Randomized Algorithms for Semi-Infinite Programming Problems

Vladislav B. Tadić[1], Sean P. Meyn[2], and 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 and the Coordinated
Science Laboratory, University of Illinois, Urbana-Champaign, IL 61801, USA
e-mail: meyn@uiuc.edu

[3] IEIIT-CNR, Politecnico di Torino, 10129 Torino, Italy
e-mail: tempo@polito.it

Summary. This paper studies the development of Monte Carlo methods to solve semi-infinite, nonlinear programming problems. An equivalent stochastic optimization problem is proposed, which leads to a class of randomized algorithms based on stochastic approximation. The main results of the paper show that almost sure convergence can be established under relatively mild conditions.

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

1 Introduction

In this paper, we consider semi-infinite programming problems consisting of a possibly uncountable number of constraints. As a special case, we also study the determination of a feasible solution to an uncountable number of inequality constraints. Computational problems of this form arise in optimal and robust control, filter design, optimal experiment design, reliability, and numerous other engineering problems in which the underlying model contains a parameterized family of inequality constraints. More specifically, these parameters may represent time, frequency, or space, and hence may vary over an uncountable set.

The class of problems considered here are known as *semi-infinite programming problems* since the number of constraints is infinite, but there is a finite number of variables (see e.g. [7, 15, 18] and references cited therein). Several deterministic numerical procedures have been proposed to solve problems of this kind. Standard approach is to approximately solve the optimization problem through discretization using a deterministic grid (for a recent survey see

[19], and also [8, 16]). The algorithms based on this approach typically suffer from the curse of dimensionality so that their computational complexity is generally exponential in the problem dimension, see e.g. [25].

This paper explores an alternative approach based on Monte Carlo methods and randomized algorithms. The use of randomized algorithms has become widespread in recent years for various problems, and is currently an active area of research. In particular, see [24] for a development of randomized algorithms for uncertain systems and robust control; [1, 5] for applications to reinforcement learning, and approximate optimal control in stochastic systems; [4, 22] for topics in mathematical physics; [13, 14] for applications in computer science and computational geometry; [6] for a treatment of Monte Carlo methods in finance; and [21] for a recent survey on Markov Chain Monte Carlo methods for approximate sampling from a complex probability distribution, and related Bayesian inference problems.

The main idea of this paper is to reformulate the semi-infinite programming problem as a stochastic optimization problem that may be solved using stochastic approximation methods [9, 11]. The resulting algorithm can be easily implemented, and is provably convergent under verifiable conditions. The general results on Monte Carlo methods as well as the theoretical results reported in this paper suggest that the computational complexity of the proposed algorithms is considerably reduced in comparison with existing deterministic methods (see also [20]).

The paper is organized as follows. Section 2 contains a description of the general semi-infinite programming problems considered in this paper, and an equivalent stochastic programming representation of these semi-infinite programming problems is proposed. Randomized algorithms to solve the stochastic programming problems are introduced in Section 3, and convergence proofs are contained in the Appendix.

2 Semi-Infinite Nonlinear Programming

The semi-infinite programming problems studied in this paper are based on a given Borel-measurable function $g : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$. This function is used to define the constraint region,

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

The problems considered in this paper concern computation of a point in D, and optimization of a given function f over this set. These problems are now described more precisely.

2.1 Two general computational problems

1. Constraint set feasibility

Is D non-empty? And, if so, how do we compute elements of D? That is, we seek algorithms to determine a solution $x^* \in \mathbb{R}^p$ to the following uncountably-infinite system of inequalities:

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

2. Optimization over D

For a given continuous function $f\colon \mathbb{R}^p \to \mathbb{R}$, how do we compute an optimizer over D? That is, a solution $x^* \in \mathbb{R}^p$ to the semi-infinite nonlinear program,

$$\begin{array}{l} \text{Minimize } f(x) \\ \text{Subject to:} \quad g(x, y) \le 0, \quad \forall y \in \mathbb{R}^q \end{array} \tag{3}$$

These two problems cover the following general examples:

Min-max problems

Consider the general optimization problem in which a function $f\colon \mathbb{R}^p \to \mathbb{R}$ is to be minimized, of the specific form,

$$f(x) = \max_y g(x, y), \quad x \in \mathbb{R}^p,$$

where $g\colon \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ is continuous. Under mild conditions, this optimization problem can be formulated as a semi-infinite optimization problem (for details see [15]).

Sup-norm minimization

Suppose that $H\colon \mathbb{R}^p \to \mathbb{R}^r$ is a given measurable function to be approximated by a family of functions $\{G_i : i = 1, \dots, p\}$ and their linear combinations. Let $G = [G_1 | \cdots | G_p]$ denote the $p \times r$ matrix-valued function on $\mathbb{R}^q$, and consider the minimization of the function,

$$f(x) = \sup_y \|H(y) - G(y)x\|, \quad x \in \mathbb{R}^p.$$

A vector x^* minimizing f provides the best approximation of H in the supremum norm. The components of x^* are then interpreted as basis weights. This is clearly a special case of the min-max problem in which $g(x, y) = \|H(y) - G(y)x\|$.

Common Lyapunov functions

Consider a set of parameterized real Hurwitz matrices $\mathcal{A} = \{A(y) : y \in Y \subseteq \mathbb{R}^q\}$. In this case, the feasibility problem is checking the existence of a symmetric positive definite matrix $P > 0$ which satisfies the Lyapunov strict inequalities

$$PA(y) + A^T(y)P < 0, \quad \forall y \in Y \subseteq \mathbb{R}^q.$$

This is equivalent to verify the existence of $P > 0$ which satisfies

$$PA(y) + A^T(y)P + Q \leq 0, \quad \forall y \in Y \subseteq \mathbb{R}^q,$$

where $Q > 0$ is arbitrary. Clearly, the existence of a feasible solution $P > 0$ implies that the quadratic function $V(x) = x^T Px$ is a common Lyapunov function for the family of asymptotically stable linear systems

$$\dot{z} = A(y)z, \quad \forall y \in Y \subseteq \mathbb{R}^q.$$

This feasibility problem can be reformulated as follows: Determine the existence of a symmetric positive definite matrix $P > 0$ to the system of scalar inequalities

$$\lambda_{\max}(PA(y) + A^T(y)P + Q) \leq 0, \quad \forall y \in Y \subseteq \mathbb{R}^q,$$

where $\lambda_{\max}(A)$ denotes the largest eigenvalue of a real symmetric matrix A. This observation follows from the fact that $\lambda_{\max}(A) \leq 0$ if and only if $A \leq 0$. We also notice that $\lambda_{\max}$ is a convex function and, if $\lambda_{\max}$ is a simple eigenvalue, it is also differentiable, see details in [10]. In the affirmative case when this common Lyapunov problem is feasible, clearly the objective is to find a solution $P > 0$.

2.2 Equivalent Stochastic Programming Representation

Algorithms to solve the semi-infinite programming problems (2), (3) may be constructed based on an equivalent stochastic programming representation. This section contains details on this representation and some key assumptions.

