



Stochastic proximal gradient methods for nonconvex problems in Hilbert spaces

Caroline Geiersbach¹ · Teresa Scarinci²

Received: 5 January 2020 / Accepted: 17 December 2020 / Published online: 12 January 2021
© The Author(s) 2021

Abstract

For finite-dimensional problems, stochastic approximation methods have long been used to solve stochastic optimization problems. Their application to infinite-dimensional problems is less understood, particularly for nonconvex objectives. This paper presents convergence results for the stochastic proximal gradient method applied to Hilbert spaces, motivated by optimization problems with partial differential equation (PDE) constraints with random inputs and coefficients. We study stochastic algorithms for nonconvex and nonsmooth problems, where the nonsmooth part is convex and the nonconvex part is the expectation, which is assumed to have a Lipschitz continuous gradient. The optimization variable is an element of a Hilbert space. We show almost sure convergence of strong limit points of the random sequence generated by the algorithm to stationary points. We demonstrate the stochastic proximal gradient algorithm on a tracking-type functional with a L^1 -penalty term constrained by a semilinear PDE and box constraints, where input terms and coefficients are subject to uncertainty. We verify conditions for ensuring convergence of the algorithm and show a simulation.

Keywords Stochastic programming · Nonsmooth and nonconvex optimization · Differential inclusions · Mathematical programming methods · Partial differential equations with randomness · Optimal control problems involving partial differential equations

The work of the both authors was funded by the Austrian Science Fund (FWF) within Project No. W1260-N35. The second author acknowledges the support by the Istituto Nazionale di Alta Matematica (INdAM), through the GNAMPA Research Project 2019: “Control of PDEs for physical models and life sciences” coordinated by Prof. Fragnelli.

✉ Caroline Geiersbach
caroline.geiersbach@wias-berlin.de

Teresa Scarinci
teresa.scarinci@univaq.it

¹ Weierstrass Institute, Mohrenstrasse 39, 10117 Berlin, Germany

² Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, Via Vetoio - Loc. Coppito, 67010 L'Aquila, Italy

1 Introduction

In this paper, we focus on stochastic approximation methods for solving a stochastic optimization problem on a Hilbert space H of the form

$$\min_{u \in H} \{f(u) = j(u) + h(u)\}, \quad (\text{P})$$

where the expectation $j(u) = \mathbb{E}[J(u, \xi)]$ is generally nonconvex with a Lipschitz continuous gradient and h is a proper, lower semicontinuous, and convex function that is generally nonsmooth.

Our work is motivated by applications to PDE-constrained optimization under uncertainty, where a nonlinear PDE constraint can lead to an objective function that is nonconvex with respect to the Hilbert-valued variable. To handle the (potentially infinite-dimensional) expectation, algorithmic approaches for solving such problems involve either some discretization of the stochastic space or an ensemble-based approach with sampling or carefully chosen quadrature points. Stochastic discretization includes polynomial chaos and the stochastic Galerkin method; cf. [24, 30, 34, 47]. For ensemble-based methods, the simplest method is sample average approximation (SAA), where the original problem is replaced by a proxy problem with a fixed set of samples, which can then be solved using a deterministic solver. A number of standard improvements to Monte Carlo sampling have been applied to optimal control problems in, e.g., [1, 54]. Another ensemble-based approach is the stochastic collocation method, which has been used in optimal control problems in e.g. [47, 51]. Sparse-tensor discretization has been used for optimal control problems in, for instance, [28, 29].

The approach we use is an ensemble-based approach called stochastic approximation, which is fundamentally different in the sense that sampling takes place dynamically as part of the optimization procedure, leading to an algorithm with low complexity and computational effort when compared to other approaches. Stochastic approximation originated in a groundbreaking paper by [45], where an iterative method to find the root of an unknown function using noisy estimates was proposed. The authors of [25] used this idea to solve a regression problem using finite differences subject to noise. Algorithms of this kind, with bias in addition to stochastic noise, are sometimes called stochastic quasi-gradient methods; see, e.g., [17, 53]. Basic versions of these algorithms rely on positive step sizes t_n of the form $\sum_{n=1}^{\infty} t_n = \infty$ and $\sum_{n=1}^{\infty} t_n^2 < \infty$. The (almost sure) asymptotic convergence of stochastic approximation algorithms for convex problems is classical in finite dimensions; we refer to the texts by [16, 33].

There have been a number of contributions with proofs of convergence of the stochastic gradient method for unconstrained nonconvex problems; see [6, 7, 49, 56]. Fewer results exist for constrained and/or nonsmooth nonconvex problems. A randomized stochastic algorithm was proposed by Ghadimi et al. [21]; this scheme involves running a stochastic approximation process and randomly choosing an iterate from the generated sequence. There have been some contributions involving constant step sizes with increasing sampling; see [35, 44]. Convergence

of projection-type methods for nonconvex problems was shown in [32] and for prox-type methods by Davis et al. [13].

As far as stochastic approximation on function spaces is concerned, many contributions were motivated by applications with nonparametric statistics. Perhaps the oldest example is from [55]. Goldstein [22] studied an infinite-dimensional version of the Kiefer–Wolfowitz procedure. A significant contribution for unconstrained problems was by Yin and Zhu [58]. Projection-type methods were studied by [3, 10, 12, 40].

In this paper, we prove convergence results for nonconvex and nonsmooth problems in Hilbert spaces. We present convergence analysis that is based on the recent contributions in [13, 35]. Applications of the stochastic gradient method to PDE-constrained optimization have already been explored by [19, 37]. In these works, however, convexity of the objective function is assumed, leaving the question of convergence in the more general case entirely open. We close that gap by making the following contributions:

- For an objective function that is the sum of a smooth, generally nonconvex expectation and a convex, nonsmooth term, we prove that strong accumulation points of iterates generated by the method are stationary points.
- We show that convergence holds even in the presence of systematic additive bias, which is relevant for the application in mind.
- We demonstrate the method on an application to PDE-constrained optimization under uncertainty and verify conditions for convergence.

The paper is organized as follows. In Sect. 2, notation and background is given. Convergence of two related algorithms is proven in Sect. 3. In Sect. 4, we introduce a problem in PDE-constrained optimization under uncertainty, where coefficients in the semilinear PDE constraint are subject to uncertainty. The problem is shown to satisfy conditions for convergence, and numerical experiments demonstrate the method. We finish the paper with closing remarks in Sect. 5.

2 Notation and background

We recall some notation and background from convex analysis and stochastic processes; see [4, 11, 38, 43].

Let H be a Hilbert space with the scalar product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. The symbols \rightarrow and \rightharpoonup denote strong and weak convergence, respectively. The set of proper, convex, and lower semicontinuous functions $h : H \rightarrow (-\infty, \infty]$ is denoted by $\Gamma_0(H)$. Given a function $h \in \Gamma_0(H)$ and $t > 0$, the proximity operator $\text{prox}_{th} : H \rightarrow H$ is given by

$$\text{prox}_{th}(u) := \arg \min_{v \in H} \left(h(v) + \frac{1}{2t} \|v - u\|^2 \right).$$

We recall that for a proper function $h : H \rightarrow (-\infty, \infty]$, the subdifferential (in the sense of convex analysis) is the set-valued operator

$$\partial h : H \rightrightarrows H : u \mapsto \{v \in H : \langle y - u, v \rangle + h(u) \leq h(y) \quad \forall y \in H\}.$$

For any $h \in \Gamma_0(H)$, the subdifferential ∂h is maximally monotone. The domain of h is denoted by $\text{dom}(h)$. The indicator function of a set C is denoted by δ_C , where $\delta_C(u) = 0$ if $u \in C$ and $\delta_C(u) = \infty$ otherwise. The sum of two sets A and B with $\lambda \in \mathbb{R}$ is given by $A + \lambda B := \{a + \lambda b : a \in A, b \in B\}$. The distance of a point u to a nonempty, closed set A is denoted by $d(u, A) := \inf_{a \in A} \|u - a\|$ and the diameter of A is denoted by the symbol $\text{diam}(A) := \sup_{u, v \in A} \|u - v\|$. For a nonempty and convex set C , the normal cone $N_C(u)$ at $u \in C$ is defined by

$$N_C(u) := \{z \in H : \langle z, w - u \rangle \leq 0, \quad \forall w \in C\}.$$

We set $N_C(u) := \emptyset$ if $u \notin C$. We recall that $\partial \delta_C(u) = N_C(u)$ for all $u \in C$. If $h_1, h_2 \in \Gamma_0(H)$ and $\text{dom}(h_2) = H$, then $\partial[h_1(u) + h_2(u)] = \partial h_1(u) + \partial h_2(u)$. If h is proper and $\text{dom}(h)$, then $\partial h(u)$ is closed and convex. We recall that the graph of ∂h for a function $h \in \Gamma_0(H)$, given by the set $\text{gra}(\partial h) = \{(u, \partial h(u)) : u \in H\}$, is sequentially closed in the strong-to-weak topology, meaning that for $u_n \rightarrow u$, $\zeta_n \in \partial h(u_n)$, and $\zeta_n \rightarrow \zeta$, it follows that $\zeta \in \partial h(u)$. The normal cone $N_C(u)$ is strong-to-weak sequentially closed if C is convex.

Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ will denote a probability space, where Ω represents the sample space, $\mathcal{F} \subset 2^\Omega$ is the σ -algebra of events on the power set of Ω , denoted by 2^Ω , and $\mathbb{P} : \Omega \rightarrow [0, 1]$ is a probability measure. Given a random vector $\xi : \Omega \rightarrow \mathcal{E} \subset \mathbb{R}^m$, we write $\xi \in \mathcal{E}$ to denote a realization of the random vector. The operator $\mathbb{E}[\cdot]$ denotes the expectation with respect to this distribution; for a parametrized functional $J : H \times \mathcal{E} \rightarrow \mathbb{R}$, this is defined as the integral over all elements in Ω , i.e.,

$$\mathbb{E}[J(u, \xi)] = \int_{\Omega} J(u, \xi(\omega)) \, d\mathbb{P}(\omega).$$

A filtration is a sequence $\{\mathcal{F}_n\}$ of sub- σ -algebras of \mathcal{F} such that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}$. We define a discrete H -valued stochastic process as a collection of H -valued random variables indexed by n , in other words, the set $\{\beta_n : \Omega \rightarrow H \mid n \in \mathbb{N}\}$. The stochastic process is said to be adapted to a filtration $\{\mathcal{F}_n\}$ if and only if β_n is \mathcal{F}_n -measurable for all n . The natural filtration is the filtration generated by the sequence $\{\beta_n\}$ and is given by $\mathcal{F}_n = \sigma(\{\beta_1, \dots, \beta_n\})$.¹ If for an event $F \in \mathcal{F}$ it holds that $\mathbb{P}(F) = 1$, or equivalently, $\mathbb{P}(\Omega \setminus F) = 0$, we say F occurs almost surely (a.s.). Sometimes we also say that such an event occurs with probability one. A sequence of random variables $\{\beta_n\}$ is said to converge almost surely to a random variable β if and only if

¹ The σ -algebra generated by a random variable $\beta : \Omega \rightarrow \mathbb{R}$ is given by $\sigma(\beta) = \{\beta^{-1}(B) : B \in \mathcal{B}\}$, where \mathcal{B} is the Borel σ -algebra on \mathbb{R} . Analogously, the σ -algebra generated by the set of random variables $\{\beta_1, \dots, \beta_n\}$ is the smallest σ -algebra such that β_i is measurable for all $i = 1, \dots, n$.

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} \beta_n(\omega) = \beta(\omega)\right\}\right) = 1.$$

For an integrable random variable $\beta : \Omega \rightarrow \mathbb{R}$, the conditional expectation is denoted by $\mathbb{E}[\beta | \mathcal{F}_n]$, which is itself a random variable that is \mathcal{F}_n -measurable and which satisfies $\int_A \mathbb{E}[\beta | \mathcal{F}_n](\omega) d\mathbb{P}(\omega) = \int_A \beta(\omega) d\mathbb{P}(\omega)$ for all $A \in \mathcal{F}_n$. Almost sure convergence of H -valued stochastic processes and conditional expectation are defined analogously.

Given a random operator $F : X \times \Omega \rightarrow Y$, where X and Y are Banach spaces, we will sometimes use the notation $F_\omega := F(\cdot, \omega) : X \rightarrow Y$ for a fixed (but arbitrary) $\omega \in \Omega$. For a Banach space $(X, \|\cdot\|_X)$, the Bochner space $L^p(\Omega, X)$ is the set of all (equivalence classes of) strongly measurable functions $u : \Omega \rightarrow X$ having finite norm, where the norm is defined by

$$\|u\|_{L^p(\Omega, X)} := \begin{cases} (\int_\Omega \|u(\omega)\|_X^p d\mathbb{P}(\omega))^{1/p}, & p < \infty \\ \text{ess sup}_{\omega \in \Omega} \|u(\omega)\|_X, & p = \infty \end{cases}.$$

A sequence $\{\beta_n\}$ in $L^1(\Omega, X)$ is called a martingale if a filtration $\{\mathcal{F}_n\}$ exists such that β_n is \mathcal{F}_n -measurable and $\mathbb{E}[\beta_{n+1} | \mathcal{F}_n] = \beta_n$ is satisfied for all n .

For an open subset U of a Banach space X and a function $J_\omega : U \rightarrow \mathbb{R}$, we denote the Gâteaux derivative at $u \in U$ in the direction $v \in X$ by $dJ_\omega(u; v)$. The Fréchet derivative at u is denoted by $J'_\omega : U \rightarrow \mathcal{L}(X, \mathbb{R})$, where $\mathcal{L}(X, \mathbb{R})$ is the set of bounded and linear operators mapping X to \mathbb{R} . We recall this is none other than the dual space X^* and we denote the dual pairing by $\langle \cdot, \cdot \rangle_{X^*, X}$. For an open subset U of a Hilbert space H and a Fréchet differentiable function $j : U \rightarrow \mathbb{R}$, the gradient $\nabla j : U \rightarrow H$ is the Riesz representation of $j' : U \rightarrow H^*$, i.e., it satisfies $\langle \nabla j(u), v \rangle = \langle j'(u), v \rangle_{H^*, H}$ for all $u \in U$ and $v \in H$. In Hilbert spaces, the Riesz representation relates elements of the dual space to the Hilbert space itself, allowing us to drop the dual pairing notation and use simply $\langle \cdot, \cdot \rangle$.

The notation $C_L^{1,1}(U)$ is used to denote the set of continuously differentiable functions on $U \subset H$ with an L -Lipschitz gradient, meaning $\|\nabla j(u) - \nabla j(v)\| \leq L\|u - v\|$ is satisfied for all $u, v \in U$. The following lemma gives a classical Taylor estimate for such functions.

Lemma 2.1 *Suppose $j \in C_L^{1,1}(U)$, $U \subset H$ open and convex. Then for all $u, v \in U$,*

$$j(v) + \langle \nabla j(v), u - v \rangle - \frac{L}{2} \|u - v\|^2 \leq j(u) \leq j(v) + \langle \nabla j(v), u - v \rangle + \frac{L}{2} \|u - v\|^2.$$

3 Asymptotic convergence results

In this section, we show asymptotic convergence results for two variants of the stochastic proximal gradient method in Hilbert spaces for solving Problem (P). Let $G : H \times \Xi \rightarrow H$ be a parametrized operator (the *stochastic gradient*) approximating

(in a sense to be specified later) the gradient $\nabla j : H \rightarrow H$ and let t_n be a positive step size. Both algorithms in this section will share the basic iterative form

$$u_{n+1} := \text{prox}_{t_n h}(u_n - t_n G(u_n, \xi_n)),$$

where h is the nonsmooth term from Problem (P). The following assumptions will be in force in all sections.

