x \in R^p$, $\{\delta_n \psi(\cdot)\}_{n\geq 1}$, can be used as penalty functions for (17). Therefore, the following gradient algorithm (with penalty functions) can be used to solve the problem (3):

$$x_{n+1} = x_n - \gamma_{n+1}(\nabla f(x_n) + \delta_{n+1}\psi(x_n)), \quad n \geq 0,$$

where $\{\gamma_n\}_{n\geq 1}$ is a sequence of positive reals. However, it is difficult to determine analytically $\nabla\psi(\cdot)$. As $h'(g(x,Y))\nabla_x g(x,Y)$ can be used as an unbiased estimate of $\nabla\psi(\cdot)$ (due to (14)), it is quite natural to use the following stochastic gradient algorithm (with penalty functions) to search for the minima of $\psi(\cdot)$:

$$\begin{aligned} X_{n+1} = X_n - \gamma_{n+1}(\nabla f(X_n) + \delta_{n+1} h'(g(X_n, Y_{n+1})) \\ \cdot \nabla_x g(X_n, Y_{n+1})), \quad n \geq 0. \end{aligned} \tag{18}$$

$\{\gamma_n\}_{n\geq 1}$ is a sequence of positive reals, while Y_n, $n \geq 1$, are independent random samples from the probability measure $\mu(\cdot)$, i.e., $\{Y_n\}_{n\geq 1}$ is a sequence of R^q-valued i.i.d. random variable satisfying

$$P(Y_n \in B) = \mu(B), \quad n \geq 1,$$

for all $B \in \mathcal{B}^q$.

For the analysis of the algorithm (18), the following assumptions are needed:

C1 $\lim_{n\to\infty} \gamma_n = 0$, $\sum_{n=1}^{\infty} \gamma_n = \infty$ *and* $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$.

C2 $\lim_{n\to\infty} \delta_n = \infty$, $\lim_{n\to\infty} \gamma_n \delta_n = 0$ *and* $\lim_{n\to\infty} \gamma_{n+1}^{-1} \delta_{n+1}^{-1} (\delta_{n+1} - \delta_n) = 0$.

C3 $f(\cdot)$ *is convex and* $\nabla f(\cdot)$ *is locally Lipschitz continuous.*

C4 $h(\cdot)$ *is convex and* $g(\cdot, y)$ *is convex for all* $y \in R^q$. *For all* $\rho \in [1, \infty)$, *there exists a Borel-measurable function* $\phi_\rho : R^q \to [1, \infty)$ *and such that*

$$\int \phi_\rho^4(y)\mu(dy) < \infty,$$

and

$$\max\{|h(g(x,y))|, |h'(g(x,y))|, \|\nabla_x g(x,y)\|\} \leq \phi_\rho(y),$$

$$|h'(g(x',y)) - h'(g(x'',y))| \leq \phi_\rho(y)\|x' - x''\|,$$

$$\|\nabla_x g(x',y) - \nabla_x g(x'',y)\| \leq \phi_\rho(y)\|x' - x''\|$$

for all $x, x', x'' \in B_\rho^p$, $y \in R^q$.

C5 $f(\cdot)$ *has a finite minimum* x_* *on* D.

Theorem 3 *Let A1 – A3 and C1 – C5 hold. Then,* $\lim_{n\to\infty} X_n = x_*$ *w.p.*1 *on the event* $\{\sup_{0\leq n} \|X_n\| < \infty\}$.

For a proof see [11].

Acknowledgements

This research was performed while Vladislav B. Tadić was with the Department of Electrical and Electronic Engineering of the University of Melbourne. His research work was supported by the Australian Research Council (ARC).

Research conducted by Sean P. Meyn is supported by the National Science Foundation under Award Nos. ECS 02 17836 and DMI 00 85165. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Part of this research was performed while Roberto Tempo was visiting the Coordinated Science Laboratory of the University of Illinois as a CSL Visiting Research Professor, the support is gratefully acknowledged.

References

[1] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximation*, Springer Verlag, 1990.

[2] M. A. Goberna and M. A. Lopez (Eds.), *Semi-Infinite Programming: Recent Advances*, Kluwer Academic, 2001.

[3] R. Hettich and K. O. Kortanek, "Semi-infinite programming: theory, methods, and applications," *SIAM Review*, vol. 35, pp. 380–429, 1993.

[4] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer Verlag, 1997.

[5] E. Polak, *Optimization : Algorithms and Consistent Approximations*, Springer Verlag, 1997.

[6] E. Polak, “On the use of consistent approximations in the solution of semi-infinite optimization and optimal control problems,” *Mathematical Programming*, vol. 62, pp. 385–414, 1993.

[7] B. T. Polyak and R. Tempo, “Probabilistic robust design with linear quadratic regulator,” *System and Control Letters*, vol. 43, pp. 343–353, 2001.

[8] R. Reemtsen and J.-J. Rückmann (Eds.), *Semi-Infinite Programming*, Kluwer Academic, 1998.

[9] R. Reemtsen, “Numerical methods for semi-infinite programming: A survey,” in R. Reemtsen and J.-J. Rückmann (Eds.), *Semi-Infinite Programming*, Kluwer Academic, 1998.

[10] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Verlag, 1999.

[11] V. B. Tadić, S. P. Meyn and R. Tempo, “Simulation based algorithms for solving semi-infinite programming problems”, in preparation.

[12] R. Tempo and F. Dabbene, “Randomized algorithms for analysis and control of uncertain systems: An overview,” in *Perspectives in Robust Control*, Springer Verlag, 2001.