We adopt the following notation throughout the paper: The standard Euclidean norm is denoted $\|\cdot\|$, while $d(\cdot,\cdot)$ stands for the associated metric. For an integer $r \geq 1$ and $z \in \mathbb{R}^r$, $\rho \in (0, \infty)$, the associated closed balls are defined as

$$B^r_\rho(z) = \{x' \in \mathbb{R}^r : \|z - z'\| \leq \rho\}, \quad B^r_\rho = B^r_\rho(0),$$

while $\mathcal{B}^r$ denotes the class of Borel-measurable sets on $\mathbb{R}^r$.

Throughout the paper, a probability measure μ on $\mathcal{B}^q$ is fixed, where $q \geq 1$ is the integer used in (1). It is assumed that its support is full in the sense that

$$\mu(A) > 0 \text{ for any non-empty open set } A \subset \mathbb{R}^q. \tag{4}$$

In applications one will typically take μ of the form $\mu(dy) = p(y)\,dy$, $y \in \mathbb{R}^q$, where p is continuous, and strictly positive. A continuous function $h\colon \mathbb{R} \to \mathbb{R}_+$ is fixed with support equal to $(0, \infty)$, in the sense that

$$h(t) = 0 \text{ for all } t \in (-\infty, 0], \text{ and } h(t) > 0 \text{ for all } t \in (0, \infty). \tag{5}$$

For example, the function $h(t) = (\max\{0, t\})^2$, $t \in \mathbb{R}$, is a convex, C^1 solution to (5).

Equivalent stochastic programming representations of (2) and (3) are based on the probability distribution μ, the function h, and the following condition on the function g that determines the constraint region D:

$$g(x, \cdot) \text{ is continuous on } \mathbb{R}^q \text{ for each } x \in \mathbb{R}^p. \tag{6}$$

The following conditional average of g is the focus of the algorithms and results in this paper,

$$\psi(x) = \int h(g(x, y))\mu(dy), \quad x \in \mathbb{R}^p. \tag{7}$$

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

Theorem 1. *Under the assumptions of this section, the constraint region may be expressed as*

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

Proof. Suppose that $x \in D$. By definition, the following equation then holds for all $y \in \mathbb{R}^q$,

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

This and (7) establish the inclusion,

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

Conversely, if $x \notin D$, then there exists $y \in \mathbb{R}^q$ such that $g(x, y) > 0$. Continuity of the functions g and h implies that there exist constants $\delta, \varepsilon \in (0, \infty)$ such that

$$h(g(x, y')) \geq \varepsilon, \quad \text{for all } y' \in B^q_\delta(y).$$

This combined with the support assumption (4) implies that

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

which gives the reverse inclusion,

$$D^c \subseteq \{x \in \mathbb{R}^p : \psi(x) \neq 0\},$$

where D^c denotes the complement of D.

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

$$\mathsf{P}(Y \in B) = \mu(B), \quad B \in \mathcal{B}^q.$$

It follows that ψ may be expressed as the expectation,

$$\psi(x) = \mathsf{E}(h(g(x, Y))), \quad x \in \mathbb{R}^p, \tag{8}$$

and the following corollaries are then a direct consequence of Theorem 1:

Corollary 1. *A vector $x \in \mathbb{R}^p$ solves the semi-infinite problem (2) if and only if it solves the equation*

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

Corollary 2. *The semi-infinite optimization problem problem (3) is equivalent to the constrained stochastic optimization problem*

$$\begin{aligned} &\text{Minimize } f(x) \\ &\text{Subject to:} \quad \mathsf{E}(h(g(x, Y))) = 0. \end{aligned} \tag{9}$$

Corollaries 1 and 2 motivate the development of Monte-Carlo methods to solve the semi-infinite problems (2) and (3). The search for a feasible or optimal $x \in D$ may be performed by sampling $\mathbb{R}^q$ using the probability measure μ. Specific algorithms are proposed in the next section.

3 Algorithms

We begin with consideration of the constraint satisfaction problem (2).

3.1 Systems of Infinitely Many Inequalities

Theorem 1 implies that solutions of (2) are characterized as global minima for the function ψ. If the functions h and g are differentiable, then minimization of ψ may be performed using a gradient algorithm, described as follows: Given a vanishing sequence $\{\gamma_n\}_{n \geq 1}$ of positive reals, we consider the recursion

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

Unfortunately, apart from some special cases (see e.g., [17]), it is impossible to determine analytically the gradient $\nabla \psi$.

On the other hand, under mild regularity conditions, (8) implies that the gradient may be expressed as the expectation,

$$\nabla \psi(x) = \mathsf{E}(h'(g(x, Y)) \nabla_x g(x, Y)), \quad x \in \mathbb{R}^p, \tag{10}$$

where h' denotes the derivative of h. This provides motivation for the 'stochastic approximation' of $\nabla\psi$ given by

$$h'(g(x, Y))\nabla_x g(x, Y),$$

and the following stochastic gradient algorithm to search for the minima of ψ:

$$X_{n+1} = X_n - \gamma_{n+1} h'(g(X_n, Y_{n+1}))\nabla_x g(X_n, Y_{n+1}), \quad n \geq 0. \tag{11}$$

In this recursion, $\{\gamma_n\}_{n\geq 1}$ again denotes a sequence of positive reals. The i.i.d. sequence $\{Y_n\}_{n\geq 1}$ has common marginal distribution μ, so that

$$\mathsf{P}(Y_n \in B) = \mu(B), \quad B \in \mathcal{B}^q, \ n \geq 1.$$

Depending upon the specific assumptions imposed on the functions h and g, analysis of the stochastic approximation recursion (11) may be performed following the general theory of, say, [2, 9, 5, 23].

The asymptotic behavior of the algorithm (11) is analyzed under the following assumptions:

A1 *$\gamma_n > 0$ for each $n \geq 1$, $\sum_{n=1}^{\infty} \gamma_n = \infty$, and $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$.*

A2 *For each $\rho \in [1, \infty)$, there exists a Borel-measurable function $\phi_\rho : \mathbb{R}^q \to [1, \infty)$ such that*

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

and for each $x, x', x'' \in B_\rho^p$, $y \in \mathbb{R}^q$,

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

A3 *$\nabla\psi(x) \neq 0$ for all $x \notin D$.*

Assumption A1 holds if the step-size sequence is of the usual form $\gamma_n = n^{-c}$, $n \geq 1$, where the constant c lies in the interval $(1/2, 1]$.

Assumption A2 corresponds to the properties of the functions g and h. It ensures that ψ is well-defined, finite and differentiable, and that $\nabla\psi$ is locally Lipschitz continuous. This assumption is satisfied under appropriate assumptions on the function g, provided the function h is carefully chosen. Consider the special case in which h is the piecewise quadratic, $h(t) = (\max\{0, t\})^2$, $t \in \mathbb{R}$. Then, Assumption A2 holds under the following general assumptions on g: for each $\rho \in [1, \infty)$, there exists a Borel-measurable function $\varphi_\rho : \mathbb{R}^q \to [1, \infty)$ such that

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

and for each $x, x', x'' \in B_\rho^p$, $y \in \mathbb{R}^q$,

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

In the special case in which g is linear in x, so that there exist Borel-measurable functions $a : \mathbb{R}^q \to \mathbb{R}^p$, $b : \mathbb{R}^q \to \mathbb{R}$, with

$$g(x,y) = a^T(y)x + b(y), \quad x \in \mathbb{R}^p,\ y \in \mathbb{R}^q,$$

a bounding function φ_ρ may be constructed for each $\rho \in [1,\infty)$ provided

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

Assumption A3 corresponds to the properties of the stationary points of ψ. Consider the important special case in which $g(\cdot, y)$ is convex for each $y \in \mathbb{R}^q$. We may assume that h is convex and non-decreasing (notice that h is non-decreasing if it is convex and satisfies (5)), and it then follows that the function ψ is also convex in this special case. Moreover, since ψ is non-negative valued, convex, and vanishes only on D, it follows that $\nabla\psi(x) \ne 0$ for $x \in D^c$, so that Assumption A3 holds.