Assumption 3.1 Let $\{\mathcal{F}_n\}$ be a filtration and let $\{u_n\}$ and $\{G(u_n, \xi_n)\}$ be sequences of iterates and stochastic gradients. We assume

- (i) The sequence $\{u_n\}$ is a.s. contained in a bounded set $V \subset H$ and u_n is adapted to \mathcal{F}_n for all n .
- (ii) On an open and convex set U such that $V \subset U \subset H$, the expectation $j \in C_L^{1,1}(U)$ is bounded below.
- (iii) For all n , the H -valued random variable $r_n := \mathbb{E}[G(u_n, \xi_n)|\mathcal{F}_n] - \nabla j(u_n)$ is adapted to \mathcal{F}_n and for $K_n := \text{ess sup}_{\omega \in \Omega} \|r_n(\omega)\|$, $\sum_{n=1}^\infty t_n K_n < \infty$ and $\sup_n K_n < \infty$ are satisfied.
- (iv) For all n , $\mathfrak{w}_n := G(u_n, \xi_n) - \mathbb{E}[G(u_n, \xi_n)|\mathcal{F}_n]$ is an H -valued random variable.

Remark 3.2 The assumption that the sequence $\{u_n\}$ stays bounded with probability one is by no means automatically fulfilled, but can be verified or enforced in different ways. We refer to [6, Section 5.2] and [13, Section 6.1] for conditions on the function, constraint set, and/or regularizers that ensure boundedness of iterates. The conditions in Assumption 3.1 allow for additive bias r_n in the stochastic gradient in addition to zero-mean error \mathfrak{w}_n . The requirement that u_n and r_n are adapted to \mathcal{F}_n is automatically fulfilled if $\{\mathcal{F}_n\}$ is chosen to be the natural filtration generated by $\{\xi_1, \dots, \xi_n\}$. Together, Assumption 3.1(iii) and Assumption 3.1(iv) imply

$$G(u_n, \xi_n) = \nabla j(u_n) + r_n + \mathfrak{w}_n$$

and $\mathbb{E}[\mathfrak{w}_n|\mathcal{F}_n] = 0$. Notice that a single realization $\xi_n \in \Xi$ can be replaced by m_n independently drawn realizations $\xi_n^1, \dots, \xi_n^{m_n} \in \Xi$ since

$$\mathbb{E}[G(u_n, \xi_n)|\mathcal{F}_n] = \frac{1}{m_n} \mathbb{E} \left[\sum_{i=1}^{m_n} G(u_n, \xi_n^i) | \mathcal{F}_n \right].$$

This set of m_n samples is sometimes called a ‘‘batch’’; batches clearly reduce the variance of the stochastic gradient.

The result in Sect. 3.1 shows asymptotic convergence of the proximal gradient method with constant step sizes and increasing sampling. In Sect. 3.2, we switch to the versatile ordinary differential equation (ODE) method to prove convergence of the stochastic proximal gradient method with decreasing step sizes. We emphasize that the convergence results generalize existing convergence theory from the finite-dimensional case. Our analysis includes convergence in possibly infinite-dimensional Hilbert spaces. Additionally, we allow for stochastic gradients subject

to additive bias, which is not covered by existing results. This theory can be used to develop mesh refinement strategies in applications with PDEs [20].

3.1 Variance-reduced stochastic proximal gradient method

In this section, we show under what conditions the variance-reduced stochastic proximal gradient method converges to stationary points for Problem (P). With $\xi_n = (\xi_n^1, \dots, \xi_n^{m_n})$, the stochastic gradient is given by the average

$$G(u_n, \xi_n) = \frac{\sum_{i=1}^{m_n} G(u_n, \xi_n^i)}{m_n}$$

over an *increasing* number of samples m_n . The algorithm is presented below, which uses constant step sizes $t_n \equiv t$ depending on the Lipschitz constant L from Assumption 3.1(ii).

Algorithm 1 Variance-Reduced Stochastic Proximal Gradient Method

Initialization: $u_1 \in H, 0 < t < \frac{1}{2L}$
for $n = 1, 2, \dots$ **do**
 Generate independent $\xi_n^1, \dots, \xi_n^{m_n} \in \Xi$, independent of $\xi_1^1, \dots, \xi_{n-1}^{m_{n-1}}$
 $u_{n+1} := \text{prox}_{th} \left(u_n - t \frac{\sum_{i=1}^{m_n} G(u_n, \xi_n^i)}{m_n} \right)$
end for

Remark 3.3 If $h(u) = \delta_C(u)$ and π_C denotes the projection onto C , then the algorithm reduces to $u_{n+1} := \pi_C \left(u_n - t \frac{\sum_{i=1}^{m_n} G(u_n, \xi_n^i)}{m_n} \right)$, i.e., the variance-reduced projected stochastic gradient method.

In addition to Assumption 3.1, the following assumptions will be in force in this section.

Assumption 3.4 Let $\{u_n\}$ and $\{G(u_n, \xi_n)\}$ be generated by Algorithm 1. We assume

- (i) The function h satisfies $h \in \Gamma_0(H)$.
- (ii) For all n ,

$$w_n := \frac{\sum_{i=1}^{m_n} G(u_n, \xi_n^i)}{m_n} - \nabla j(u_n)$$

is an H -valued random variable and there exists an $M \geq 0$ such that $\mathbb{E}[\|w_n\|^2 | \mathcal{F}_n] \leq \frac{M}{m_n}$ and $\sum_{n=1}^{\infty} \frac{1}{m_n} < \infty$.

Remark 3.5 We use assumptions similar to those found in [35], but we do not require the effective domain of h to be bounded; we instead use boundedness of the iterates

by Assumption 3.1(i). Notice that $w_n = r_n + \mathfrak{w}_n$ from Assumption 3.1(iv), hence Assumption 3.4(ii) also provides a condition on the rate at which r_n and \mathfrak{w}_n must decay.

For the convergence result, we need the following lemma [46].

Lemma 3.6 (Robbins–Siegmund) *Assume that $\{\mathcal{F}_n\}$ is a filtration and v_n, a_n, b_n, c_n are nonnegative random variables adapted to \mathcal{F}_n . If*

$$\mathbb{E}[v_{n+1}|\mathcal{F}_n] \leq v_n(1 + a_n) + b_n - c_n \quad \text{a.s.}$$

and $\sum_{n=1}^\infty a_n < \infty, \sum_{n=1}^\infty b_n < \infty$ a.s., then with probability one, $\{v_n\}$ is convergent and $\sum_{n=1}^\infty c_n < \infty$.

To show convergence, we first present a technical lemma.

Lemma 3.7 *Let $u \in U$ and $t > 0$. Suppose $v := \text{prox}_{th}(u - tg) \in U$ for a given $g \in H$. Then for any $z \in U$,*

$$\begin{aligned} f(v) \leq & f(z) + \langle v - z, \nabla j(u) - g \rangle + \left(\frac{L}{2} - \frac{1}{2t}\right) \|v - u\|^2 \\ & + \left(\frac{L}{2} + \frac{1}{2t}\right) \|z - u\|^2 - \frac{1}{2t} \|v - z\|^2. \end{aligned} \tag{3.1}$$

Proof We first claim that for all $y, z \in H, t > 0$ and $p = \text{prox}_{th}(y)$,

$$h(p) + \frac{1}{2t} \|p - y\|^2 \leq h(z) + \frac{1}{2t} \|z - y\|^2 - \frac{1}{2t} \|p - z\|^2. \tag{3.2}$$

This follows by definition of the prox operator. Indeed, for $t > 0, p = \text{prox}_{th}(y)$ if and only if for all $z \in H$,

$$h(z) \geq h(p) + \frac{1}{t} \langle y - p, z - p \rangle. \tag{3.3}$$

It is straightforward to verify the following equality (the law of cosines)

$$\|z - y\|^2 = \|z - p\|^2 + \|p - y\|^2 - 2\langle y - p, z - p \rangle. \tag{3.4}$$

Multiplying (3.4) by $\frac{1}{2t}$ and adding it to (3.3), we get (3.2). Now, since $j \in C_L^{1,1}(U)$, it follows by Lemma 2.1 for $u, v, z \in U$ that

$$j(v) \leq j(u) + \langle \nabla j(u), v - u \rangle + \frac{L}{2} \|v - u\|^2, \tag{3.5}$$

$$j(u) \leq j(z) + \langle \nabla j(u), u - z \rangle + \frac{L}{2} \|z - u\|^2. \tag{3.6}$$

Combining (3.5) and (3.6), we get

$$j(v) \leq j(z) + \langle \nabla j(u), v - z \rangle + \frac{L}{2} \|v - u\|^2 + \frac{L}{2} \|z - u\|^2. \tag{3.7}$$

Now, by (3.2) applied to $v = \text{prox}_{th}(u - tg)$,

$$h(v) + \frac{1}{2t} \|v - (u - tg)\|^2 \leq h(z) + \frac{1}{2t} \|z - (u - tg)\|^2 - \frac{1}{2t} \|v - z\|^2$$

if and only if

$$\begin{aligned} h(v) + \frac{1}{2t} \|v - u\|^2 + \langle v - u, g \rangle \\ \leq h(z) + \frac{1}{2t} \|z - u\|^2 + \langle z - u, g \rangle - \frac{1}{2t} \|v - z\|^2. \end{aligned} \tag{3.8}$$

Finally, adding (3.7) and (3.8), and using that $f = j + h$, we get (3.1). □

In the following, we define

$$\bar{u}_{n+1} := \text{prox}_{th}(u_n - t\nabla j(u_n)) \tag{3.9}$$

as the iterate at $n + 1$ if the true gradient were used.

Lemma 3.8 *For all n ,*

$$\mathbb{E}[f(u_{n+1}) | \mathcal{F}_n] \leq f(u_n) - \left(\frac{1}{2t} - L\right) \|\bar{u}_{n+1} - u_n\|^2 + \frac{t}{2} \mathbb{E}[\|w_n\|^2 | \mathcal{F}_n] \quad \text{a.s.} \tag{3.10}$$

Proof Using Lemma 3.7 with $v = \bar{u}_{n+1}$, $u = z = u_n$, and $g = \nabla j(u_n)$, we have

$$f(\bar{u}_{n+1}) \leq f(u_n) + \left(\frac{L}{2} - \frac{1}{t}\right) \|\bar{u}_{n+1} - u_n\|^2. \tag{3.11}$$

Again using Lemma 3.7, with $v = u_{n+1}$, $z = \bar{u}_{n+1}$, $u = u_n$, and $g = \nabla j(u_n) + w_n$, we get

$$\begin{aligned} f(u_{n+1}) \leq f(\bar{u}_{n+1}) - \langle u_{n+1} - \bar{u}_{n+1}, w_n \rangle + \left(\frac{L}{2} - \frac{1}{2t}\right) \|u_{n+1} - u_n\|^2 \\ + \left(\frac{L}{2} + \frac{1}{2t}\right) \|\bar{u}_{n+1} - u_n\|^2 - \frac{1}{2t} \|u_{n+1} - \bar{u}_{n+1}\|^2. \end{aligned} \tag{3.12}$$

By Young's inequality, $\langle u_{n+1} - \bar{u}_{n+1}, w_n \rangle \leq \frac{1}{2t} \|u_{n+1} - \bar{u}_{n+1}\|^2 + \frac{t}{2} \|w_n\|^2$, so combining (3.11) and (3.12), we obtain since $0 < t < \frac{1}{2L}$ that

$$\begin{aligned} f(u_{n+1}) \leq f(u_n) + \left(L - \frac{1}{2t}\right) \|\bar{u}_{n+1} - u_n\|^2 + \left(\frac{L}{2} - \frac{1}{2t}\right) \|u_{n+1} - u_n\|^2 \\ + \frac{t}{2} \|w_n\|^2 \\ \leq f(u_n) + \left(L - \frac{1}{2t}\right) \|\bar{u}_{n+1} - u_n\|^2 + \frac{t}{2} \|w_n\|^2. \end{aligned} \tag{3.13}$$

Taking conditional expectation on both sides of (3.13), and noting that \bar{u}_{n+1} is \mathcal{F}_n -measurable by \mathcal{F}_n -measurability of u_n , we get (3.10). \square

Remark 3.9 Any bounded sequence $\{u_n\}$ in H contains a weakly convergent subsequence $\{u_{n_k}\}$ such that $u_{n_k} \rightharpoonup u$ for a $u \in H$. Generally this convergence is not strong, so we cannot conclude from $\|\bar{u}_{n+1} - u_n\|^2 \rightarrow 0$ that there exists a \tilde{u} such that, for a subsequence $\{u_{n_k}\}$, $\lim_{k \rightarrow \infty} \bar{u}_{n_k+1} = \lim_{k \rightarrow \infty} u_{n_k} = \tilde{u}$. Therefore, to obtain convergence to stationary points, we will assume that $\{u_n\}$ has a strongly convergent subsequence.

We are ready to state the convergence result for sequences generated by Algorithm 1.

Theorem 3.10 *Let Assumptions 3.1 and 3.4 hold. Then*

1. *The sequence $\{f(u_n)\}$ converges a.s.*
2. *The sequence $\{\|\bar{u}_{n+1} - u_n\|\}$ converges to zero a.s.*
3. *Every strong accumulation point of $\{u_n\}$ is a stationary point with probability one.*

Proof The sequence $\{u_n\}$ is contained in a bounded set V by Assumption 3.1(i). By Assumption 3.4(i), $h \in \Gamma_0(H)$ must therefore be bounded below on V [4, Corollary 9.20]; j is bounded below by Assumption 3.1(ii). W.l.o.g. we can thus assume $f \geq 0$. Since $\frac{1}{2r} > L$ and $\sum_{n=1}^\infty \mathbb{E}[\|w_n\|^2 | \mathcal{F}_n] < \infty$ by Assumption 3.4(ii), we can apply Lemma 3.6 to (3.10) to conclude that $f(u_n)$ converges almost surely. The second statement follows immediately, since by Lemma 3.6,

$$\sum_{n=1}^\infty \|\bar{u}_{n+1} - u_n\|^2 < \infty \quad \text{a.s.}, \tag{3.14}$$

which implies that for almost every sample path, $\lim_{n \rightarrow \infty} \|\bar{u}_{n+1} - u_n\|^2 = 0$.

For the third statement, we have that there exists a subsequence $\{u_{n_k}\}$ such that $u_{n_k} \rightarrow u$. We argue that then $\bar{u}_{n_k+1} \rightarrow u$. Since $\{\bar{u}_{n_k+1}\}$ is bounded, there exists a weak limit point \tilde{u} (potentially on a subsequence with the same labeling). Then, using weak lower semicontinuity of the norm as well as the rule $\langle a_n, b_n \rangle \rightarrow \langle a, b \rangle$ for $a_n \rightarrow a$ and $b_n \rightarrow b$,

$$\begin{aligned} 0 &= \lim_{k \rightarrow \infty} \|\bar{u}_{n_k+1} - u_{n_k}\|^2 = \lim_{k \rightarrow \infty} \|\bar{u}_{n_k+1}\|^2 - 2\langle \bar{u}_{n_k+1}, u_{n_k} \rangle + \|u_{n_k}\|^2 \\ &= \liminf_{k \rightarrow \infty} \|\bar{u}_{n_k+1}\|^2 - 2\langle \tilde{u}, u \rangle + \|u\|^2 \\ &\geq \|\tilde{u}\|^2 - 2\langle \tilde{u}, u \rangle + \|u\|^2 = \|\tilde{u} - u\|^2 \geq 0, \end{aligned}$$

implying $u = \tilde{u}$. It follows $\bar{u}_{n_k+1} \rightarrow u$ by assuming $\lim_{k \rightarrow \infty} \|\bar{u}_{n_k+1}\|^2 \neq \|u\|^2$ and arriving at a contradiction. Now, by definition of the prox operator,

$$\begin{aligned} \bar{u}_{n_k+1} &= \text{prox}_{tH}(u_{n_k} - t\nabla j(u_{n_k})) \\ &= \arg \min_{v \in H} \left\{ h(v) + \frac{1}{2t} \|v - u_{n_k} + t\nabla j(u_{n_k})\|^2 \right\} \\ &= \arg \min_{v \in H} \left\{ h(v) + \langle \nabla j(u_{n_k}), v \rangle + \frac{1}{2t} \|v\|^2 - \frac{1}{t} \langle v, u_{n_k} \rangle =: H(v) \right\}. \end{aligned}$$

Clearly, $\partial H(v) = \partial h(v) + \nabla j(u_{n_k}) + \frac{1}{t}(v - u_{n_k})$. By optimality of \bar{u}_{n_k+1} (see Fermat’s rule, [4, Theorem 16.2]), $0 \in \partial H(\bar{u}_{n_k+1})$, or equivalently,

$$-\frac{1}{t}(\bar{u}_{n_k+1} - u_{n_k}) \in \nabla j(u_{n_k}) + \partial h(\bar{u}_{n_k+1}).$$

Taking the limit as $k \rightarrow \infty$, and using continuity of ∇j , we conclude by strong-to-weak sequential closedness of $\text{gra}(\partial h)$ that

$$0 \in \nabla j(u) + \partial h(u), \tag{3.15}$$

so therefore u is a stationary point. □

3.2 Stochastic proximal gradient method: decreasing step sizes

An obvious drawback of Algorithm 1 is the fact that step sizes are restricted to small steps bounded by a factor depending on the Lipschitz constant, which in applications might be difficult to determine. Additionally, the algorithm requires increasing batch sizes to dampen noise, which is unattractive from a complexity standpoint. In this section, we obtain convergence with a nonsmooth and convex term h using the step size rule

$$t_n \geq 0, \quad \sum_{n=1}^{\infty} t_n = \infty, \quad \sum_{n=1}^{\infty} t_n^2 < \infty. \tag{3.16}$$

This step size rule dampens noise enough so that increased sampling is not necessary.

We observe Problem (P) with

$$h(u) := \eta(u) + \delta_C(u).$$

For asymptotic arguments, it will be convenient to treat the term δ_C separately. To that end, we define

$$\varphi(u) := j(u) + \eta(u)$$

and note that $f(u) = \varphi(u) + \delta_C(u)$. The stochastic gradient $G(u, \xi) : H \times \Xi \rightarrow H$ can be comprised of one or more samples as in the unconstrained case; see Remark 3.2. The algorithm is now stated below.

Algorithm 2 Stochastic Proximal Gradient Method

Initialization: $u_1 \in C$
for $n = 1, 2, \dots$ **do**
 Generate $\xi_n \in \Xi$, independent of ξ_1, \dots, ξ_{n-1}
 Choose t_n satisfying (3.16)
 $u_{n+1} := \text{prox}_{t_n h}(u_n - t_n G(u_n, \xi_n))$
end for

To prove convergence of Algorithm 2, we will use the ODE method, which dates back to [33, 36]. While we use many ideas from [13], we emphasize that we generalize results to (possibly infinite-dimensional) Hilbert spaces and moreover, we handle the case when j is the expectation.

We define the set-valued map $S : C \rightrightarrows H$ by

$$S(u) := -\nabla j(u) - \partial\eta(u) - N_C(u).$$

Additionally, we define the sequence of (single-valued) maps $S_n : C \rightarrow H$ for all n by

$$S_n(u) := -\nabla j(u) - \frac{1}{t_n} \mathbb{E}[u - t_n G(u, \xi) - \text{prox}_{t_n h}(u - t_n G(u, \xi))].$$

In addition to Assumption 3.1, the following assumptions will apply in this section.

Assumption 3.11 Let $\{u_n\}$ and $\{G(u_n, \xi_n)\}$ be generated by Algorithm 2. We assume

- (i) The set C is nonempty, bounded, convex, and closed.
- (ii) The function $\eta \in \Gamma_0(H)$ with $\text{dom}(\eta) = H$ is locally Lipschitz and bounded below on C , and there exists a function $L_\eta : H \rightarrow \mathbb{R}$, which is bounded on bounded sets, satisfying

$$L_\eta(u) \geq \sup_{z: \eta(z) \leq \eta(u)} \frac{\eta(u) - \eta(z)}{\|u - z\|}. \tag{3.17}$$

- (iii) There exists a function $M : H \rightarrow [0, \infty)$, which is bounded on bounded sets, such that $\mathbb{E}[\|G(u, \xi)\|^2] \leq M(u)$.
- (iv) For any strongly convergent sequence $\{u_n\}$, $\mathbb{E}[\sup_n \|G(u_n, \xi)\|] < \infty$ holds.
- (v) The set of critical values $\{f(u) : 0 \in \partial f(u)\}$ does not contain any segment of nonzero length.

Remark 3.12 To handle the infinite-dimensional case, we use assumptions that are generally more restrictive than in [13]; we restrict ourselves to the case where C and η are convex and we assume higher regularity of j in Assumption 3.1(ii) to handle the case $j(u) = \mathbb{E}[J(u, \xi)]$. However, we allow for bias r_n , which is not covered

in [13]. We note that C does not need to be bounded if η is Lipschitz continuous over C . Assumption 3.11(ii) is satisfied if $\text{dom}(\partial\eta) = H$ and $\partial\eta$ maps bounded sets to bounded sets; see also [4, Proposition 16.17] for equivalent conditions. The last assumption is technical but standard; see [48, Assumption H4].

The main result is the following, which we will prove in several parts. Throughout, we use the notation $g_n := G(u_n, \xi_n)$.

Theorem 3.13 *Let Assumptions 3.1 and 3.11 hold. Then*

1. *The sequence $\{f(u_n)\}$ converges a.s.*
2. *Every strong accumulation point u of the sequence $\{u_n\}$ is a stationary point with probability one, namely, $0 \in \partial f(u)$ a.s.*

Lemma 3.14 *The sequence $\{u_n\}$ satisfies the recursion*

$$u_{n+1} = u_n + t_n(y_n - r_n + w_n), \tag{3.18}$$

where $y_n = S_n(u_n)$ and $w_n = -\frac{1}{t_n}\mathbb{E}[\text{prox}_{t_n h}(u_n - t_n g_n)|\mathcal{F}_n] + \frac{1}{t_n}\text{prox}_{t_n h}(u_n - t_n g_n)$.

Proof Note that u_n and r_n are \mathcal{F}_n -measurable, so $\mathbb{E}[g_n|\mathcal{F}_n] = \nabla j(u_n) + r_n$. Then

$$\begin{aligned} u_{n+1} - u_n &= \text{prox}_{t_n h}(u_n - t_n g_n) - u_n \\ &= -t_n \mathbb{E}[g_n|\mathcal{F}_n] - \mathbb{E}[u_n - t_n g_n - \text{prox}_{t_n h}(u_n - t_n g_n)|\mathcal{F}_n] \\ &\quad - \mathbb{E}[\text{prox}_{t_n h}(u_n - t_n g_n)|\mathcal{F}_n] + \text{prox}_{t_n h}(u_n - t_n g_n) \\ &= t_n S_n(u_n) - t_n r_n - \mathbb{E}[\text{prox}_{t_n h}(u_n - t_n g_n)|\mathcal{F}_n] + \text{prox}_{t_n h}(u_n - t_n g_n), \end{aligned}$$

where we used that ξ_n is independent from ξ_1, \dots, ξ_{n-1} , so

$$\begin{aligned} &\mathbb{E}[u_n - t_n g_n - \text{prox}_{t_n h}(u_n - t_n g_n)|\mathcal{F}_n] \\ &= \mathbb{E}[u_n - t_n G(u_n, \xi) - \text{prox}_{t_n h}(u_n - t_n G(u_n, \xi))]. \end{aligned} \tag{3.19}$$

By definition of y_n and w_n , we arrive at the conclusion. □

Lemma 3.15 *For any $u \in C$, $g \in H$ and $t > 0$, we have for $\bar{u} = \text{prox}_{th}(u - tg)$ that*

$$\frac{1}{t} \|\bar{u} - u\| \leq 2L_\eta(u) + 2\|g\|.$$

Proof By definition of the proximity operator,

$$\eta(\bar{u}) + \delta_C(\bar{u}) + \frac{1}{2t} \|\bar{u} - (u - tg)\|^2 \leq \eta(u) + \delta_C(u) + \frac{1}{2t} \|u - (u - tg)\|^2,$$

or equivalently (note $\bar{u}, u \in C$),

$$\eta(\bar{u}) + \frac{1}{2t} \|\bar{u} - u\|^2 + \langle \bar{u} - u, g \rangle \leq \eta(u).$$

By (3.17), in the case $\eta(u) \geq \eta(\bar{u})$, we obtain

$$\frac{1}{t} \|\bar{u} - u\|^2 \leq 2(\eta(u) - \eta(\bar{u})) - 2\langle \bar{u} - u, g \rangle \leq 2L_\eta(u)\|\bar{u} - u\| + 2\|\bar{u} - u\| \|g\|. \tag{3.20}$$

Notice that the last inequality (3.20) is trivial whenever $\eta(u) \leq \eta(\bar{u})$. This yields the conclusion. \square

Lemma 3.16 *The sequence $\{y_n\}$ is bounded a.s.*

Proof By the characterization of $y_n = S_n(u_n)$ from Lemma 3.14 and (3.19), followed by Jensen’s inequality, and the application of Lemma 3.15 in the fourth inequality, we get

$$\begin{aligned} \|y_n\| &\leq \|\nabla j(u_n)\| + \|\frac{1}{t_n} \mathbb{E}[u_n - t_n g_n - \text{prox}_{t_n h}(u_n - t_n g_n) | \mathcal{F}_n]\| \\ &\leq \|\nabla j(u_n)\| + \mathbb{E}\left[\left\|\frac{1}{t_n} (u_n - t_n g_n - \text{prox}_{t_n h}(u_n - t_n g_n))\right\| \middle| \mathcal{F}_n\right] \\ &\leq \|\nabla j(u_n)\| + \mathbb{E}[\|g_n\| | \mathcal{F}_n] + \mathbb{E}\left[\left\|\frac{1}{t_n} (u_n - \text{prox}_{t_n h}(u_n - t_n g_n))\right\| \middle| \mathcal{F}_n\right] \\ &\leq \|\nabla j(u_n)\| + \mathbb{E}[\|g_n\| | \mathcal{F}_n] + 2L_\eta(u_n) + 2\mathbb{E}[\|g_n\| | \mathcal{F}_n] \\ &\leq \|\nabla j(u_n)\| + 3\sqrt{M(u_n)} + 2L_\eta(u_n). \end{aligned} \tag{3.21}$$

The last step follows by $\mathbb{E}[\|g_n\| | \mathcal{F}_n] = \mathbb{E}[\|G(u_n, \xi)\|]$ and Assumption 3.11(iii) with Jensen’s inequality. We have from Assumption 3.1(i) that $\{u_n\}$ is bounded a.s.; therefore, all terms on the right-hand side of (3.21) are bounded a.s. \square

For Lemma 3.18, we need the following result, which is a generalization of a convergence theorem for quadratic variations from [57, p. 111] to Bochner spaces. The proof can be found in Sect. A.

Lemma 3.17 *Let $\{v_n\}$ be an H -valued martingale. Then $\{v_n\}$ is bounded in $L^2(\Omega, H)$ if and only if*

$$\sum_{n=1}^{\infty} \mathbb{E}[\|v_{n+1} - v_n\|^2] < \infty, \tag{3.22}$$

and when this is satisfied, $v_n \rightarrow v_\infty$ a.s. as $n \rightarrow \infty$.

Lemma 3.18 *The series $\sum_{j=1}^N t_j w_j$ a.s. converges to a limit as $N \rightarrow \infty$.*

Proof Recall the elementary inequality $\mathbb{E}[\|X - \mathbb{E}[X | \mathcal{F}_n]\|^2 | \mathcal{F}_n] \leq \mathbb{E}[\|X\|^2 | \mathcal{F}_n]$, which holds for any random variable X . By Lemma 3.14 with

$$X := \frac{1}{t_n}(\text{prox}_{t_n h}(u_n - t_n g_n) - u_n),$$

followed by Lemma 3.15 and Assumption 3.11(iii), we get

$$\begin{aligned} \mathbb{E}[\|w_n\|^2 | \mathcal{F}_n] &\leq \frac{1}{t_n^2} \mathbb{E}[\|\text{prox}_{t_n h}(u_n - t_n g_n) - u_n\|^2 | \mathcal{F}_n] \\ &\leq 4(L_\eta(u_n))^2 + 4M(u_n) < \infty. \end{aligned} \tag{3.23}$$

Let $v_n := \sum_{j=1}^n t_j w_j$. We show that v_n is a square integrable martingale, i.e., $v_n \in L^2(\Omega, H)$ for every n and $\sup_n \mathbb{E}[\|v_n\|^2] < \infty$. It is clearly a martingale, since for all n , $\mathbb{E}[w_n | \mathcal{F}_n] = 0$ and thus

$$\mathbb{E}[v_n | \mathcal{F}_n] = \mathbb{E}[t_n w_n | \mathcal{F}_n] + \sum_{j=1}^{n-1} t_j w_j = v_{n-1}.$$

To show that v_n is square integrable, we use (3.23) and the fact that $\mathbb{E}[v_n] = 0$ for all n to conclude that its quadratic variations are bounded. Indeed,

$$A_n := \sum_{j=2}^n \mathbb{E}[\|v_j - v_{j-1}\|^2 | \mathcal{F}_j] = \sum_{j=2}^n t_j^2 \mathbb{E}[\|w_j\|^2 | \mathcal{F}_j].$$

Because of the condition (3.16), we have that $\sup_n \mathbb{E}[A_n] < \infty$. We have obtained that $\{v_n\}$ is square integrable, so by Lemma 3.17, it follows that $\{v_n\}$ converges a.s. to a limit as $n \rightarrow \infty$. □

Lemma 3.19 *The following is true with probability one:*

$$\lim_{n \rightarrow \infty} \|u_{n+1} - u_n\| = 0. \tag{3.24}$$

Proof This is a simple consequence of (3.18) and a.s. boundedness of y_n , r_n , and w_n for all n by Lemmas 3.16, 3.18, and Assumption 3.1(iii), respectively. □