Theorem 2 states that stability of the algorithm implies convergence under the assumptions imposed here. General conditions to verify stability, so that $\sup_{0\le n}\|X_n\| < \infty$ holds w.p.1., are included in [2, 5, 9].

Theorem 2. *Suppose that Assumptions A1–A3 hold. Then, on the event* $\{\sup_{0\le n}\|X_n\| < \infty\}$*, we have convergence:*

$$\lim_{n\to\infty} d(X_n, D) = 0 \quad \textit{w.p.1.}$$

A proof of Theorem 2 is included in Appendix.

In many practical situations, a solution to the semi-infinite problem (2) is known to lie in a predetermined bounded set Q. In this case one may replace the iteration (11) with the following projected stochastic gradient algorithm:

$$X_{n+1} = \Pi_Q(X_n - \gamma_{n+1}h'(g(X_n, Y_{n+1}))\nabla_x g(X_n, Y_{n+1})), \quad n \ge 0. \tag{12}$$

It is assumed that the constraint set $Q \subset \mathbb{R}^p$ is compact and convex, and $\Pi_Q(\cdot)$ is the projection on Q (i.e., $\Pi_Q(x) = \arg\inf_{x'\in Q}\|x - x'\|$ for $x \in \mathbb{R}^p$). The step-size sequence and i.i.d. sequence $\{Y_n\}_{n\ge1}$ are defined as above.

Under additional assumptions on g and h, we can prove that the algorithm (12) converges.

Theorem 3. *Let $\{X_n\}_{n\geq 0}$ be generated by (12), and suppose that Assumptions A1 and A2 hold. Suppose that $D \cap Q \neq \emptyset$, h is convex, and $g(\cdot, y)$ is convex for each $y \in \mathbb{R}^q$. Then,*

$$\lim_{n\to\infty} d(X_n, D) = 0 \quad w.p.1.$$

A proof of Theorem 3 is presented in Appendix.

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

$$\begin{aligned}&\text{Minimize } f(x)\\&\text{Subject to: } \quad \psi(x) = 0.\end{aligned} \tag{13}$$

Suppose that the gradient $\nabla\psi$ could be computed explicitly. Then, under general conditions on the function f and the set D, the constrained optimization problem (3) could be solved using the following penalty-function approach (see e.g., [3, 15]). 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 $x \in \mathbb{R}^p$, this sequence can be used as penalty parameters for (13) in the following gradient algorithm:

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

where $\{\gamma_n\}_{n\geq 1}$ is a sequence of positive reals.

However, since the gradient is typically unavailable, we may instead use (10) to obtain the estimate of $\nabla\psi$, given by $h'(g(x, Y))\nabla_x g(x, Y)$, and it is then quite natural to use the following stochastic gradient algorithm to search for the minima of the function f over D:

$$X_{n+1} = X_n - \gamma_{n+1}(\nabla f(X_n) + \delta_{n+1}h'(g(X_n, Y_{n+1}))\nabla_x g(X_n, Y_{n+1})), \;\; n \geq 0, \tag{14}$$

where $\{\gamma_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ have the same meaning as in the case of the algorithm (11).

The following assumptions are required in the analysis of the algorithm (14):

B1 $\gamma_n > 0$ *for each* $n \geq 1$, $\sum_{n=1}^{\infty} \gamma_n = \infty$, *and* $\sum_{n=1}^{\infty} \gamma_n^2 \delta_n^2 < \infty$.

B2 f *is convex and* ∇f *is locally Lipschitz continuous.*

B3 *h is convex and $g(\cdot, y)$ is convex for each $y \in \mathbb{R}^q$. For all $\rho \in [1, \infty)$, there exists a Borel-measurable function $\phi_\rho : \mathbb{R}^q \to [1, \infty)$ and such that*

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

and, for all $x, x', x'' \in B_\rho^p$, $y \in \mathbb{R}^q$,

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

B4 *$\eta^* := \inf_{x \in D} f(x) > -\infty$, and the set of optimizers given by $D^* := \{x \in D : f(x) = \eta^*\}$ is non-empty.*

Assumption B3 ensures that ψ is well-defined, finite, convex and differentiable. It also implies that $\nabla\psi$ is locally Lipschitz continuous. Assumption B4 is satisfied if D is bounded or f is coercive (i.e. the sublevel set $\{x : f(x) \leq N\}$ is bounded for each $N \geq 1$).

Theorem 4. *Let $\{X_n\}_{n \geq 0}$ be generated by (14), and suppose that Assumptions B1–B4 hold. Then, on the event $\{\sup_{0 \leq n} \|X_n\| < \infty\}$,*

$$\lim_{n \to \infty} d(X_n, D^*) = 0 \quad \textit{w.p.1.}$$

A proof of Theorem 4 is presented in Section 4 of Appendix.

4 Conclusion

The main contribution of this paper is to reformulate a given semi-infinite program as a stochastic optimization problem. One can then apply Monte Carlo and stochastic approximation methods to generate efficient algorithms, and provide a foundation for analysis. Under standard convexity assumptions, and additional relatively mild conditions, the proposed algorithms provide a convergent solution with probability one.

The next step is to test these algorithms in practical, non-trivial applications. In particular, we are interested in application to specific optimization problems, and to robust control. It is of interest to see how these randomization approaches compare to their deterministic counterparts. We also expect that the algorithms may be refined and improved within a particular application context.

Acknowledgements

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.

Appendix

Here we provide proofs of the main results of the paper. The following notation is fixed in this Appendix: Let $\mathcal{F}_0 = \sigma\{X_0\}$, while $\mathcal{F}_n = \sigma\{X_0, Y_1, \dots, Y_n\}$ for $n \geq 1$.

Proof of Theorem 2

We begin with the representation,

$$\begin{aligned} X_{n+1} &= X_n - \gamma_{n+1}\nabla\psi(X_n) + \xi_{n+1}, \\ \psi(X_{n+1}) &= \psi(X_n) - \gamma_{n+1}\|\nabla\psi(X_n)\|^2 + \varepsilon_{n+1}, \quad n \geq 1, \end{aligned} \tag{15}$$

where the error terms in (15) are defined as,

$$\xi_{n+1} = \gamma_{n+1}(\nabla\psi(X_n) - h'(g(X_n, Y_{n+1})\nabla_x g(X_n, Y_{n+1})),$$

$$\begin{aligned} \varepsilon_{1,n+1} &:= (\nabla\psi(X_n))^T \xi_{n+1}, \\ \varepsilon_{2,n+1} &:= \int_0^1 (\nabla\psi(X_n + t(X_{n+1} - X_n)) - \nabla\psi(X_n))^T (X_{n+1} - X_n) dt, \\ \varepsilon_{n+1} &:= \varepsilon_{1,n+1} + \varepsilon_{2,n+1}, \quad n \geq 0. \end{aligned}$$

The first step in our analysis of (15) is to establish the asymptotic properties of $\{\xi_n\}_{n\geq 1}$ and $\{\varepsilon_n\}_{n\geq 1}$:

Lemma 1. *Suppose that Assumptions A1 and A2 hold. Then,* $\sum_{n=1}^{\infty} \xi_n$, $\sum_{n=1}^{\infty} \varepsilon_n$ *converge w.p.1 on the event* $\{\sup_{0\leq n} \|X_n\| < \infty\}$.

Proof. Fix $\rho \in [1, \infty)$, and let $K_\rho < \infty$ serve as an upper bound on $\|\nabla\psi\|$, and a Lipschitz constant for $\nabla\psi$ on the set B_ρ^p. Due to A1,

$$\begin{aligned} \|\xi_{n+1}\| I_{\{\|X_n\|\leq\rho\}} &\leq 2K_\rho\gamma_{n+1}\phi_\rho^2(Y_{n+1}) \\ |\varepsilon_{1,n+1}| I_{\{\|X_n\|\leq\rho\}} &\leq K_\rho \|\xi_{n+1}\| I_{\{\|X_n\|\leq\rho\}}, \\ |\varepsilon_{2,n+1}| I_{\{\|X_n\|\leq\rho\}} &\leq K_\rho \|X_{n+1} - X_n\|^2 I_{\{\|X_n\|\leq\rho\}} \\ &\leq 2K_\rho^3\gamma_{n+1}^2 + 2K_\rho\|\xi_{n+1}\| I_{\{\|X_n\|\leq\rho\}}, \quad n \geq 0. \end{aligned}$$

Consequently,

$$\mathsf{E}\left(\sum_{n=0}^{\infty} \|\xi_{n+1}\|^2 I_{\{\|X_n\|\leq\rho\}}\right) \leq 4K_\rho^2 \sum_{n=1}^{\infty} \gamma_n^2 \mathsf{E}(\phi_\rho^4(Y_n)) < \infty, \tag{16}$$

$$\mathsf{E}\left(\sum_{n=0}^{\infty} |\varepsilon_{1,n+1}|^2 I_{\{\|X_n\|\leq\rho\}}\right) \leq K_\rho^2 \mathsf{E}\left(\sum_{n=0}^{\infty} \|\xi_{n+1}\|^2 I_{\{\|X_n\|\leq\rho\}}\right) < \infty, \tag{17}$$

$$\begin{aligned} \mathsf{E}\left(\sum_{n=0}^{\infty} |\varepsilon_{2,n+1}|^2 I_{\{\|X_n\|\leq\rho\}}\right) \leq & 2K_\rho \mathsf{E}\left(\sum_{n=0}^{\infty} \|\xi_{n+1}\|^2 I_{\{\|X_n\|\leq\rho\}}\right) \\ & + 2K_\rho^3 \sum_{n=1}^{\infty} \gamma_n^2 < \infty. \end{aligned} \tag{18}$$

Since X_n is measurable with respect to $\mathcal{F}_n$ and independent of Y_{n+1}, we have

$$\mathsf{E}\left(\xi_{n+1}\|\xi_{n+1}\|^2 I_{\{\|X_n\|\leq\rho\}}|\mathcal{F}_n\right) = 0 \ \textit{w.p.1.},$$
$$\mathsf{E}\left(\varepsilon_{1,n+1}\|\xi_{n+1}\|^2 I_{\{\|X_n\|\leq\rho\}}|\mathcal{F}_n\right) = 0 \ \textit{w.p.1.}, \quad n \geq 0.$$

Then, Doob's martingale convergence theorem (see e.g., [12]) and (16)–(18) imply that $\sum_{n=1}^{\infty}\xi_n$, $\sum_{n=1}^{\infty}\varepsilon_{1,n}$, $\sum_{n=1}^{\infty}\varepsilon_{2,n}$ converge w.p.1 on the event $\{\sup_{0\leq n}\|X_n\| \leq \rho\}$. Since ρ can be arbitrary large, it can easily be deduced that $\sum_{n=1}^{\infty}\xi_n$, $\sum_{n=1}^{\infty}\varepsilon_n$ converge w.p.1 on the event $\{\sup_{0\leq n}\|X_n\| < \infty\}$.

Proof of Theorem 2. We again fix $\rho \in [1, \infty)$, and let $K_\rho < \infty$ serve as an upper bound on $\|\nabla\psi\|$, and a Lipschitz constant for $\nabla\psi$ on the set B_ρ^p. Fix an arbitrary sample $\omega \in \Omega$ from the event on which $\sup_{0\leq n}\|X_n\| \leq \rho$, and both $\sum_{n=1}^{\infty}\xi_n$ and $\sum_{n=1}^{\infty}\varepsilon_n$ are convergent (for the sake of notational simplicity, ω does not explicitly appear in the relations which follow in the proof). Due to Lemma 1 and the fact that $\nabla\psi$ is continuous, it is sufficient to show $\lim_{n\to\infty}\|\nabla\psi(X_n)\| = 0$. We proceed by contradiction.

If $\|\nabla\psi(X_n)\|$ does not vanish as $n \to \infty$, then we may find $\varepsilon > 0$ such that

$$\limsup_{n\to\infty}\|\nabla\psi(X_n)\| > 2\varepsilon. \tag{19}$$

On the other hand, (15) yields

$$\sum_{i=0}^{n-1}\gamma_{i+1}\|\nabla\psi(X_i)\|^2 = \psi(X_0) - \psi(X_n) + \sum_{i=1}^{n}\xi_i \leq K_\rho + \sum_{i=1}^{n}\xi_i, \quad n \geq 1.$$

Consequently,

$$\sum_{n=0}^{\infty}\gamma_{n+1}\|\nabla\psi(X_n)\|^2 < \infty, \tag{20}$$

and A1 then implies that $\liminf_{n\to\infty}\|\nabla\psi(X_n)\| = 0$. Otherwise, there would exist $\delta \in (0,\infty)$ and $j_0 \geq 1$ (both depending on ω) such that $\|\nabla\psi(X_n)\| \geq \delta$, $n \geq j_0$, which combined with A1 would yield

$$\sum_{n=0}^{\infty}\gamma_{n+1}\|\nabla\psi(X_n)\|^2 \geq \delta^2\sum_{n=j_0+1}^{\infty}\gamma_n = \infty.$$

Let $m_0 = n_0 = 0$ and

$$m_{k+1} = \{n \geq n_k : \|\nabla\psi(X_n)\| \geq 2\varepsilon\},$$
$$n_{k+1} = \{n \geq m_{k+1} : \|\nabla\psi(X_n)\| \leq \varepsilon\}, \quad k \geq 0.$$

Obviously, $\{m_n\}_{k\geq 0}$, $\{n_k\}_{k\geq 0}$ are well-defined, finite, and satisfy $m_k < n_k < m_{k+1}$ for $k \geq 1$. Moreover,

$$\|\nabla\psi(X_{m_k})\| \geq 2\varepsilon, \quad \|\nabla\psi(X_{n_k})\| \leq \varepsilon, \quad k \geq 1, \tag{21}$$

and

$$\|\nabla\psi(X_n)\| \geq \varepsilon, \quad \text{for } m_k \leq n < n_k,\ k \geq 1. \tag{22}$$