Lemma 3.20 *For any sequence $\{z_n\}$ in C such that $z_n \rightarrow z$ as $n \rightarrow \infty$, it follows that*

$$\lim_{m \rightarrow \infty} d\left(\frac{1}{m} \sum_{n=1}^m S_n(z_n), S(z)\right) = 0 \quad \text{a.s.} \tag{3.25}$$

Proof Notice that C is closed, so $z \in C$. The fact that $S(z)$ is nonempty, closed, and convex follows by these properties of $\nabla j(z)$, $\partial \eta(z)$, and $N_C(z)$. We define $g_n^\xi := G(z_n, \xi)$ and

$$\tilde{S}_n(z_n, \xi) := -\nabla j(z_n) - \frac{1}{t_n}(z_n - t_n g_n^\xi - \text{prox}_{t_n h}(z_n - t_n g_n^\xi)). \tag{3.26}$$

Clearly, $\mathbb{E}_\xi[\tilde{S}_n(z_n, \xi)] = S_n(z_n)$. Now, by Jensen's inequality and convexity of the mapping $u \mapsto d(u, S(z))$,

$$\begin{aligned}
 d\left(\frac{1}{m} \sum_{n=1}^m S_n(z_n), S(z)\right) &\leq \frac{1}{m} \sum_{n=1}^m d(S_n(z_n), S(z)) \\
 &\leq \frac{1}{m} \sum_{n=1}^m \mathbb{E}_\xi [d(\tilde{S}_n(z_n, \xi), S(z))].
 \end{aligned}$$

Notice that $\bar{z} = \text{prox}_{th}(u)$ if and only if $0 \in \partial\eta(\bar{z}) + N_C(\bar{z}) + \frac{1}{t}(\bar{z} - u)$, so with

$$\bar{z}_n := \text{prox}_{t_n h}(z_n - t_n g_n^\xi), \tag{3.27}$$

there exist $\zeta_{\eta,n} \in \partial\eta(\bar{z}_n)$ and $\zeta_{C,n} \in N_C(\bar{z}_n)$ such that

$$-(\zeta_{\eta,n} + \zeta_{C,n}) = \frac{1}{t_n}(\bar{z}_n - z_n + t_n g_n^\xi). \tag{3.28}$$

Because $\{z_n\}$ converges, it is contained in a bounded set. Hence, by Lemma 3.15, we get

$$\|\zeta_{\eta,n} + \zeta_{C,n}\| = \frac{1}{t_n} \|\bar{z}_n - z_n + t_n g_n^\xi\| \leq 2L_\eta(z_n) + 3\|g_n^\xi\|, \tag{3.29}$$

which must be almost surely finite by Assumption 3.11(iv). Now, by (3.26) and (3.27), followed by (3.28),

$$\begin{aligned}
 d(\tilde{S}_n(z_n, \xi), S(z)) &= d(-\nabla j(z_n) + \frac{1}{t_n}(\bar{z}_n - z_n + t_n g_n^\xi), S(z)) \\
 &= d(-\nabla j(z_n) - \zeta_{\eta,n} - \zeta_{C,n}, S(z)).
 \end{aligned}$$

By the simple rule $d(u + v, A + B) \leq d(u, A) + d(v, B)$ for sets A and B and points $u, v \in H$, we get by definition of $S(z)$ that

$$d(\tilde{S}_n(z_n, \xi), S(z)) \leq \|\nabla j(z_n) - \nabla j(z)\| + d(\zeta_{\eta,n}, \partial\eta(z)) + d(\zeta_{C,n}, N_C(z)).$$

By strong-to-weak sequential closedness of $\text{gra}(\partial\eta)$ and $\text{gra}(N_C)$ as well as continuity of ∇j , it follows that

$$\lim_{n \rightarrow \infty} d(\tilde{S}_n(z_n, \xi), S(z)) = 0 \quad \text{a.s.} \tag{3.30}$$

We show that $d(\tilde{S}_n(z_n, \xi), S(z))$ is almost surely bounded by an integrable function $\tilde{M}(z)$ for all n . Using elementary arguments and (3.29) in the third inequality,

$$\begin{aligned}
 &d(\tilde{S}_n(z_n, \xi), S(z)) \\
 &\leq d(-\nabla j(z_n) - \zeta_{\eta,n} - \zeta_{C,n}, S(z)) \\
 &\leq \|\nabla j(z_n) - \nabla j(z)\| + d(\zeta_{\eta,n} + \zeta_{C,n}, \partial\eta(z) + N_C(z)) \\
 &\leq \|\nabla j(z_n) - \nabla j(z)\| + 2L_\eta(z_n) + 3\|g_n^\xi\| + d(0, \partial\eta(z) + N_C(z)) \\
 &\leq \sup_{n \in \mathbb{N}} \{ \|\nabla j(z_n) - \nabla j(z)\| + 2L_\eta(z_n) + 3\|g_n^\xi\| + d(0, \partial\eta(z) + N_C(z)) \},
 \end{aligned}$$

which is almost surely bounded by Assumption 3.11(ii) and Assumption 3.11(iv). By the dominated convergence theorem, it follows by (3.30) that as $n \rightarrow \infty$, $\mathbb{E}_\xi[d(\tilde{S}_n(z_n, \xi), S(z))] \rightarrow 0$. Finally, (3.25) follows from the fact that if $a_n \rightarrow 0$ as $n \rightarrow \infty$, it follows that $\frac{1}{m} \sum_{n=1}^m a_n \rightarrow 0$ as $m \rightarrow \infty$. \square

Now we will show a compactness result, adapted from [15], namely that in the limit, the time shifts of the linear interpolation of the sequence $\{u_n\}$ can be made arbitrarily close to trajectories, or solutions, of the differential inclusion

$$\dot{z}(t) \in S(z(t)). \tag{3.31}$$

The set $C(I, H)$ denotes the space of continuous functions from I to H . We recall that if $z(\cdot) \in C([0, \infty), H)$ satisfies (3.31) and is absolutely continuous on any compact interval $[a, b] \subset (0, \infty)$, it is called a strong solution. The existence and uniqueness of this solution is guaranteed by the following result.

Proposition 3.21 *For every $z_0 = z(0) \in C$ there exists a unique strong solution $z \in C([0, \infty), H)$ to the differential inclusion (3.31).*

Proof The function $u \mapsto \eta(u) + \delta_C(u)$ is proper, convex, and lower semicontinuous and $B := -\nabla j$ is Lipschitz continuous. Therefore, by [8, Proposition 3.12], the statement follows. \square

For the next result, we set $s_n := \sum_{j=1}^{n-1} t_j$ and define the linear interpolation $u : [0, \infty) \rightarrow H$ of iterates as well as the piecewise constant extension $y : [0, \infty) \rightarrow H$ of the sequence $\{y_n\}$ via

$$u(t) := u_n + \frac{t - s_n}{s_{n+1} - s_n} (u_{n+1} - u_n), \quad y(t) := y_n, \quad \forall t \in [s_n, s_{n+1}), \forall n \in \mathbb{N}. \tag{3.32}$$

The time shifts of $u(\cdot)$ are denoted by $u(\cdot + \tau)$ for $\tau > 0$. We define $u^\tau : [0, \infty) \rightarrow H$ by

$$u^\tau(t) := u(\tau) + \int_\tau^t y(s) \, ds \tag{3.33}$$

as the solution to the ODE

$$\dot{u}^\tau(\cdot) = y(\cdot), \quad u^\tau(\tau) = u(\tau),$$

which is guaranteed to exist by [9, Theorem 1.4.35].

Theorem 3.22 *For any $T > 0$ and any nonnegative sequence $\{\tau_n\}$, the sequence of the time shifts $\{u(\cdot + \tau_n)\}$ is relatively compact in $C([0, T], H)$. If $\tau_n \rightarrow \infty$, all limit points $\bar{u}(\cdot)$ of the time shifts $\{u(\cdot + \tau_n)\}$ are in $C([0, T], H)$ and there exists a $\bar{y} : [0, T] \rightarrow H$ such that $\bar{y}(t) \in S(\bar{u}(t))$ and $\bar{u}(t) = \bar{u}(0) + \int_0^t \bar{y}(s) \, ds$.*

Proof Relative compactness of time shifts. We first claim that for all $T > 0$,

$$\lim_{\tau \rightarrow \infty} \sup_{t \in [\tau, \tau+T]} \|u^\tau(t) - u(t)\| = 0 \quad \text{a.s.} \tag{3.34}$$

We consider a fixed (but arbitrary) sample path $\omega = (\omega_1, \omega_2, \dots)$ throughout the proof. Let $p := \min\{n : s_n \geq \tau\}$ and $q := \max\{n : s_n \leq t\}$. By (3.33) and (3.32),

$$u^\tau(t) = u(\tau) + \int_\tau^t y(s) \, ds = u(\tau) + \int_\tau^{s_p} y(s) \, ds + \sum_{\ell=p}^{q-1} t_\ell y_\ell + \int_{s_q}^t y(s) \, ds. \tag{3.35}$$

Notice that due to the recursion (3.18),

$$\sum_{\ell=p}^{q-1} t_\ell y_\ell = u_q - u_p - \sum_{\ell=p}^{q-1} t_\ell (w_\ell - r_\ell). \tag{3.36}$$

Plugging (3.36) into (3.35), we get

$$\begin{aligned} u^\tau(t) - u(t) &= u(\tau) + u_q - u_p - u(t) + \int_\tau^{s_p} y(s) \, ds \\ &\quad - \sum_{\ell=p}^{q-1} t_\ell (w_\ell - r_\ell) + \int_{s_q}^t y(s) \, ds. \end{aligned}$$

Therefore,

$$\begin{aligned} \|u^\tau(t) - u(t)\| &\leq \left\| u(\tau) - u_p + \int_\tau^{s_p} y(s) \, ds \right\| + \left\| u_q - u(t) + \int_{s_q}^t y(s) \, ds \right\| \\ &\quad + \left\| \sum_{\ell=p}^{q-1} t_\ell w_\ell \right\| + \left\| \sum_{\ell=p}^{q-1} t_\ell r_\ell \right\|. \end{aligned}$$

Note that by (3.32), it follows that

$$\begin{aligned} \|u(\tau) - u_p\| &\leq \|u_{p-1} - u_p\| = t_{p-1} \|y_{p-1} - r_{p-1} + w_{p-1}\|, \\ \|u_q - u(t)\| &\leq \|u_q - u_{q+1}\| = t_q \|y_q - r_q + w_q\|. \end{aligned}$$

Moreover, by (3.32), we have

$$\left\| \int_\tau^{s_p} y(s) \, ds \right\| \leq t_{p-1} \|y_{p-1}\| \quad \text{and} \quad \left\| \int_{s_q}^t y(s) \, ds \right\| \leq t_q \|y_q\|.$$

Therefore,

$$\begin{aligned} \|u^\tau(t) - u(t)\| &\leq t_{p-1}(2\|y_{p-1}\| + \|r_{p-1}\| + \|w_{p-1}\|) \\ &\quad + t_q(2\|y_q\| + \|r_q\| + \|w_q\|) + \left\| \sum_{\ell=p}^{q-1} t_\ell w_\ell \right\| + \left\| \sum_{\ell=p}^{q-1} t_\ell r_\ell \right\|. \end{aligned} \tag{3.37}$$

We take the limit $p, q \rightarrow \infty$ on the right-hand side of (3.37) and observe that by Lemma 3.16, $\lim_{n \rightarrow \infty} \sup_{m \geq n} t_m \|y_m\| = 0$ and by Lemma 3.18, we have $\lim_{n \rightarrow \infty} \sup_{m \geq n} \left\| \sum_{\ell=n}^{m-1} t_\ell w_\ell \right\| = 0$ as well as $\lim_{n \rightarrow \infty} \sup_{m \geq n} t_m \|w_m\| = 0$. By Assumption 3.1(iii), we have $\lim_{n \rightarrow \infty} \sup_{m \geq n} \left\| \sum_{\ell=n}^{m-1} t_\ell r_\ell \right\| = 0$. We have shown (3.34), so it follows that the set

$$A := \{u^\tau(\cdot) : \tau \in [0, \infty)\}$$

is a family of equicontinuous functions.

To invoke the Arzelà–Ascoli theorem, we first show that the set

$$A(t) := \{u^\tau(t) : \tau \in [0, \infty)\}$$

is relatively compact for all $t \in [0, T], T > 0$. We show this by proving that arbitrary sequences in $A(t)$ have a Cauchy subsequence, which converge in H by completeness of H . To this end, let $\varepsilon > 0$ be arbitrary and observe first the case $\tau_n \rightarrow \infty$. Let n_k be the index such that $\tau_k \in [s_{n_k}, s_{n_k+1})$ and

$$u^{\tau_k}(t) = u_{n_k} + \frac{\tau_k - s_{n_k}}{s_{n_k+1} - s_{n_k}}(u_{n_k+1} - u_{n_k}) + \int_{\tau_k}^t y(s) \, ds.$$

Similarly, let m_j be the index such that $\tau_j \in [s_{m_j}, s_{m_j+1})$. Thus we have

$$\begin{aligned} &\|u^{\tau_k}(t) - u^{\tau_j}(t)\| \\ &\leq \left\| \frac{\tau_k - s_{n_k}}{s_{n_k+1} - s_{n_k}}(u_{n_k+1} - u_{n_k}) - \frac{\tau_j - s_{m_j}}{s_{m_j+1} - s_{m_j}}(u_{m_j+1} - u_{m_j}) \right\| \\ &\quad + \left\| u_{n_k} - u_{m_j} + \int_{\tau_k}^{\tau_j} y(s) \, ds \right\|. \end{aligned} \tag{3.38}$$

Using (3.36), we get (w.l.o.g. $\tau_k \leq \tau_j$)

$$\begin{aligned} \left\| u_{n_k} - u_{m_j} + \int_{\tau_k}^{\tau_j} y(s) \, ds \right\| &\leq \|u_{n_k} - u_{n_k+1}\| + \left\| \int_{\tau_k}^{s_{n_k+1}} y(s) \, ds \right\| \\ &\quad + \left\| \int_{s_{m_j}}^{\tau_j} y(s) \, ds \right\| + \left\| \sum_{\ell=n_k+1}^{m_j-1} t_\ell (w_\ell - r_\ell) \right\|. \end{aligned} \tag{3.39}$$

Combining (3.38) and (3.39), and observing that $\left| \frac{\tau_k - s_{n_k}}{s_{n_k+1} - s_{n_k}} \right| \leq 1$ as well as

$$\left| \frac{\tau_j - s_{m_j}}{s_{m_j+1} - s_{m_j}} \right| \leq 1, \text{ we obtain}$$

$$\begin{aligned} \|u^{\tau_k}(t) - u^{\tau_j}(t)\| &\leq 2\|u_{n_{k+1}} - u_{n_k}\| + \|u_{m_{j+1}} - u_{m_j}\| + t_{n_k}\|y_{n_k}\| \\ &\quad + t_{m_j}\|y_{m_j}\| + \left\| \sum_{\ell=n_k+1}^{m_j-1} t_\ell(w_\ell - r_\ell) \right\|. \end{aligned} \tag{3.40}$$

By Lemma 3.19 as well as convergence of the other terms on the right-hand side of (3.40), for $\varepsilon > 0$ there exists a N such that for all $k, j > N, \|u^{\tau_k}(t) - u^{\tau_j}(t)\| \leq \varepsilon$ for all $k, j > N$ and thus $\{u^{\tau_n}(t)\}$ has a Cauchy subsequence for $\tau_n \rightarrow \infty$. Now we observe the case where the sequence $\{\tau_n\}$ is bounded. Then $\tau_n \rightarrow \bar{\tau}$ for some $\bar{\tau} > 0$ at least on a subsequence (with the same labeling). By convergence of $\{\tau_n\}$ we get that $m_j = n_k$ for $k, j \geq N$ and N large enough. Therefore (3.38) reduces to

$$\|u^{\tau_k}(t) - u^{\tau_j}(t)\| \leq \left| \frac{\tau_k - \tau_j}{s_{n_k+1} - s_{n_k}} \right| \|u_{n_{k+1}} - u_{n_k}\| + \left\| \int_{\tau_k}^{\tau_j} y(s) ds \right\|. \tag{3.41}$$

We can bound terms on the right-hand side of (3.41) as before to obtain that $\{u^{\tau_n}(t)\}$ has a Cauchy subsequence. We have shown that $A(t)$ is relatively compact for all $t \in [0, T], T > 0$, so by the Arzelà–Ascoli theorem, it follows that the set A is relatively compact.

Now, the relative compactness of the set of time shifts $\{u(\cdot + \tau) : \tau \in [0, \infty)\}$ follows from the relative compactness of the set A . Indeed, for any sequence $\{u^{\tau_n}(\cdot + \tau_n)\}$ there exists a convergent subsequence such that $u^{\tau_{n_k}}(\cdot + \tau_{n_k}) \rightarrow \bar{u}(\cdot)$ for some $\bar{u}(\cdot) \in C([0, T], H)$. Now, for the time shift $u(\cdot + \tau_{n_k})$, we have

$$\begin{aligned} &\sup_{t \in [0, T]} \|u(t + \tau_{n_k}) - \bar{u}(t)\| \\ &\leq \sup_{t \in [0, T]} \|u(t + \tau_{n_k}) - u^{\tau_{n_k}}(t + \tau_{n_k})\| + \sup_{t \in [0, T]} \|u^{\tau_{n_k}}(t + \tau_{n_k}) - \bar{u}(t)\|, \end{aligned}$$

so it follows that $u(\cdot + \tau_{n_k}) \rightarrow \bar{u}(\cdot)$ in $C([0, T], H)$ as $\tau_{n_k} \rightarrow \infty$ by convergence of $u^{\tau_{n_k}}(\cdot)$ and (3.34). If $\tau_{n_k} \rightarrow \bar{\tau}$, then $u(\cdot + \tau_{n_k}) \rightarrow u(\cdot + \bar{\tau})$ by uniform continuity of $u(\cdot)$ on $[0, \bar{\tau} + T]$.

Limit points are trajectories of the differential inclusion. Let $\{\tau_n\}$ be a sequence such that as $\tau_n \rightarrow \infty, u^{\tau_n}(\cdot + \tau_n) \rightarrow \bar{u}(\cdot)$ in $C([0, T], H)$ (potentially on a subsequence). The sequence $\{y(\cdot + \tau_n)\} \subset L^2([0, T], H)$ is bounded by boundedness of $\{y_n\}$, and since $L^2([0, T], H)$ is a Hilbert space, there exists a subsequence $\{n_k\}$ such that $y(\cdot + \tau_{n_k}) \rightharpoonup \bar{y}(\cdot)$ in $L^2([0, T], H)$ for some $\bar{y} \in L^2([0, T], H)$. Notice that for $\{\tau_{n_k}\}$, by (3.33) it follows that

$$u^{\tau_{n_k}}(t + \tau_{n_k}) = u^{\tau_{n_k}}(\tau_{n_k}) + \int_0^t y(s + \tau_{n_k}) ds. \tag{3.42}$$

By (3.34), $u^{\tau_{n_k}}(\cdot + \tau_{n_k}) \rightarrow \bar{u}(\cdot)$ in $C([0, T], H)$ as $k \rightarrow \infty$. Taking $k \rightarrow \infty$ on both sides of (3.42) we get, due to $y(\cdot + \tau_{n_k}) \rightharpoonup \bar{y}(\cdot)$ for $t \in [0, T]$, that

$$\bar{u}(t) = \bar{u}(0) + \int_0^t \bar{y}(s) ds.$$

Now, we will show that $\bar{y}(t) \in S(\bar{u}(t))$ for a.e. $t \in [0, T]$. By the Banach-Saks theorem (cf. [42]), there exists a subsequence of $\{y(\cdot + \tau_{n_k})\}$ (where we use the same notation for the sequence as its subsequence) such that

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m y(\cdot + \tau_{n_k}) = \bar{y}(\cdot). \tag{3.43}$$

Recall that $y_n = S_n(u_n)$ by Lemma 3.14 and set $\ell_k^t := \max\{\ell : s_\ell \leq t + \tau_{n_k}\}$. Then we have

$$y(t + \tau_{n_k}) = y(s_{\ell_k^t}) = y_{\ell_k^t} = S_{\ell_k^t}(u_{\ell_k^t}).$$

Therefore, since $t + \tau_{n_k} \in [\ell_k^t, \ell_k^t + 1]$,

$$\begin{aligned} \|u(s_{\ell_k^t}) - \bar{u}(t)\| &\leq \|u(s_{\ell_k^t}) - u(t + \tau_{n_k})\| + \|u(t + \tau_{n_k}) - \bar{u}(t)\| \\ &\leq \|u(s_{\ell_k^t}) - u(s_{\ell_k^t+1})\| + \|u(t + \tau_{n_k}) - \bar{u}(t)\| \\ &\leq t_{\ell_k^t} (\|y_{\ell_k^t}\| + \|r_{\ell_k^t}\| + \|w_{\ell_k^t}\|) + \|u(t + \tau_{n_k}) - \bar{u}(t)\|, \end{aligned} \tag{3.44}$$

which a.s. converges to zero as $k \rightarrow \infty$, since $u(\cdot + \tau_{n_k}) \rightarrow \bar{u}(\cdot)$ and the fact that $t_n \rightarrow 0$ by (3.16) (combined a.s. boundedness of $y_n, r_n,$ and w_n for all n by Lemma 3.16, Assumption 3.1(iii), and Lemma 3.18, respectively). Now, using $y(t + \tau_{n_k}) = y_{\ell_k^t}$, we get

$$\begin{aligned} &d(\bar{y}(t), S(\bar{u}(t))) \\ &\leq \left\| \frac{1}{m} \sum_{k=1}^m y(t + \tau_{n_k}) - \bar{y}(t) \right\| + d\left(\frac{1}{m} \sum_{k=1}^m y(t + \tau_{n_k}), S(\bar{u}(t)) \right) \\ &\leq \left\| \frac{1}{m} \sum_{k=1}^m y(t + \tau_{n_k}) - \bar{y}(t) \right\| + d\left(\frac{1}{m} \sum_{k=1}^m S_{\ell_k^t}(u(s_{\ell_k^t})), S(\bar{u}(t)) \right), \end{aligned}$$

which converges to zero as $m \rightarrow \infty$ by (3.43) and Lemma 3.20, where we note that $u(s_{\ell_k^t}) \rightarrow \bar{u}(t)$ as $k \rightarrow \infty$ by (3.44). Since $S(\bar{u}(t))$ is a closed set and the sample path was chosen to be arbitrary, we have that the statement must be true with probability one. □

Now, we show that there is always a strict decrease in φ along a trajectory that originates at a noncritical point $z(0)$.

Lemma 3.23 *Whenever $z : [0, \infty) \rightarrow C$ is a trajectory satisfying the differential inclusion (3.31) and $0 \notin S(z(0))$, then there exists a $T > 0$ such that*

$$\varphi(z(T)) < \sup_{t \in [0, T]} \varphi(z(t)) \leq \varphi(z(0)). \tag{3.45}$$

Proof We modify the proof from [13, Lemma 5.2]. Let δ, τ satisfying $0 < \delta < \tau$ be fixed but arbitrary. From Theorem 3.21 we have that z is absolutely continuous on $[\delta, \tau]$. It is straightforward to show that $\varphi \circ z : [\delta, \tau] \rightarrow \mathbb{R}$ is absolutely continuous,

since C is bounded and φ is a composition of a locally Lipschitz map with an absolutely continuous function. Therefore, by Rademacher’s theorem, it is differentiable for almost every $t \in [\delta, \tau]$. On the other hand, notice that since η is locally Lipschitz near $z(t)$ and convex, it is Clarke regular, so the chain rule $\partial(\eta \circ z)(t) = \partial\eta(z(t)) \circ \dot{z}(t)$ holds by [11, Theorem 2.3.10]. The chain rule for j holds by differentiability. Therefore for almost every t , it follows for all $v \in \partial\varphi(z(t))$ that

$$(\varphi \circ z)'(t) = \partial(\varphi \circ z)(t) = (\nabla j(z(t)) + \partial\eta(z(t))) \circ \dot{z}(t) = \langle v, \dot{z}(t) \rangle. \tag{3.46}$$

We now observe the following property for the subdifferential of δ_C , namely,

$$\langle v, \dot{z}(t) \rangle = 0 \quad \forall v \in N_C(z(t)). \tag{3.47}$$

Indeed, since $z(\cdot)$ takes values in C and by definition of the subdifferential, for all $r \geq 0$ it follows that

$$0 = \delta_C(z(t+r)) - \delta_C(z(t)) \geq \langle v, z(t+r) - z(t) \rangle.$$

Hence,

$$0 \geq \lim_{r \rightarrow 0^+} \left\langle v, \frac{z(t+r) - z(t)}{r} \right\rangle = \langle v, \dot{z}(t) \rangle.$$

The reverse inequality can be obtained by using the left limit of the difference quotient, and we get (3.47). By (3.46) and (3.47), we obtain for a.e. t that

$$\langle v, \dot{z}(t) \rangle = \partial(\varphi \circ z)(t) \quad \forall v \in -S(z(t)). \tag{3.48}$$

We now show that $\|\dot{z}(t)\| = d(0, S(z(t)))$. Trivially, $d(0, S(z(t))) \leq \|\zeta - 0\|$ for all $\zeta \in S(z(t))$, so it follows that $d(0, S(z(t))) \leq \|\dot{z}(t)\|$. Notice that for all $v, w \in \partial\varphi(z(t))$, by (3.46), $0 = \langle v - w, \dot{z}(t) \rangle$. Setting $W := \text{span}(\partial\varphi(z(t)) - \partial\varphi(z(t)))$, we get $\dot{z}(t) \in W^\perp$. Clearly, $-\dot{z}(t) \in (-\dot{z}(t) + W) \cap W^\perp$ so $\|\dot{z}(t)\| \leq d(0, -\dot{z}(t) + W)$. Since $\partial\varphi(z(t)) \subset \dot{z}(t) + W$, it follows $\|\dot{z}(t)\| \leq d(0, \partial\varphi(z(t)))$ and we get $\|\dot{z}(t)\| = d(0, S(z(t)))$.

Now, notice that by (3.48) and the fact that $\dot{z}(t) \in S(z(t))$, we have for a.e. t that

$$\partial(\varphi \circ z)(t) = -\|\dot{z}(t)\|^2 = -d(0, S(z(t)))^2.$$

Since $\varphi \circ z$ is absolutely continuous on $[\delta, \tau]$,

$$\varphi(z(\tau)) = \varphi(z(\delta)) - \int_\delta^\tau d(0, S(z(s)))^2 ds \tag{3.49}$$

and hence $\varphi(z(\delta)) \geq \varphi(z(\tau))$. Using the continuity of $\varphi \circ z$, and the fact that $0 < \delta < \tau$ were arbitrarily chosen, we get $\varphi(z(0)) \geq \varphi(z(t))$ for all $t > 0$. To finish the proof, we must find some $T > 0$ such that $\varphi(z(T)) < \sup_{t \in [0, T]} \varphi(z(t))$. Suppose that $d(0, S(z(t))) = 0$ for a.e. $t \in [0, T]$ for all $T > 0$. Since $\|\dot{z}(t)\| = d(0, S(z(t)))$ then $z \equiv z(0)$. This is a contradiction, since $\dot{z}(\cdot) \in S(z(\cdot))$ and $0 \notin S(z(0))$. By (3.49), we conclude that there exists a $T > 0$ such that (3.45) holds. \square

The following proof is standard, but we need to make several arguments differently in the infinite-dimensional setting. We will proceed as in [13]. We define the level sets of φ as

$$\mathcal{L}_r := \{u \in H : \varphi(u) \leq r\}.$$

Proposition 3.24 *For all $\varepsilon > 0$ there exists a N such that for all $n \geq N$, if $u_n \in \mathcal{L}_\varepsilon$, then $u_{n+1} \in \mathcal{L}_{2\varepsilon}$ a.s.*

Proof First, we remark that φ is uniformly continuous on V , since $\eta(\cdot)$ satisfies (3.17) and, in turn, is Lipschitz continuous on V , as well as the fact that j is Lipschitz continuous on V . Therefore, for any $\varepsilon > 0$ there exists a $\delta > 0$ such that if $\|u_{n+1} - u_n\| < \delta$, then $|\varphi(u_{n+1}) - \varphi(u_n)| < \varepsilon$. Now, we choose N such that $\|u_{n+1} - u_n\| < \delta$ for all $n \geq N$, which is possible by Lemma 3.19. Then it must follow that $|\varphi(u_{n+1}) - \varphi(u_n)| < \varepsilon$ for all $n \geq N$ as well. Now, since $u_n \in \mathcal{L}_\varepsilon$, it follows that $\varphi(u_{n+1}) \leq 2\varepsilon$, so therefore $u_{n+1} \in \mathcal{L}_{2\varepsilon}$. □

Lemma 3.25 *The following equalities hold.*

$$\liminf_{n \rightarrow \infty} \varphi(u_n) = \liminf_{t \rightarrow \infty} \varphi(u(t)) \quad \text{and} \quad \limsup_{n \rightarrow \infty} \varphi(u_n) = \limsup_{t \rightarrow \infty} \varphi(u(t)). \tag{3.50}$$

Proof We argue that $\liminf_{n \rightarrow \infty} \varphi(u_n) \leq \liminf_{t \rightarrow \infty} \varphi(u(t))$; the other direction is clear by construction of $u(\cdot)$ from (3.32). Let $\{\tau_n\}$ be a sequence such that $\tau_n \rightarrow \infty$, $\lim_{n \rightarrow \infty} u(\tau_n) = \bar{u}$ for some $\bar{u} \in H$, and $\liminf_{n \rightarrow \infty} \varphi(u(\tau_n)) = \varphi(\bar{u})$. With $k_n := \max\{n : t_k \leq \tau_n\}$, we get

$$\|u_{k_n} - \bar{u}\| \leq \|u_{k_n} - u(\tau_n)\| + \|u(\tau_n) - \bar{u}\| \leq \|u_{k_n} - u_{k_{n+1}}\| + \|u(\tau_n) - \bar{u}\|,$$

which converges to zero as $n \rightarrow \infty$ by (3.24) and convergence of the sequence $\{u(\tau_n)\}$. Therefore $u_{k_n} \rightarrow \bar{u}$ and so by continuity of φ , it follows that

$$\liminf_{t \rightarrow \infty} \varphi(u(t)) = \varphi(\bar{u}) = \lim_{n \rightarrow \infty} \varphi(u_{k_n}) \geq \liminf_{n \rightarrow \infty} \varphi(u_n).$$