Due to (20), (22),

$$\varepsilon^2 \sum_{k=1}^{\infty} \sum_{i=m_k}^{n_k-1} \gamma_{i+1} \leq \sum_{k=1}^{\infty} \sum_{i=m_k}^{n_k-1} \gamma_{i+1} \|\nabla\psi(X_i)\|^2 \leq \sum_{n=0}^{\infty} \gamma_{n+1} \|\nabla\psi(X_n)\|^2 < \infty.$$

Therefore,

$$\lim_{k\to\infty} \sum_{i=m_k+1}^{n_k} \gamma_i = 0, \tag{23}$$

while (15) yields, for each $k \geq 1$,

$$\begin{aligned}
\varepsilon \leq \|\nabla\psi(X_{n_k}) - \nabla\psi(X_{m_k})\| &\leq K_\rho \|X_{n_k} - X_{m_k}\| \\
&= K_\rho \left\| -\sum_{i=m_k}^{n_k-1} \gamma_{i+1} \nabla\psi(X_i) + \sum_{i=m_k+1}^{n_k} \xi_i \right\| \\
&\leq K_\rho^2 \sum_{i=m_k+1}^{n_k} \gamma_i + \left\| \sum_{i=m_k+1}^{n_k} \xi_i \right\|.
\end{aligned} \tag{24}$$

However, this is not possible, since (23) and the limit process $k \to \infty$ applied to (24) yield $\varepsilon \leq 0$. Hence, $\lim_{n\to\infty} \|\nabla\psi(X_n)\| = 0$. This completes the proof.

Proof of Theorem 3

Let $C = D \cap Q$, and let $\Pi_C(\cdot)$ denote the projection operator onto the set C. Moreover, the sequence $\{\xi_n\}_{n\geq 0}$ has the same meaning as in the previous section, while

$$\begin{aligned}
Z_{n+1} &= X_n - \gamma_{n+1} h'(g(X_n, Y_{n+1})) \nabla_x g(X_n, Y_{n+1}), \\
\varepsilon_{1,n+1} &= 2(X_n - \Pi_C(X_n))^T \xi_{n+1}, \\
\varepsilon_{2,n+1} &= \|Z_{n+1} - X_n\|^2, \\
\varepsilon_{n+1} &= \varepsilon_{1,n+1} + \varepsilon_{2,n+1}, \quad n \geq 0.
\end{aligned}$$

Since ψ is convex (under the conditions of Theorem 3) and $\Pi_C(\cdot)$, $\Pi_Q(\cdot)$ are non-expansive, we have

$$\begin{aligned}
(X_n - \Pi_C(X_n))^T \nabla\psi(X_n) &\geq \ \psi(X_n) - \psi(\Pi_C(X_n)) = \psi(X_n), \\
\|X_{n+1} - \Pi_C(X_{n+1})\| &\leq \|X_{n+1} - \Pi_C(X_n)\| \\
&= \|\Pi_Q(Z_{n+1}) - \Pi_Q(\Pi_C(X_n))\| \\
&\leq \|Z_{n+1} - \Pi_C(X_n)\|
\end{aligned}$$

for $n \geq 0$. Then, it is straightforward to demonstrate that for all $n \geq 0$,

$$Z_{n+1} = X_n - \gamma_{n+1} \nabla \psi(X_n) + \xi_{n+1}, \tag{25}$$

and moreover,

$$\begin{aligned}
\|X_{n+1} - &\Pi_C(X_{n+1})\|^2 \\
&\leq \|(X_n - \Pi_C(X_n)) + (Z_{n+1} - X_n)\|^2 \\
&= \|X_n - \Pi_C(X_n)\|^2 + 2(X_n - \Pi_C(X_n))^T (Z_{n+1} - X_n) \\
&\quad + \|Z_{n+1} - X_n\|^2 \\
&= \|X_n - \Pi_C(X_n)\|^2 - 2\gamma_{n+1}(X_n - \Pi_C(X_n))^T \nabla \psi(X_n) + \varepsilon_{n+1} \\
&\leq \|X_n - \Pi_C(X_n)\|^2 - 2\gamma_{n+1}\psi(X_n) + \varepsilon_{n+1}.
\end{aligned} \tag{26}$$

Lemma 2. *Suppose that Assumptions A1 and A2 hold. Then,* $\sum_{n=1}^{\infty} \xi_n$, $\sum_{n=1}^{\infty} \varepsilon_n$ *converge w.p.1.*

Proof. Let $K \in [1, \infty)$ denote an upper bound of $\|\cdot\|$, $\|\Pi_C(\cdot)\|$, $\|\nabla \psi\|$ on Q. Due to A2,

$$\begin{aligned}
\|\xi_{n+1}\| &\leq 2K\phi_K^2(Y_{n+1}), \\
|\varepsilon_{1,n+1}| &\leq 4K\|\xi_{n+1}\|, \\
|\varepsilon_{2,n+1}| &\leq 2K^2\gamma_{n+1}^2 + 2K\|\xi_{n+1}\|^2, \quad n \geq 0,
\end{aligned}$$

and this implies the following bounds,

$$\begin{aligned}
\mathsf{E}\left(\sum_{n=1}^{\infty} \|\xi_n\|^2\right) &\leq 4K^2 \sum_{n=1}^{\infty} \gamma_n^2 \mathsf{E}(\phi_K^2(Y_n)) < \infty, \\
\mathsf{E}\left(\sum_{n=1}^{\infty} \|\varepsilon_{1,n}\|^2\right) &\leq 16K^2 \mathsf{E}\left(\sum_{n=1}^{\infty} \|\xi_n\|^2\right) < \infty, \\
\mathsf{E}\left(\sum_{n=1}^{\infty} \|\varepsilon_{2,n}\|^2\right) &\leq 2K^2 \sum_{n=1}^{\infty} \gamma_n^2 + 2\mathsf{E}\left(\sum_{n=1}^{\infty} \|\xi_n\|^2\right) < \infty.
\end{aligned}$$

Then, using the same arguments as in the proof of Lemma 1, it can easily be deduced that $\sum_{n=1}^{\infty} \xi_n$, $\sum_{n=1}^{\infty} \varepsilon_n$ converge *w.p.1.*

Proof of Theorem 3. Let $K_\rho \in [1, \infty)$ denote a simultaneous upper bound for $\|\cdot\|$, $\|\Pi_C(\cdot)\|$, $\|\nabla \psi\|$ on the set Q, and a Lipschitz constant for ψ on the same set. Moreover, let ω be an arbitrary sample from the event where $\sum_{n=1}^{\infty} \xi_n$, $\sum_{n=1}^{\infty} \varepsilon_n$ converge (for the sake of notational simplicity, ω does not explicitly appear in the relations which follow in the proof). Due to Lemma 2 and the fact that ψ is continuous and strictly positive on $Q \backslash D$, it is sufficient to show $\lim_{n\to\infty} \psi(X_n) = 0$. Suppose the opposite. Then, there exists $\varepsilon \in (0, \infty)$ (depending on ω) such that $\limsup_{n\to\infty} \psi(X_n) > 2\varepsilon$. On the other hand, (26) yields

$$\begin{aligned}\sum_{i=0}^{n-1}\gamma_{i+1}\psi(X_i)\leq&\|X_0-\Pi_C(X_0)\|^2-\|X_n-\Pi_C(X_n)\|^2\\&+\sum_{i=1}^{n}\xi_i\leq K_\rho+\sum_{i=1}^{n}\xi_i,\quad n\geq 1.\end{aligned}$$