Analogous arguments can be made for the claim

$$\limsup_{n \rightarrow \infty} \varphi(u_n) = \limsup_{t \rightarrow \infty} \varphi(u(t)).$$

□

Lemma 3.26 *Only finitely many iterates $\{u_n\}$ are contained in $H \setminus \mathcal{L}_{2\varepsilon}$.*

Proof We choose $\varepsilon > 0$ such that $\varepsilon \notin \varphi(S^{-1}(0))$, which is possible for arbitrarily small ε by Assumption 3.11(v), where we note that $\varphi(S^{-1}(0)) = f(S^{-1}(0))$. We construct the process given by the recursion

$$\begin{aligned}
 i_1 &:= \min\{n : u_n \in \mathcal{L}_\varepsilon \text{ and } u_{n+1} \in \mathcal{L}_{2\varepsilon} \setminus \mathcal{L}_\varepsilon\}, \\
 e_1 &:= \min\{n : n > i_1 \text{ and } u_n \in H \setminus \mathcal{L}_{2\varepsilon}\}, \\
 i_2 &:= \min\{n : n > e_1 \text{ and } u_n \in \mathcal{L}_\varepsilon\},
 \end{aligned}$$

and so on. We argue by contradiction and recall that $s_n = \sum_{j=1}^{n-1} t_j$. Suppose infinitely many $\{u_n\}$ are in $H \setminus \mathcal{L}_{2\varepsilon}$, then it must follow that $i_j \rightarrow \infty$ as $j \rightarrow \infty$. By Theorem 3.22, $\{u(\cdot + s_j)\}$ is relatively compact in $C([0, T], H)$ for all $T > 0$ and there exists a subsequence (with the same labeling) and limit point $z(\cdot)$ such that $z(\cdot)$ is a trajectory of (3.31). Now, since by construction $\varphi(u_{i_j}) \leq \varepsilon$ and $\varphi(u_{i_j+1}) > \varepsilon$, it follows that

$$\begin{aligned}
 \varepsilon &\geq \varphi(u_{i_j}) = \varphi(u_{i_j+1}) + \varphi(u_{i_j}) - \varphi(u_{i_j+1}) \\
 &\geq \varepsilon + \varphi(u_{i_j}) - \varphi(u_{i_j+1}).
 \end{aligned} \tag{3.51}$$

Recall that $\lim_{j \rightarrow \infty} u_{i_j} = u(\cdot + s_j) = z(0)$. Taking the limit $j \rightarrow \infty$ on both sides of (3.51), by continuity of φ , we get

$$\lim_{j \rightarrow \infty} \varphi(u_{i_j}) = \varphi(z(0)) = \varepsilon,$$

meaning $z(0)$ is not a critical point of φ . Thus we can invoke Lemma 3.23 to get the existence of a $T > 0$ such that

$$\varphi(z(T)) < \sup_{t \in [0, T]} \varphi(z(t)) \leq \varphi(z(0)) = \varepsilon. \tag{3.52}$$

By uniform convergence of $u(\cdot + s_j)$ to $z(\cdot)$, it follows for j sufficiently large that

$$\sup_{t \in [0, T]} |\varphi(u(t + s_j)) - \varphi(z(t))| < \varepsilon,$$

so

$$\sup_{t \in [0, T]} \varphi(u(t + s_j)) \leq \sup_{t \in [0, T]} |\varphi(u(t + s_j)) - \varphi(z(t))| + \sup_{t \in [0, T]} \varphi(z(t)) \leq 2\varepsilon.$$

Therefore it must follow that

$$s_{e_j} > s_{i_j} + T \tag{3.53}$$

for j sufficiently large. We now find a contradiction to the statement (3.53). This is done by observing the sequence $\ell_j := \max\{\ell : s_{i_j} \leq s_\ell \leq s_{i_j} + T\}$. From (3.52), we have that there exists a $\delta > 0$ such that $\varphi(z(T)) \leq \varepsilon - 2\delta$. Observe that

$$\|u_{\ell_j} - u(T + s_{i_j})\| = \|u(s_{\ell_j}) - u(T + s_{i_j})\| \leq \|u_{\ell_j} - u_{\ell_j+1}\| \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

Therefore $u_{\ell_j} \rightarrow u(T + s_{i_j})$ and hence $u_{\ell_j} \rightarrow z(T)$ as $j \rightarrow \infty$. By continuity, we get $\lim_{j \rightarrow \infty} \varphi(u_{\ell_j}) = \varphi(z(T))$. Thus $\varphi(u_{\ell_j}) < \varepsilon - \delta$ for j sufficiently large, a contradiction to (3.53). □

Proposition 3.27 *The limit $\lim_{t \rightarrow \infty} \varphi(u(t))$ exists.*

Proof W.l.o.g. assume $\liminf_{t \rightarrow \infty} \varphi(u(t)) = 0$; this is possible by the fact that j and η are bounded below. Choosing $\varepsilon > 0$ such that $\varepsilon \notin \varphi(S^{-1}(0))$, we have by Lemma 3.26 that for N sufficiently large, $u_n \in \mathcal{L}_{2\varepsilon}$ for all $n \geq N$. Since ε can be chosen to be arbitrarily small, we conclude that $\lim_{t \rightarrow \infty} \varphi(u(t)) = 0$. \square

Proof of Theorem 3.13 The fact that $\{\varphi(u_n)\}$ converges follows from Proposition 3.27 and Lemma 3.25. Since $\{u_n\} \subset C$, it trivially follows that $\{f(u_n)\}$ converges a.s. Let \bar{u} be a limit point of $\{u_n\}$ and suppose that $0 \notin S(\bar{u})$. Let $\{u_{n_k}\}$ be a subsequence converging to \bar{u} and let $z(\cdot)$ be the limit of $\{u(\cdot + s_{n_k})\}$. Then, by Lemma 3.23, there exists a $T > 0$ such that

$$\varphi(z(T)) < \sup_{t \in [0, T]} \varphi(z(t)) \leq \varphi(\bar{u}). \tag{3.54}$$

However, it follows from Proposition 3.27 that

$$\varphi(z(T)) = \lim_{k \rightarrow \infty} \varphi(u(T + s_{n_k})) = \lim_{t \rightarrow \infty} \varphi(u(t)) = \varphi(\bar{u}),$$

which is a contradiction to (3.54). \square

4 Application to PDE-constrained optimization under uncertainty

In this section, we apply the algorithm presented in Sect. 3.2 to a nonconvex problem from PDE-constrained optimization under uncertainty. In Sect. 4.1, we set up the problem and verify conditions for convergence of the stochastic proximal gradient method. We show numerical experiments in Sect. 4.2.

4.1 Model problem

We first introduce notation and concepts specific to our application; see [18, 52]. Let $D \subset \mathbb{R}^d$, $d \leq 3$ be an open and bounded Lipschitz domain. The inner product between vectors $x, y \in \mathbb{R}^d$ is denoted by $x \cdot y = \sum_{i=1}^d x_i y_i$. For a function $v : \mathbb{R}^d \rightarrow \mathbb{R}$, let $\nabla v(x) = (\partial v(x)/\partial x_1, \dots, \partial v(x)/\partial x_d)^\top$ denote the gradient and for $w : \mathbb{R}^d \rightarrow \mathbb{R}^d$, let $\nabla \cdot w(x) = \partial w_1(x)/\partial x_1 + \dots + \partial w_d(x)/\partial x_d$ denote the divergence. We define the Sobolev space $H^1(D) = \{u \in L^2(D) \text{ having weak derivatives } \partial u/\partial x_i \in L^2(D), i = 1, \dots, d\}$ and the closure of $C_c^\infty(D)$ in $H^1(D)$ by $H_0^1(D)$.

We will focus on a semilinear diffusion-reaction equation with uncertainties, which describes transport phenomena at equilibrium and is motivated by [41]. We assume that there exist random fields $a : D \times \Omega \rightarrow \mathbb{R}$ and $r : D \times \Omega \rightarrow \mathbb{R}$, which are the diffusion and reaction coefficients, respectively. To facilitate simulation, we will make a standard finite-dimensional noise assumption, meaning the random field has the form

$$a(x, \omega) = a(x, \xi(\omega)), \quad r(x, \omega) = r(x, \xi(\omega)) \quad \text{in } D \times \Omega,$$

where $\xi(\omega) = (\xi_1(\omega), \dots, \xi_m(\omega))$ is a vector of real-valued uncorrelated random variables $\xi_i : \Omega \rightarrow \Xi_i \subset \mathbb{R}$. The support of the random vector will be denoted by $\Xi := \prod_{i=1}^m \Xi_i$. We consider the following PDE constraint, to be satisfied for almost every $\xi \in \Xi$:

$$\begin{aligned} -\nabla \cdot (a(x, \xi)\nabla y(x, \xi)) + r(x, \xi)(y(x, \xi))^3 &= u(x), & (x, \xi) \in D \times \Xi, \\ y(x, \xi) &= 0, & (x, \xi) \in \partial D \times \Xi. \end{aligned} \tag{4.1}$$

Optimal control problems with semilinear PDEs involving random coefficients have been studied in, for instance, [26, 27]. We include a nonsmooth term as in [14] with the goal of obtaining sparse solutions. In the following, we assume that $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, and $y_D \in L^2(D)$. The model problem we solve is given by

$$\begin{aligned} \min_{u \in C} \quad & \left\{ \varphi(u) := \frac{1}{2} \mathbb{E}[\|y(\xi) - y_D\|_{L^2(D)}^2] + \frac{\lambda_2}{2} \|u\|_{L^2(D)}^2 + \lambda_1 \|u\|_{L^1(D)} \right\} \\ \text{s.t.} \quad & -\nabla \cdot (a(x, \xi)\nabla y) + r(x, \xi)y^3 = u(x), \quad (x, \xi) \in D \times \Xi, \\ & y = 0, \quad (x, \xi) \in \partial D \times \Xi, \\ & C := \{u \in L^2(D) : u_a(x) \leq u(x) \leq u_b(x) \text{ a.e. } x \in D\}. \end{aligned} \tag{P'}$$

The following assumptions will apply in this section. In particular, we do not require uniform bounds on the coefficient $a(\cdot, \xi)$, which allow for modeling with log-normal random fields.

Assumption 4.1 We assume $y_D \in L^2(D)$, $u_a, u_b \in L^2(D)$, and $u_a \leq u_b$. There exist $a_{\min}(\cdot), a_{\max}(\cdot)$ such that $0 < a_{\min}(\xi) < a(\cdot, \xi) < a_{\max}(\xi) < \infty$ in D a.s. and $a_{\min}^{-1}, a_{\max} \in L^p(\Xi)$ for all $p \in [1, \infty)$. Furthermore, there exists $r_{\max}(\cdot)$ such that $0 \leq r(\cdot, \xi) \leq r_{\max}(\xi) < \infty$ a.s. and $r_{\max} \in L^p(\Xi)$ for all $p \in [1, \infty)$.

Existence of a solution to Problem (P') follows by applying [27, Proposition 3.1]. The following result holds by [27, Proposition 2.1] combined with standard a priori estimates for a fixed realization ξ to obtain (4.2) and (4.3).

Lemma 4.2 For almost every $\xi \in \Xi$, (4.1) has a unique solution $y(\xi) = y(\cdot, \xi) \in H_0^1(D)$ and there exists a positive random variable $C_1 \in L^p(\Xi)$ for all $p \in [1, \infty)$ independent of u such that for almost every $\xi \in \Xi$,

$$\|y(\xi)\|_{L^2(D)} \leq C_1(\xi)\|u\|_{L^2(D)}. \tag{4.2}$$

Additionally, for $y_1(\xi)$ and $y_2(\xi)$ solving (4.1) with $u = u_1$ and $u = u_2$, respectively, we have for almost every $\xi \in \Xi$ that

$$\|y_1(\xi) - y_2(\xi)\|_{L^2(D)} \leq C_1(\xi)\|u_1 - u_2\|_{L^2(D)}. \tag{4.3}$$

By Lemma 4.2, the control-to-state operator $T(\xi) : L^2(D) \rightarrow H_0^1(D), u \mapsto T(\xi)u$ is well-defined for almost every ξ and all $u \in L^2(D)$. Additionally, for almost

every $\xi \in \Xi$, this mapping is in fact continuously Fréchet differentiable; this can be argued by verifying [23, Assumption 1.47] as in [23, pp. 76-78]. With that, we define the reduced functional $J : L^2(D) \times \Xi \rightarrow \mathbb{R}$ by $J(u, \xi) := \frac{1}{2} \|T(\xi)u - y_D\|_{L^2(D)}^2 + \frac{\lambda_2}{2} \|u\|_{L^2(D)}^2$ and we can define the stochastic gradient.

Proposition 4.3 $J : L^2(D) \times \Xi \rightarrow \mathbb{R}$ is continuously Fréchet differentiable and the stochastic gradient is given by

$$G(u, \xi) := \lambda_2 u - p(\cdot, \xi), \tag{4.4}$$

where, given a solution $y = y(\cdot, \xi)$ to (4.1), the function $p = p(\cdot, \xi) \in H_0^1(D)$ is the solution to the adjoint equation

$$\begin{aligned} -\nabla \cdot (a(x, \xi)\nabla p) + 3r(x, \xi)y^2 p &= y_D - y, & (x, \xi) \in D \times \Xi \\ p &= 0, & (x, \xi) \in \partial D \times \Xi. \end{aligned} \tag{4.5}$$

Furthermore, for almost every $\xi \in \Xi$, with the same $C_1 \in L^p(\Xi)$ for all $p \in [1, \infty)$ as in Lemma 4.2,

$$\|p(\cdot, \xi)\|_{L^2(D)} \leq C_1(\xi) \|y_D - y(\xi)\|_{L^2(D)}. \tag{4.6}$$

Additionally, for $p_1(\xi)$ and $p_2(\xi)$ solving (4.5) with $y = y_1(\xi)$ and $y = y_2(\xi)$, respectively (where $y_i(\xi)$ solves (4.1) with $u = u_i$),

$$\|p_1(\xi) - p_2(\xi)\|_{L^2(D)} \leq C_1(\xi) \|y_1(\xi) - y_2(\xi)\|_{L^2(D)}. \tag{4.7}$$

The proofs of the above and following proposition are in Sect. B. We define $j : L^2(D) \rightarrow \mathbb{R}$ by $j(u) := \mathbb{E}[J(u, \xi)]$ for all $u \in L^2(D)$ and show that it is continuously Fréchet differentiable in the following proposition.

Proposition 4.4 The function $j : L^2(D) \rightarrow \mathbb{R}$ is continuously Fréchet differentiable and $\mathbb{E}[G(u, \xi)] = \nabla j(u)$ for all $u \in L^2(D)$.

Now, we present the main result of this section, which is the verification of assumptions for the convergence of Algorithm 2.

Theorem 4.5 Problem (P') satisfies Assumption 3.1(ii) as well as Assumptions 3.11(i)–3.11(iv).