Consequently,

$$\sum_{n=0}^{\infty}\gamma_{n+1}\psi(X_n)<\infty. \tag{27}$$

Then, A1 implies $\liminf_{n\to\infty}\psi(X_n)=0$. Otherwise, there would exist $\delta\in(0,\infty)$ and $j_0\geq 1$ (both depending on ω) such that $\psi(X_n)\geq\delta$, $n\geq j_0$, which combined with A1 would yield

$$\sum_{n=0}^{\infty}\gamma_{n+1}\psi(X_n)\geq\delta\sum_{n=j_0+1}^{\infty}\gamma_n=\infty.$$

Let $m_0=n_0=0$ and

$$\begin{aligned}m_{k+1}&=\{n\geq n_k:\psi(X_n)\geq 2\varepsilon\},\\n_{k+1}&=\{n\geq m_{k+1}:\psi(X_n)\leq\varepsilon\}\end{aligned}$$

for $k\geq 0$. Obviously, $\{m_n\}_{k\geq 0}$, $\{n_k\}_{k\geq 0}$ are well-defined, finite and satisfy $m_k<n_k<m_{k+1}$ for $k\geq 1$. Moreover,

$$\psi(X_{m_k})\geq 2\varepsilon,\quad\psi(X_{n_k})\leq\varepsilon \tag{28}$$

for $k\geq 1$, and

$$\psi(X_n)\geq\varepsilon,\quad\text{for } m_k\leq n<n_k,\ k\geq 0. \tag{29}$$

Due to (27), (29),

$$\varepsilon^2\sum_{k=1}^{\infty}\sum_{i=m_k}^{n_k-1}\gamma_{i+1}\leq\sum_{k=1}^{\infty}\sum_{i=m_k}^{n_k-1}\gamma_{i+1}\psi(X_i)\leq\sum_{n=0}^{\infty}\gamma_{n+1}\psi(X_n)<\infty.$$

Therefore,

$$\lim_{k\to\infty}\sum_{i=m_k+1}^{n_k}\gamma_i=0, \tag{30}$$

while (25), (28) yield

$$\begin{aligned}\varepsilon\leq|\psi(X_{n_k})-\psi(X_{m_k})|\leq&K\|X_{n_k}-X_{m_k}\|\\&=K\left\|-\sum_{i=m_k}^{n_k-1}\gamma_{i+1}\nabla\psi(X_i)+\sum_{i=m_k+1}^{n_k}\xi_i\right\|\\&\leq K\sum_{i=m_k+1}^{n_k}\gamma_i+\left\|\sum_{i=m_k+1}^{n_k}\xi_i\right\|\end{aligned} \tag{31}$$

for $k \geq 1$. However, this is not possible, since (30) and the limit process $k \to \infty$ applied to (31) yield $\varepsilon \leq 0$. Hence, $\lim_{n\to\infty} \|\nabla\psi(X_n)\| = 0$. This completes the proof.

Proof of Theorem 4

Let $f_n(x) = f(x) + \delta_n\psi(x)$ for $x \in \mathbb{R}^p$, $n \geq 1$, while $\Pi_{D^*}(\cdot)$ denotes the projection on the set D^* (i.e., $\Pi_{D^*}(x) = \arg\inf_{x'\in D^*} \|x - x'\|$ for $x \in \mathbb{R}^p$). The following error sequences are defined for $n \geq 0$,

$$\begin{aligned}
\xi_{n+1} &= \gamma_{n+1}\delta_{n+1}(\nabla\psi(X_n) - h'(g(X_n, Y_{n+1}))\nabla_x g(X_n, Y_{n+1})),\\
\varepsilon_{1,n+1} &= 2(X_n - \Pi_{D^*}(X_n))^T\xi_{n+1},\\
\varepsilon_{2,n+1} &= \|X_{n+1} - X_n\|^2,\\
\varepsilon_{n+1} &= \varepsilon_{1,n+1} + \varepsilon_{2,n+1}\,.
\end{aligned}$$

It is straightforward to verify that the following recursion holds for $n \geq 0$,

$$X_{n+1} = X_n - \gamma_{n+1}\nabla f_n(X_n) + \xi_{n+1},$$

and this implies the following bounds,

$$\begin{aligned}
&\|X_{n+1} - \Pi_{D^*}(X_{n+1})\|^2\\
&\quad\leq \|X_{n+1} - \Pi_{D^*}(X_n)\|^2\\
&\quad= \|X_n - \Pi_{D^*}(X_n)\|^2 + 2(X_n - \Pi_{D^*}(X_n))^T(X_{n+1} - X_n)\\
&\qquad + \|X_{n+1} - X_n\|^2\\
&\quad= \|X_n - \Pi_{D^*}(X_n)\|^2\\
&\qquad - 2\gamma_{n+1}(X_n - \Pi_{D^*}(X_n))^T\nabla f_n(X_n) + \varepsilon_{n+1}, \quad n \geq 0.
\end{aligned}$$

Lemma 3. *Suppose that Assumptions B3 and B4 hold. Moreover, let $\{x_k\}_{k\geq 0}$ be a bounded sequence from $\mathbb{R}^p$, while $\{n_k\}_{k\geq 0}$ is an increasing sequence of positive integers. Suppose that*

$$\liminf_{k\to\infty} \|x_k - \Pi_{D^*}(x_k)\| > 0.$$

Then,

$$\liminf_{k\to\infty} (x_k - \Pi_{D^*}(x_k))^T\nabla f_{n_k}(x_k) > 0.$$

Proof. Since $\{f_{n_k}\}_{k\geq 0}$ are convex and $f_{n_k}(\Pi_{D^*}(x_k)) = f(\Pi_{D^*}(x_k)) = \eta^*$ for $k \geq 0$, we have

$$(x_k - \Pi_{D^*}(x_k))^T\nabla f_{n_k}(x_k) \geq f_{n_k}(x_k) - f(\Pi_{D^*}(x_k)) = f_{n_k}(x_k) - \eta^*, \quad k \geq 0.$$

Therefore, it is sufficient to show that $\liminf_{k\to\infty} f_{n_k}(x_k) > \eta^*$. Suppose the opposite. Then, there exists $\varepsilon \in (0, \infty)$, $\tilde{x} \in \mathbb{R}^p$ and a subsequence $\{\tilde{x}_k, \tilde{n}_k\}_{k\geq 0}$

of $\{x_k, n_k\}_{k\geq 0}$ such that $\lim_{k\to\infty} \tilde{x}_k = \tilde{x}$, $\limsup_{k\to\infty} f_{\tilde{n}_k}(\tilde{x}_k) \leq \eta^*$ and $\|\tilde{x}_k - \Pi_{D^*}(\tilde{x}_k)\| \geq \varepsilon$ for $k \geq 0$. Consequently,

$$\begin{aligned} f(\tilde{x}) &= \lim_{k\to\infty} f(\tilde{x}_k) \leq \limsup_{k\to\infty} f_{\tilde{n}_k}(\tilde{x}_k) \leq \eta^*, \\ d(\tilde{x}, D^*) &= \|\tilde{x} - \Pi_{D^*}(\tilde{x})\| = \lim_{k\to\infty} \|\tilde{x}_k - \Pi_{D^*}(\tilde{x}_k)\| \geq \varepsilon. \end{aligned} \tag{32}$$