Proof For Assumption 3.1(ii), we note that by Proposition 4.4, j is continuously Fréchet differentiable and $\mathbb{E}[G(u, \xi)] = \nabla j(u)$ for all $u \in L^2(D)$. Now, for arbitrary $u_1, u_2 \in L^2(D)$, we have by Jensen’s inequality, (4.4), and Hölder’s inequality applied to (4.7) and (4.3) that

$$\begin{aligned} \|\nabla j(u_1) - \nabla j(u_2)\|_{L^2(D)} &\leq \mathbb{E}[\|G(u_1, \xi) - G(u_2, \xi)\|_{L^2(D)}] \\ &\leq \lambda_2 \|u_1 - u_2\|_{L^2(D)} + \mathbb{E}[\|p_1(\xi) - p_2(\xi)\|_{L^2(D)}] \\ &\leq \lambda_2 \|u_1 - u_2\|_{L^2(D)} + (\mathbb{E}[(C_1(\xi))^2])^{1/2} \left(\mathbb{E}[\|y_1(\xi) - y_2(\xi)\|_{L^2(D)}^2] \right)^{1/2} \\ &\leq \lambda_2 \|u_1 - u_2\|_{L^2(D)} + \|C_1\|_{L^2(\Xi)}^2 \|u_1 - u_2\|_{L^2(D)}. \end{aligned}$$

Since $\|C_1\|_{L^2(\Xi)}^2 < \infty$ it follows that $j \in C_L^{1,1}(L^2(D))$.

Assumption 3.11(i) is obviously satisfied. For Assumption 3.11(ii), we have that the function $\eta(u) = \lambda_1 \|u\|_{L^1(D)} \in \Gamma_0(L^2(D))$ and is clearly bounded below; additionally, η is globally Lipschitz and therefore satisfies (3.17). For Assumption 3.11(iii), we have by (4.4), (4.6), and (4.2) the bound

$$\|G(u, \xi)\|_{L^2(D)} \leq \lambda_2 \|u\|_{L^2(D)} + C_1(\xi) \|y_D\|_{L^2(D)} + (C_1(\xi))^2 \|u\|_{L^2(D)} \tag{4.8}$$

and furthermore $\mathbb{E}[\|G(u, \xi)\|_{L^2(D)}^2] =: M(u) < \infty$ by integrability of $\xi \mapsto C_1(\xi)$. Assumption 3.11(iv) follows for any $u \in C$ (and hence any convergent sequence $\{u_n\}$ in C) by (4.8). □

The last assumption from Assumption 3.11 is technical and difficult to verify for general functions in infinite dimensions. Indeed, [31] gave an example of a C^∞ -function whose critical values make up a set of measure greater than zero. In finite dimensions the story is easier: the Morse–Sard theorem guarantees that Assumption 3.11(v) holds if $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $f \in C^k$ for $k \geq n$. In infinite dimensions, certain well-behaved functions, in particular Fredholm operators, see [50], satisfy this assumption.

4.2 Numerical experiments

In this section, we demonstrate Algorithm 2 on Problem (P'). Simulations were run using FEniCS by [2] on a laptop with Intel Core i7 Processor (8 x 2.6 GHz) with 16 GB RAM. Let the domain be given by $D = (0, 1) \times (0, 1)$ and the constraint set be given by $C = \{u \in L^2(D) \mid -0.5 \leq u(x) \leq 0.5 \ \forall x \in D\}$. We modify [14, Example 6.1], with $y_D(x) = \sin(2\pi x_1) \sin(2\pi x_2) \exp(2x_1)/6$, $\lambda_1 = 0.008$, and $\lambda_2 = 0.001$. We generate random fields using a Karhunen–Loève expansion, with means $a_0 = 0.5$ and $r_0 = 0.5$, number of summands $m = 20$, and $\xi^{a,i}, \xi^{r,i} \sim U(-\sqrt{0.5}, \sqrt{0.5})$, where $U(a, b)$ denotes the uniform distribution between real numbers a and b , $a < b$. The eigenfunctions and eigenvalues are given by

$$\tilde{\phi}_{j,k}(x) := 2 \cos(j\pi x_2) \cos(k\pi x_1), \quad \tilde{\lambda}_{k,j} := \frac{1}{4} \exp(-\pi(j^2 + k^2)l^2), \quad j, k \geq 1,$$

where we reorder terms so that the eigenvalues appear in descending order (i.e., $\phi_1 = \tilde{\phi}_{1,1}$ and $\lambda_1 = \tilde{\lambda}_{1,1}$) and we choose correlation length $l = 0.5$. Thus

$$a(x, \xi) = a_0 + \sum_{i=1}^m \sqrt{\lambda_i} \phi_i \xi^{a,i}, \quad r(x, \xi) = r_0 + \sum_{i=1}^m \sqrt{\lambda_i} \phi_i \xi^{r,i}. \tag{4.9}$$

For Algorithm 2, we generate samples with $\xi_n = (\xi_n^{a,1}, \dots, \xi_n^{a,m}, \xi_n^{r,a}, \dots, \xi_n^{r,m})$ at each iteration n . The step size is chosen to be $t_n = \theta/n$ with $\theta = 100$, where the scaling was chosen such that $\theta \approx 1/\|G(u_1, \xi_1)\|$. The initial point was $u_1(x) = \sin(4\pi x_1) \sin(4\pi x_2)$.

A uniform mesh \mathcal{T} with 9800 shape regular triangles T was used. We denote the mesh fineness with $\hat{h} = \max_{T \in \mathcal{T}} \text{diam}(T)$. The state and adjoint were discretized using piecewise linear finite elements, (where \mathcal{P}_i denotes the space of polynomials of degree up to i), given by the set

$$V_{\hat{h}} := \{v \in H_0^1(D) : v|_T \in \mathcal{P}_1(T) \text{ for all } T \in \mathcal{T}\}.$$

For the controls, we choose a discretization of $L^2(D)$ by piecewise constants, given by the set

$$U_{\hat{h}} := \{u \in L^2(D) : v|_T \in \mathcal{P}_0(T) \text{ for all } T \in \mathcal{T}\}, \quad C_{\hat{h}} := U_{\hat{h}} \cap C.$$

We use the L^2 -projection $P_{\hat{h}} : L^2(D) \rightarrow U_{\hat{h}}$ defined for each $v \in L^2(D)$ by

$$P_{\hat{h}}(v)|_T := \frac{1}{|T|} \int_T v \, dx.$$

This is done to project the stochastic gradient onto the $L^2(D)$ space as in [20]. Hence, the last line of Algorithm 2 is given by the expression $u_{n+1} := \text{prox}_{t_n h}(u_n - t_n P_{\hat{h}} G(u_n, \xi_n))$. For the computation of the proximity operator $\text{prox}_{t(\eta+\delta_C)}(z) = \arg \min_{-0.5 \leq v \leq 0.5} \{\lambda_1 \|v\|_{L^1(D)} + \frac{1}{2t} \|v - z\|_{L^2(D)}^2\}$, we use the formula from [5, Example 6.22], defined piecewise on each element of the mesh. For each $T \in \mathcal{T}$, it is given by

$$\text{prox}_{t(\eta+\delta_C)}(z|_T) = \min\{\max\{|z|_T| - t\lambda_1, 0\}, 0.5\} \text{sgn}(z|_T).$$

For convergence plots, we use a heuristic to approximate the objective function and the measure of stationarity by increasing sampling as the control reaches stationarity. To be more precise, we use a sequence of sample sizes $\{m_n\}$ with $m_n = 10 \lfloor \frac{n}{50} \rfloor + 1$ newly generated i.i.d. samples $(\xi_{n,1}, \dots, \xi_{n,m_n})$ and compute

$$\hat{f}_n := \frac{1}{m_n} \sum_{j=1}^{m_n} J(u_n, \xi_{n,j}) + \eta(u_n),$$

$$r_n := \left\| u_n - \text{prox}_{\eta+\delta_C} \left(u_n - \frac{1}{m_n} \sum_{j=1}^{m_n} P_{\hat{h}} G(u_n, \xi_{n,j}) \right) \right\|_{L^2(D)}.$$

The algorithm is terminated for $n \geq 50$ if $\hat{r}_n := \sum_{k=n-50}^n r_n \leq \text{tol}$ with $\text{tol} = 2e^{-4}$. The parameters for our heuristic termination rule were tuned, for illustration purposes only, so that the algorithm stopped after several hundred iterations. A plot of the control after termination is shown in Fig. 1. The effect of the sparse term η as well as the constraint set C can be seen clearly. Decay of the objective function value

Fig. 1 The control u after 251 iterations

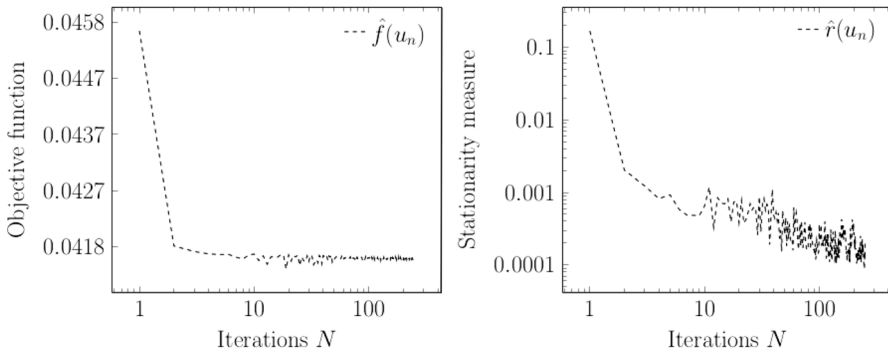
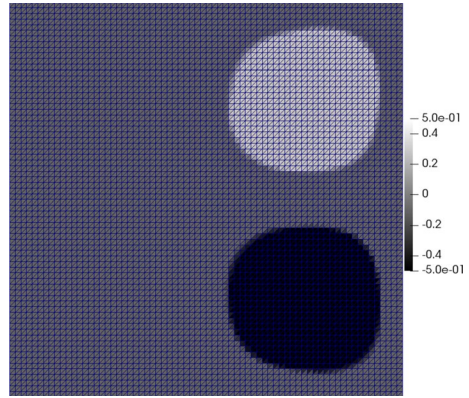


Fig. 2 Behavior of the objective function (left) and the stationarity measure (right)

and the stationarity measure are shown in Fig. 2. We see convergence of the objective function values and the stationarity measure tends to zero as expected.

Additionally, we conduct an experiment to demonstrate mesh independence of the algorithm by running the algorithm once each for different meshes and comparing the number of iterations needed until the tolerance tol is reached. In Table 1, we see that these iteration numbers are of the same order. The estimate for the objective function \hat{f}_N is also included at the final iteration N , demonstrating how solutions become more exact on finer meshes.

5 Conclusion

In this paper, we presented asymptotic convergence analysis for two variants of the stochastic proximal gradient algorithm in Hilbert spaces. The main results address the asymptotic convergence to stationary points of general functions defined over a Hilbert space. Moreover, we presented an application to the theory in the form of a problem from PDE-constrained optimization under uncertainty. Assumptions

Table 1 Experiment showing mesh independence

\hat{h}	# triangles	Objective function \hat{f}_N	# iterations N until $t_N \leq \text{tol}$
$7.1e^{-2}$	800	$4.160e^{-2}$	191
$4.7e^{-2}$	1800	$4.157e^{-2}$	295
$3.5e^{-2}$	3200	$4.157e^{-2}$	233
$2.8e^{-2}$	5000	$4.156e^{-2}$	257
$2.4e^{-2}$	7200	$4.156e^{-2}$	271
$2.0e^{-2}$	9800	$4.155e^{-2}$	251

for convergence were verified for a tracking-type problem with a L^1 -penalty term subject to a semilinear elliptic PDE with random coefficients and box constraints. Numerical experiments demonstrated the effectiveness of the method.

The ODE method from Sect. 3.2 allowed us to prove a more general result with weaker assumptions on the objective function. However, we needed to introduce an assumption on the set of critical values in the form of Assumption 3.11(v). While we did not verify this assumption for our model problem, it would be interesting to know whether this assumption is verifiable for this class of problems. We had to be slightly more restrictive on the nonsmooth term in Sect. 3.2 than we were in Sect. 3.1. The advantages in terms of computational cost of Algorithm 2 over Algorithm 1 are clear: the use of decreasing step sizes in Algorithm 2 means that increased sampling is not needed. Additionally, there is no need to determine the Lipschitz constant for the gradient, which in the application depends on (among other things) the Poincaré constant and the lower bound on the random fields, and thus lead to a prohibitively small constant step size. This phenomenon has been demonstrated in [20].

How to scale the decreasing step size t_n remains an open question. In practice, the scaling of the step size can be tuned offline. An improper choice of the scaling c in the step size $t_n = c/n^\alpha$ for $0.5 < \alpha \leq 1$ can lead to arbitrarily slow convergence; this was demonstrated in [39]. While this was not the focus of our work, efficiency estimates for nonconvex problems might also be possible following the work by [7, 21, 35]. In lieu of efficiency estimates, it would be desirable to have better termination conditions that do not rely on increased sampling as our heuristic did in the numerical experiments. Finally, it would be natural to investigate mesh refinement strategies as in [20]. For more involved choices of nonsmooth terms, the prox computation is also subject to numerical error and should be treated.

Acknowledgements The authors would like to thank the anonymous reviewers for their thoughtful comments and efforts towards improving our manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article

are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Auxiliary results

To prove Lemma 3.17, we first need the following result.

Proposition A.1 *For $1 \leq p \leq \infty$, every H -valued martingale that is bounded in the Bochner space $L^p(\Omega, H)$ converges a.s.*

Proof Since H is a Hilbert space, it is reflexive and therefore has the Radon–Nikodym property by [43, Corollary 2.11]. The rest of the proof can be found in [43, Theorem 2.5]. \square

Proof of Lemma 3.17 It is straightforward to show that $\mathbb{E}[\|v_n\|^2] = \mathbb{E}[\|v_1\|^2] + \sum_{k=1}^{n-1} \mathbb{E}[\|v_{k+1} - v_k\|^2]$, and therefore boundedness of v_n for all n follows from (3.22) and vice versa. Supposing now that (3.22) holds, the fact that $\{v_n\}$ converges to a limit v_∞ follows by Proposition A.1. \square

B Auxiliary proofs for application

Proof of Proposition 4.3 Continuous differentiability of $J(\cdot, \xi) : L^2(D) \rightarrow \mathbb{R}$ follows from continuous differentiability of $u \mapsto T(\xi)u$ and the fact that $(u, y) \mapsto \tilde{J}(u, y) = \frac{1}{2}\|y - y_D\|_{L^2(D)}^2 + \frac{\lambda_2}{2}\|u\|_{L^2(D)}^2$ is continuously Fréchet differentiable. One obtains (4.4) and (4.5) by fixing a realization $\xi \in \Xi$ and computing the derivative of $u \mapsto J(u, \xi)$ as in, e.g., [23, pp. 58–59]. Bounds (4.6) and (4.7) follow from standard a priori estimates. \square

Proof of Proposition 4.4 We verify the conditions of Lemma C.3 from Sect. C. Fréchet differentiability of $J : L^2(D) \times \Xi \rightarrow \mathbb{R}$ for almost every ξ follows from Proposition 4.3. The function j is well-defined and finite-valued for all $u \in L^2(D)$, since

$$j(u) = \frac{1}{2}\mathbb{E}[\|y - y_D\|_{L^2(D)}^2] + \frac{\lambda_2}{2}\|u\|_{L^2(D)}^2 \leq \mathbb{E}[\|T(\xi)u\|_{L^2(D)}^2] + \|y_D\|_{L^2(D)}^2 + \frac{\lambda_2}{2}\|u\|_{L^2(D)}^2$$

is finite by $T(\xi)u = y(\xi)$ and (4.2) along with the assumption that $y_D \in L^2(D)$. Now, for every $v \in C$, there exists a $y_v(\xi)$ satisfying (4.1) with $u = v$ and a $p_v(\xi)$ satisfying (4.5) with $y = y_v(\xi)$. Thus by (4.6) followed by (4.2),

$$\begin{aligned} \|G(v, \xi)\|_{L^2(D)} &= \|\lambda_2 v - p_v(\xi)\|_{L^2(D)} \leq \lambda_2 \|v\|_{L^2(D)} + C_1(\xi) \|y_D - y_v(\xi)\|_{L^2(D)} \\ &\leq \lambda_2 \|v\|_{L^2(D)} + C_1(\xi) \|y_D\|_{L^2(D)} + (C_1(\xi))^2 \|v\|_{L^2(D)} =: C(\xi). \end{aligned}$$

Notice that $C \in L^p(\Xi)$ for all $p \in [1, \infty)$ by nature of the mapping $\xi \mapsto C_1(\xi)$. Therefore, the conditions of Lemma C.3 are satisfied and we have proven Fréchet differentiability of j . □

C Differentiability of expectation functionals

Let $(X, \|\cdot\|_X)$ be a Banach space and let $J : X \times \Omega \rightarrow \mathbb{R}$ be a random variable functional. We summarize under what conditions we can exchange the integral and the derivative for the functional $j : X \rightarrow \mathbb{R}$, where $j(u) = \int_{\Omega} J(u, \omega) d\mathbb{P}(\omega)$.

The following definition gives the minimal requirement for exchanging the derivative and expectation, namely, requiring $J : X \times \Omega \rightarrow \mathbb{R}$ to be L^1 -Fréchet differentiable.

Definition C.1 A p -times integrable random functional $J : X \times \Omega \rightarrow \mathbb{R}$ is called L^p -Fréchet differentiable at u if for an open set $U \subset X$ there exists a bounded and linear random operator $A : U \times \Omega \rightarrow \mathbb{R}$ such that $\lim_{h \rightarrow 0} \|J_{\omega}(u + h) - J_{\omega}(u) + A(u, \omega)h\|_{L^p(\Omega)} / \|h\|_X = 0$.

By Hölder’s inequality, if $u \mapsto J(u, \cdot)$ is L^p -differentiable and $1 \leq r < p$, then it is also L^r -differentiable with the same derivative. This implies that $j : X \rightarrow \mathbb{R}$ is Fréchet differentiable at u .

The condition in Definition C.1 might be difficult to verify directly. For this reason, we consider other assumptions on an open neighborhood U of X containing u . We denote the functional $J(\cdot, \omega) : X \rightarrow \mathbb{R}$ for a fixed realization $\omega \in \Omega$ by $J_{\omega} : X \rightarrow \mathbb{R}$.

Assumption C.2

- (vi) The expectation $j(v)$ is well-defined and finite-valued for all $v \in U$.
- (vii) For almost every $\omega \in \Omega$, the functional $J_{\omega} : X \rightarrow \mathbb{R}$ is Fréchet differentiable at u . Moreover, there exists a positive random variable $C(\cdot) \in L^1(\Omega)$ such that for all $v \in U$ and almost every $\omega \in \Omega$,

$$\|J'_{\omega}(v)\|_{X^*} \leq C(\omega). \tag{C.1}$$

Lemma C.3 Suppose Assumption C.2 holds. Then j is Fréchet differentiable at u and

$$j'(u) = \mathbb{E}[J'_{\omega}(u)]. \tag{C.2}$$

Proof By the mean value theorem, for h close enough to u , there exists a z within the neighborhood containing $u + h$ and u that satisfies $|J_\omega(u + h) - J_\omega(u)| \leq \|J'_\omega(z)\|_{X^*} \|h\|_X$. Now, we have for almost every $\omega \in \Omega$ that

$$\begin{aligned} \frac{|J_\omega(u + h) - J_\omega(u) - J'_\omega(u)h|}{\|h\|_X} &\leq \frac{|J_\omega(u + h) - J_\omega(u)|}{\|h\|_X} + \frac{|J'_\omega(u)h|}{\|h\|_X} \\ &\leq \|J'_\omega(z)\|_{X^*} + \|J'_\omega(u)\|_{X^*} \leq 2C(\omega). \end{aligned}$$

By Assumption C.2(vii), $C(\cdot)$ is integrable, so by Lebesgue's dominated convergence theorem, it follows that

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\int_\Omega |J_\omega(u + h) - J_\omega(u) - J'_\omega(u)h| \, d\mathbb{P}(\omega)}{\|h\|_X} \\ = \int_\Omega \lim_{h \rightarrow 0} \frac{|J_\omega(u + h) - J_\omega(u) - J'_\omega(u)h|}{\|h\|_X} \, d\mathbb{P}(\omega) = 0, \end{aligned} \quad (\text{C.3})$$

where the last equality follows by Assumption C.2(vii). Now consider the mapping $F : h \mapsto \int_\Omega J'_\omega(u)h \, d\mathbb{P}(\omega)$. It is straightforward to show that this is a bounded and linear operator. Therefore, we use Assumption C.2(vi) to get

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{|\int_\Omega J_\omega(u + h) \, d\mathbb{P}(\omega) - \int_\Omega J_\omega(u) \, d\mathbb{P}(\omega) - F(h)|}{\|h\|_X} \\ = \lim_{h \rightarrow 0} \frac{|\int_\Omega (J_\omega(u + h) - J_\omega(u) - J'_\omega(u)h) \, d\mathbb{P}(\omega)|}{\|h\|_X} = 0, \end{aligned}$$

where the second equality holds by the triangle inequality and and (C.3). Therefore j is Fréchet differentiable at u with derivative $F = \int_\Omega J'_\omega(u) \, d\mathbb{P}(\omega)$. \square

References

1. Ali, A., Ullmann, E., Hinze, M.: Multilevel Monte Carlo analysis for optimal control of elliptic PDEs with random coefficients. *SIAM/ASA J. Uncertain. Quantif.* **5**(1), 466–492 (2017)
2. Alnæs, M., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M., Wells, G.: The FEniCS project version 1.5. *Arch. of Numerical Software*, 3(100) (2015)
3. Barty, K., Roy, J., Strugarek, C.: Hilbert-valued perturbed subgradient algorithms. *Math. Oper. Res.* **32**(3), 551–562 (2007)
4. Bauschke, H., Combettes, P.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, vol. 408. Springer, New York (2011)
5. Beck, A.: *First-Order Methods in Optimization*, vol. 25. SIAM (2017)
6. Bottou, L.: Online learning and stochastic approximations. *On-line Learn. Neural Netw.* **17**(9), 142 (1998)
7. Bottou, L., Curtis, F., Nocedal, J.: Optimization methods for large-scale machine learning. *SIAM Rev.* **60**(2), 223–311 (2018)

8. Brézis, H.: Opérateurs Maximaux Monotones et Semi-groupes de Contractions dans les Espaces de Hilbert. North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York (1973)
9. Cazenave, T., Haraux, A.: An Introduction to Semilinear Evolution Equations, vol. 13. Oxford University Press, Oxford (1998)
10. Chen, X., White, H.: Asymptotic properties of some projection-based Robbins–Monro procedures in a Hilbert space. *Stud. Nonlinear Dyn. Econom.* **6**, 1–53 (2002)
11. Clarke, F.: Optimization and Nonsmooth Analysis, vol. 5. SIAM, Philadelphia (1990)
12. Culioli, J.-C., Cohen, G.: Decomposition/coordination algorithms in stochastic optimization. *SIAM J. Control Optim.* **28**(6), 1372–1403 (1990)
13. Davis, D., Drusvyatskiy, D., Kakade, S., Lee, J.: Stochastic subgradient method converges on tame functions. *Found. Comput. Math.* **20**, 1–36 (2018)
14. De Los Reyes, J.: Numerical PDE-Constrained Optimization. Springer, New York (2015)
15. Duchi, J., Ruan, F.: Stochastic methods for composite and weakly convex optimization problems. *SIAM J. Optim.* **28**(4), 3229–3259 (2018)
16. Duflo, M.: Random Iterative Models, vol. 34. Springer, Berlin (2013)
17. Ermoliev, Y.: On the stochastic quasi-gradient method and stochastic quasi-Feyer sequences. *Kibernetika* **2**, 72–83 (1969)
18. Evans, L.: Partial Differential Equations, volume Graduate Studies in Mathematics vol. 19. American Mathematical Society, Providence, R.I. (1998)
19. Geiersbach, C., Pflug, G.: Projected stochastic gradients for convex constrained problems in Hilbert spaces. *SIAM J. Optim.* **29**(3), 2079–2099 (2019)
20. Geiersbach, C., Wollner, W.: A stochastic gradient method with mesh refinement for PDE-constrained optimization under uncertainty. *SIAM/ASA J. Sci. Comput.* **42**(5), A2750–A2772 (2020)
21. Ghadimi, S., Lan, G., Zhang, H.: Mini-batch stochastic approximation methods for nonconvex stochastic composite optimization. *Math. Program.* **155**(1–2), 267–305 (2016)
22. Goldstein, L.: Minimizing noisy functionals in Hilbert space: an extension of the Kiefer–Wolfowitz procedure. *J. Theor. Probab.* **1**(2), 189–204 (1988)
23. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE Constraints. Springer, Netherlands (2009)
24. Keshavarzadeh, V., Fernandez, F., Tortorelli, D.: Topology optimization under uncertainty via non-intrusive polynomial chaos expansion. *Comput. Methods Appl. Mech.* **318**, 120–147 (2017)
25. Kiefer, J., Wolfowitz, J.: Stochastic estimation of the maximum of a regression function. *Ann. Math. Stat.* **23**(3), 462–466 (1952)
26. Kouri, D., Surowiec, T.: Risk-averse PDE-constrained optimization using the conditional value-at-risk. *SIAM J. Optim.* **26**(1), 365–396 (2016)
27. Kouri, D., Surowiec, T.: Risk-averse optimal control of semilinear elliptic PDEs. *ESAIM Control Optim. Calc. Var.* **26**, 53 (2019)
28. Kouri, D., Heinkenschloss, M., Ridzal, D., Van Bloemen Waanders, B.: A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty. *SIAM J. Sci. Comput.* **35**(4), A1847–A1879 (2013)
29. Kouri, D.: A multilevel stochastic collocation algorithm for optimization of PDEs with uncertain coefficients. *SIAM/ASA J. Uncertain. Quantif.* **2**(1), 55–81 (2014)
30. Kunoth, A., Schwab, C.: Sparse adaptive tensor Galerkin approximations of stochastic PDE-constrained control problems. *SIAM/ASA J. Uncertain. Quantif.* **4**(1), 1034–1059 (2016)
31. Kupka, I.: Counterexample to the Morse–Sard theorem in the case of infinite-dimensional manifolds. *Proc. Am. Math. Soc.* **16**(5), 954–957 (1965)
32. Kushner, H., Yin, G.: Stochastic Approximation and Recursive Algorithms and Applications. Springer, New York (2003)
33. Kushner, H., Clark, D.: Stochastic Approximation Methods for Constrained and Unconstrained Systems. Springer, New York (1978)
34. Lee, H.-C., Lee, J.: A stochastic Galerkin method for stochastic control problems. *Commun. Comput. Phys.* **14**(1), 77–106 (2013)
35. Lei, J., Shanbhag, U.: A randomized block proximal variable sample-size stochastic gradient method for composite nonconvex stochastic optimization. arXiv preprint [arXiv:1808.02543](https://arxiv.org/abs/1808.02543) (2018)
36. Ljung, L.: Analysis of recursive stochastic algorithms. *IEEE Trans. Autom. Control* **22**(4), 551–575 (1977)

37. Martin, M., Krumscheid, S., Nobile, F.: Analysis of stochastic gradient methods for PDE-constrained optimal control problems with uncertain parameters. Technical report, École Polytechnique MATHICSE Institute of Mathematics (2018)
38. Métivier, M.: Semimartingales: A Course on Stochastic Processes, vol. 2. Walter de Gruyter, Berlin (2011)
39. Nemirovski, A., Juditsky, A., Lan, G., Shapiro, A.: Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* **19**(4), 1574–1609 (2009)
40. Nixdorf, R.: An invariance principle for a finite dimensional stochastic approximation method in a Hilbert space. *J. Multivar. Anal.* **15**, 252–260 (1984)
41. Nouy, A., Pled, F.: A multiscale method for semi-linear elliptic equations with localized uncertainties and non-linearities. *ESAIM: Math. Model. Numer.* **52**(5), 1763–1802 (2018)
42. Okada, N.: On the Banach–Saks property. *Proc. Jpn. Acad. Ser. A Math. Sci.* **60**(7), 246–248 (1984)
43. Pisier, G.: Martingales in Banach Spaces, vol. 155. Cambridge University Press, Cambridge (2016)
44. Reddi, S., Sra, S., Póczos, B., Smola, A.: Proximal stochastic methods for nonsmooth nonconvex finite-sum optimization. *Adv. Neural Inf. Process. Syst.* **29**, 1145–1153 (2016)
45. Robbins, H., Monro, S.: A stochastic approximation method. *Ann. Math. Stat.* **22**(3), 400–407 (1951)
46. Robbins, H., Siegmund, D.: A convergence theorem for non negative almost supermartingales and some applications. *Optimizing Methods in Statistics*, pp. 233–257. Academic Press, Cambridge (1971)
47. Rosseel, E., Wells, G.: Optimal control with stochastic PDE constraints and uncertain controls. *Comput. Methods Appl. Mech. Eng.* **213**, 152–167 (2012)
48. Ruszczyński, A., Syski, W.: Stochastic approximation method with gradient averaging for unconstrained problems. *IEEE Trans. Autom. Control* **28**(12), 1097–1105 (1983)
49. Shapiro, A., Wardi, Y.: Convergence analysis of stochastic algorithms. *Math. Oper. Res.* **21**(3), 615–628 (1996)
50. Smale, S.: An infinite dimensional version of Sard’s theorem. *The Collected Papers of Stephen Smale*, vol. 2, pp. 529–534. World Scientific, Singapore (2000)
51. Tiesler, H., Kirby, R., Xiu, D., Preusser, T.: Stochastic collocation for optimal control problems with stochastic PDE constraints. *SIAM J. Control Optim.* **50**(5), 2659–2682 (2012)
52. Tröltzsch, F.: *Optimale Steuerung partieller Differentialgleichungen*. Vieweg + Teubner, 2nd edition (2009)
53. Uryasev, S.P.: A stochastic quasigradient algorithm with variable metric. *Ann. Oper. Res.* **39**(1), 251–267 (1992)
54. Van Barel, A., Vandewalle, S.: Robust optimization of PDE constrained systems using a multilevel Monte Carlo method. *SIAM/ASA J. Uncertain. Quantif.* **7**(1), 174–202 (2019)
55. Venter, J.H.: On Dvoretzky stochastic approximation theorems. *Ann. Math. Stat.* **37**, 1534–1544 (1966)
56. Wardi, Y.: A stochastic algorithm using one sample point per iteration and diminishing stepsizes. *J. Optim. Theory Appl.* **61**(3), 473–485 (1989)
57. Williams, D.: *Probability with Martingales*. Cambridge University Press, Cambridge (1991)
58. Yin, G., Zhu, Y.M.: On H -valued Robbins–Monro processes. *J. Multivar. Anal.* **34**, 116–140 (1990)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.