Then, it can easily be deduced that $\tilde{x} \notin D$ (otherwise, (32) would imply $\tilde{x} \in D^*$). Therefore,

$$\lim_{k\to\infty} f_{\tilde{n}_k}(\tilde{x}_k) \geq \lim_{k\to\infty} \delta_{\tilde{n}_k} \psi(\tilde{x}_k) = \infty > \eta^*.$$

However, this is not possible. Hence, $\liminf_{k\to\infty} f_{n_k}(x_k) > \eta^*$. This completes the proof.

Lemma 4. *Suppose that Assumptions B1 and B4 hold. Then,* $\lim_{n\to\infty} \|X_{n+1} - X_n\| = 0$ *and* $\sum_{n=1}^{\infty} \varepsilon_n$ *converges w.p.1 on the event* $\{\sup_{0\leq n} \|X_n\| < \infty\}$.

Proof. Let $\rho \in [1, \infty)$, while $K_\rho \in [\rho, \infty)$ denotes an upper bound of $\|\Pi_{D^*}(\cdot)\|$, $\|\nabla\psi\|$, ∇f on B^p_ρ. Due to B4,

$$\begin{aligned} \|\xi_{n+1}\| I_{\{\|X_n\|\leq\rho\}} &\leq 2K_\rho \gamma_{n+1} \delta_{n+1} \phi^2_\rho(Y_{n+1}), \\ |\varepsilon_{1,n+1}| I_{\{\|X_n\|\leq\rho\}} &\leq 4K_\rho \|\xi_{n+1}\| I_{\{\|X_n\|\leq\rho\}}, \\ |\varepsilon_{2,n+1}| I_{\{\|X_n\|\leq\rho\}} &\leq 2K^2_\rho \gamma^2_{n+1} + 2\|\xi_{n+1}\|^2 I_{\{\|X_n\|\leq\rho\}}, \\ \|X_{n+1} - X_n\| I_{\{\|X_n\|\leq\rho\}} &\leq K_\rho \gamma_{n+1}(1 + \delta_{n+1}) + \|\xi_{n+1}\| I_{\{\|X_n\|\leq\rho\}} \end{aligned} \tag{33}$$

for $k \geq 0$. Consequently,

$$\mathsf{E}\left(\sum_{n=0}^{\infty} \|\xi_{n+1}\|^2 I_{\{\|X_n\|\leq\rho\}}\right) \leq 4K^2_\rho \sum_{n=1}^{\infty} \gamma^2_n \delta^2_n \mathsf{E}(\phi^4_\rho(Y_n)) < \infty \tag{34}$$

$$\mathsf{E}\left(\sum_{n=0}^{\infty} |\varepsilon_{1,n+1}|^2 I_{\{\|X_n\|\leq\rho\}}\right) \leq 16K^2_\rho \mathsf{E}\left(\sum_{n=0}^{\infty} \|\xi_{n+1}\|^2 I_{\{\|X_n\|\leq\rho\}}\right) < \infty, \tag{35}$$

$$\begin{aligned} \mathsf{E}\left(\sum_{n=0}^{\infty} |\varepsilon_{2,n+1}|^2 I_{\{\|X_n\|\leq\rho\}}\right) \leq & 2\mathsf{E}\left(\sum_{n=0}^{\infty} \|\xi_{n+1}\|^2 I_{\{\|X_n\|\leq\rho\}}\right) \\ & + 2K^2_\rho \sum_{n=1}^{\infty} \gamma^2_n (1 + \delta^2_n) < \infty. \end{aligned} \tag{36}$$

Owing to B1 and (33), (34), $\lim_{n\to\infty} \|X_{n+1} - X_n\| = 0$ w.p.1 on the event $\{\sup_{0\leq n} \|X_n\| \leq \rho\}$. On the other hand, using (35), (36) and the same arguments as in the proof of Lemma 16, it can be demonstrated that $\sum_{n=1}^{\infty} \varepsilon_n$ converges w.p.1 on the same event. Then, we can easily deduce that w.p.1 on the event $\{\sup_{0\leq n} \|X_n\| < \infty\}$, $\lim_{n\to\infty} \|X_{n+1} - X_n\| = 0$ and $\sum_{n=1}^{\infty} \varepsilon_n$ converges. This completes the proof.

Proof of Theorem 4. Let $\rho \in [1, \infty)$, while $K_\rho \in [\rho, \infty)$ denotes an upper bound of $\|\Pi_{D^*}(\cdot)\|$ on B_ρ^p. Moreover, let ω be an arbitrary sample from the event where $\sup_{0 \leq n} \|X_n\| \leq \rho$, $\lim_{n\to\infty} \|X_{n+1} - X_n\| = 0$ and $\sum_{n=1}^{\infty} \varepsilon_n$ converges (for the sake of notational simplicity, ω does not explicitly appear in the relations which follow in the proof). Due to Lemma 4, it is sufficient to show $\lim_{n\to\infty} \|X_n - \Pi_{D^*}(X_n)\| = 0$. Suppose the opposite. Then, there exists $\varepsilon \in (0, \infty)$ (depending on ω) such that $\limsup_{n\to\infty} \|X_n - \Pi_{D^*}(X_n)\| > 2\varepsilon$. On the other hand, (32) yields

$$\begin{aligned} 2\sum_{i=0}^{n-1} &\gamma_{i+1}(X_i - \Pi_{D^*}(X_i))^T \nabla f_i(X_i) \\ &\leq \|X_0 - \Pi_{D^*}(X_0)\|^2 - \|X_n - \Pi_{D^*}(X_n)\|^2 + \sum_{i=1}^{n} \varepsilon_i \qquad (37) \\ &\leq 4K_\rho^2 + \sum_{i=1}^{n} \varepsilon_i, \quad n \geq 1. \end{aligned}$$

Since $\liminf_{n\to\infty}(X_n - \Pi_{D^*}(X_n))^T \nabla f_n(X_n) > 0$ results from $\liminf_{n\to\infty} \|X_n - \Pi_{D^*}(X_n)\| > 0$ (due to Lemma 3), B1 and (37) imply that $\liminf_{n\to\infty} \|X_n - \Pi_{D^*}(X_n)\| = 0$. Otherwise, there would exist $\delta \in (0, \infty)$ and $j_0 \geq 1$ (both depending on ω) such that $(X_n - \Pi_{D^*}(X_n))^T \nabla f_n(X_n) \geq \delta$, $n \geq j_0$, which combined with B1 would yield

$$\begin{aligned} &\sum_{n=0}^{\infty} \gamma_{n+1}(X_n - \Pi_{D^*}(X_n))^T \nabla f_n(X_n) \\ &\geq \sum_{n=0}^{j_0} \gamma_{n+1}(X_n - \Pi_{D^*}(X_n))^T \nabla f_n(X_n) + \delta \sum_{n=j_0+1}^{\infty} \gamma_n = \infty. \end{aligned}$$

Let $l_0 = \inf\{n \geq 0 : \|X_n - \Pi_{D^*}(X_n)\| \leq \varepsilon\}$, and define for $k \geq 0$,

$$\begin{aligned} n_k &= \inf\{n \geq l_k : \|X_n - \Pi_{D^*}(X_n)\| \geq 2\varepsilon\}, \\ m_k &= \sup\{n \leq n_k : \|X_n - \Pi_{D^*}(X_n)\| \leq \varepsilon\}, \\ l_{k+1} &= \inf\{n \geq n_k : \|X_n - \Pi_{D^*}(X_n)\| \leq \varepsilon\}. \end{aligned}$$

Obviously, $\{l_k\}_{k\geq 0}$, $\{m_k\}_{k\geq 0}$, $\{n_k\}_{k\geq 0}$ are well-defined, finite and satisfy $l_k \leq m_k < n_k < l_{k+1}$ for $k \geq 0$. Moreover,

$$\|X_{m_k} - \Pi_{D^*}(X_{m_k})\| \leq \varepsilon, \quad \|X_{n_k} - \Pi_{D^*}(X_{n_k})\| \geq 2\varepsilon, \quad k \geq 0, \qquad (38)$$

and

$$\|X_n - \Pi_{D^*}(X_n)\| \geq \varepsilon \quad \text{for } m_k < n \leq n_k,\ k \geq 0. \qquad (39)$$

Due to (38), (39) and the fact that $\Pi_{D^*}(\cdot)$ is non-expansive (see e.g., [3]),

$$\begin{aligned}\varepsilon \leq \|X_{m_k+1} - \Pi_{D^*}(X_{m_k+1})\| \leq& \varepsilon + \|X_{m_k+1} - \Pi_{D^*}(X_{m_k+1})\| \\ &- \|X_{m_k} - \Pi_{D^*}(X_{m_k})\| \\ \leq& \varepsilon + \|(X_{m_k+1} - X_{m_k}) \\ &- (\Pi_{D^*}(X_{m_k+1}) - \Pi_{D^*}(X_{m_k}))\| \\ \leq& \varepsilon + 2\|X_{m_k+1} - X_{m_k}\|, \quad k \geq 0.\end{aligned}$$

Therefore, $\lim_{k\to\infty} \|X_{m_k+1} - \Pi_{D^*}(X_{m_k+1})\| = \varepsilon$. Then, (39) yields

$$\limsup_{k\to\infty}(\|X_{n_k} - \Pi_{D^*}(X_{n_k})\|^2 - \|X_{m_k+1} - \Pi_{D^*}(X_{m_k+1})\|^2) \geq \varepsilon^2. \tag{40}$$

On the other hand, Lemma 3 and (39) imply

$$\liminf_{k\to\infty} \min_{m_k<n\leq n_k} (X_n - \Pi_{D^*}(X_n))^T \nabla f_n(X_n) > 0.$$

Consequently, there exists $k_0 \geq 0$ (depending on ω) such that

$$\sum_{i=m_k+1}^{n_k-1} \gamma_{i+1}(X_i - \Pi_{D^*}(X_i))^T \nabla f_i(X_i) \geq 0 \tag{41}$$

for $k \geq k_0$. Owing to (32), (41),

$$\begin{aligned}&\|X_{n_k} - \Pi_{D^*}(X_{n_k})\|^2 - \|X_{m_k+1} - \Pi_{D^*}(X_{m_k+1})\|^2 \\ &\leq \sum_{m_k+1}^{n_k-1} \gamma_{i+1}(X_i - \Pi_{D^*}(X_i))^T \nabla f_i(X_i) + \sum_{i=m_k+2}^{n_k} \varepsilon_i \leq \sum_{i=m_k+2}^{n_k} \varepsilon_i\end{aligned} \tag{42}$$

for $k \geq k_0$. However, this is not possible, since (40) and the limit process $k \to \infty$ applied to (41) yield $\varepsilon^2 \leq 0$. Hence, $\lim_{n\to\infty} \|X_n - \Pi_{D^*}(X_n)\| = 0$. This completes the proof.

References

1. D. Bertsekas and J. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996.
2. A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximation*, Springer Verlag, 1990.
3. D. P. Bertsekas, *Nonlinear Programming*, 2nd Edition, Athena Scientific, 1999.
4. K. Binder, *Monte Carlo Methods in Statistical Physics*, Springer-Verlag, 1986.
5. V. S. Borkar and S. P. Meyn, "The O.D.E. method for convergence of stochastic approximation and reinforcement learning," *SIAM J. Control Optim.*, vol. 38, pp. 447–469, 2000.
6. P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer Verlag, 2003.
7. M. A. Goberna and M. A. Lopez (Eds.), *Semi-Infinite Programming: Recent Advances*, Kluwer Academic, 2001.

8. R. Hettich and K. O. Kortanek, “Semi-infinite programming: theory, methods, and applications,” *SIAM Review*, vol. 35, pp. 380–429, 1993.
9. H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer Verlag, 2003.
10. D. Liberzon and R. Tempo, “Common Lyapunov Functions and Gradient Algorithms,” *IEEE Transactions on Automatic Control*, vol. 49, 2004 (in press).
11. L. Ljung, G. Pflug and H. Walk, *Stochastic Approximation and Optimization of Random Systems*, Birkhäuser Verlag, 1992.
12. J. Neveu, *Discrete Parameter Martingales*, North Holland, 1975.
13. R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.
14. K. Mulmuley, *Computational Geometry: An Introduction through Randomization Algorithms*, Prentice-Hall, 1994.
15. E. Polak, *Optimization: Algorithms and Consistent Approximations*, Springer Verlag, 1997.
16. 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.
17. B. T. Polyak and R. Tempo, “Probabilistic robust design with linear quadratic regulator,” *System and Control Letters*, vol. 43, pp. 343–353, 2001.
18. R. Reemtsen and J.-J. Rückmann (Eds.), *Semi-Infinite Programming*, Kluwer Academic, 1998.
19. R. Reemtsen, “Numerical methods for semi-infinite programming: A survey,” in R. Reemtsen and J.-J. Rückmann (Eds.), *Semi-Infinite Programming*, Kluwer Academic, 1998.
20. C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Verlag, 1999.
21. G. O. Roberts and J. S. Rosenthal, “General state space Markov chains and MCMC algorithms,” submitted for publication, 2004.
22. K.K. Sabelfeld and N.A. Simonov, *Random Walks on Boundary for Solving PDEs*, VSP Intl. Science Publishers, 1994.
23. V. B. Tadić, “Stochastic Approximation with Random Truncations, State Dependent Noise and Discontinuous Dynamics,” *Stochastics and Stochastics Reports*, vol. 64, pp. 283–325, 1998.
24. R. Tempo, G. Calafiore and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems*, Springer Verlag, 2004 (in press).
25. J. F. Traub and A. G. Weschultz, *Complexity and Information*, Cambridge University Press, 1998.