Mathematical Programming (2022) 196:1065–1096
https://doi.org/10.1007/s10107-020-01593-z

FULL LENGTH PAPER

Series B

Dynamic probabilistic constraints under continuous random distributions

T. González Grandón[1] · R. Henrion[2] · P. Pérez-Aros[3]

Received: 3 March 2020 / Accepted: 3 November 2020 / Published online: 13 November 2020

Abstract
The paper investigates analytical properties of dynamic probabilistic constraints (chance constraints). The underlying random distribution is supposed to be continuous. In the first part, a general multistage model with decision rules depending on past observations of the random process is analyzed. Basic properties like (weak sequential) (semi-) continuity of the probability function or existence of solutions are studied. It turns out that the results differ significantly according to whether decision rules are embedded into Lebesgue or Sobolev spaces. In the second part, the simplest meaningful two-stage model with decision rules from L^2 is investigated. More specific properties like Lipschitz continuity and differentiability of the probability function are considered. Explicitly verifiable conditions for these properties are provided along with explicit gradient formulae in the Gaussian case. The application of such formulae in the context of necessary optimality conditions is discussed and a concrete identification of solutions presented.

Keywords Dynamic probabilistic constraints · Chance constraints · Continuous distributions · Decision rules · Stochastic programming

Mathematics Subject Classification 90C15 · 49K45

1 Introduction

1.1 Overview

The application of probabilistic constraints (or: chance constraints) to engineering problems and their numerical solution is nowadays standard. Introduced by Charnes

This work is dedicated to the memory of Shabbir Ahmed.

✉ R. Henrion
henrion@wias-berlin.de

Extended author information available on the last page of the article

et al. [5] in a simple form (individual constrains) in 1958, their systematic theoretical and algorithmic investigation has been pioneered by Prékopa and his students starting in the Seventies (see [15] and references therein). The typical form of a probabilistic constraint is the inequality

$$\mathbb{P}(g_i(x, \xi) \leq 0 \quad (i = 1, \ldots, p)) \geq p, \tag{1}$$

where x is a decision vector, ξ is a random vector, $\mathbb{P}$ a probability measure and g a random constraint mapping with finitely many components. The meaning of (1) is to define a decision x as feasible if the random inequality system $g(x, \cdot) \leq 0$ is satisfied at least with probability $p \in (0, 1]$. A modern theoretical treatment of probabilistic constraints can be found in the monograph [16, chapter 4]. The algorithmic solution of optimization problems subject to constraints (1) has been tremendously advanced within the last twenty years. Rather than providing a detailed list of references here, we want to emphasize the contribution to this development by Shabbir Ahmed (e.g., [12,13]). At the same time the traditional model (1) has been extended to broader settings such as PDE constrained optimization ([6,7,9]) or infinite random inequality systems (probust constraints, [17]).

A challenge of different nature consists in considering dynamic aspects in probabilistic constraints. Observe that (1) is a static model by nature: The decision x ('here-and now-decision') has to be taken before the randomness ξ is observed. Such model would apply, for instance, in the design of a mechanical construction (encoded by x) which is done once and for ever and has to resist unknown future random forces ξ with high probability. Many decisions, however, are time dependent. The components of x and ξ could refer to discrete time decision and random processes, respectively. In the control of a hydro reservoir, for instance, one is faced with an alternating sequence of decisions x_t (referring to water release) and realizations of randomness ξ_t (water inflow) according to the chronology

$$x_1 \curvearrowright \xi_1 \curvearrowright x_2 \curvearrowright \xi_2 \curvearrowright \cdots \curvearrowright x_T (\curvearrowright \xi_T). \tag{2}$$

Whether or not this sequence ends with a decision (final recourse action) or with the observation of randomness without the possibility of finally reacting to it, depends on the choice of a model of multistage stochastic optimization or of multistage probabilistic programming. This distinction requires some care because sometimes the term 'two-stage probabilistic constraint' is used for the addition of a probabilistic constraint (relaxing the almost sure existence of a recourse action) in a setting of two-stage stochastic programming. Such model has been first considered in [14] and is still of much interest (e.g., [11]). Here, the chronology is the one of (2) with $T = 2$ (without the final term in parantheses): $x_1 \curvearrowright \xi_1 \curvearrowright x_2$, i.e., it is a special two-stage stochastic optimization problem. In our understanding, it is not a two-stage probabilistic constraint which would end with the term in parentheses in (2): $x_1 \curvearrowright \xi_1 \curvearrowright x_2 \curvearrowright \xi_2$. In this way one would obtain a logical generalization of conventional one-stage (static) probabilistic constraints of type $x_1 \curvearrowright \xi_1$ and keep the idea, that in a probabilistic constraint one is always faced with a final unknown realization of some random vector. This idea follows a remark in [8]: '... a well-formed probabilistic constraint contains at

least one coefficient that depends on a random variable realized after the last decision is taken'.

It is clear that in (2) the dynamic character of the decision making process expresses itself by assuming all decisions being functions of past observations in order to take advantage of the gain of information obtained from the realizations of the random vector. Hence, instead of static (constant) decisions x_t one admits decision rules or policies $x_2(\xi_1)$, $x_3(\xi_1, \xi_2)$ etc. When considering continuously distributed random vectors, this approach takes the problem to infinite dimensions even though time is discrete, because policies are elements of appropriate function spaces. One may circumvent this difficulty by restricting policies to a parameterized class, linear decision rules in the simplest case. Then, one gets back to a static problem where decisions are the parameters of the policies. Several aspects of modeling linear decision rules in the context of (linear) multistage probabilistic constraints are discussed in [10]. It is not guaranteed, however, that the chosen class contains the optimal policy. Another idea to reduce the problem again to a finite-dimensional one would consist in a discrete approximation of the random distribution. A conceptual framework for dealing with dynamic probabilistic constraints without restricting the class of policies and keeping the continuous character of the given (multivariate Gaussian) distribution was presented in [4] along with applications to two- and three-stage probabilistic control of a water reservoir. Using stochastic dynamic programming rather than direct nonlinear programming, a similar problem was later analyzed and numerically solved in [2] for a significantly larger number of stages, however with a discrete random distribution.

The focus in this paper is not on the numerical solution of problems subject to dynamic probabilistic constraints but rather on analytical properties of the arising probability function. Here we assume the underlying random distribution to be continuous and keep the decision rules general as elements of some Lebesgue or Sobolev space. In Sect. 2, a general multistage model is analyzed. Basic properties like (weak sequential) (semi-) continuity of the probability function or existence of solutions are studied. In Sect. 3, the simplest meaningful two-stage model with decision rules from L^2 is investigated. More specific properties like Lipschitz continuity and differentiability of the probability function are considered. Explicitly verifiable conditions for these properties are provided along with explicit gradient formulae in the Gaussian case. The application of such formulae in the context of necessary optimality conditions is discussed and a concrete identification of solutions presented.

1.2 The general setting

In this paper we study optimization problems of the type

$$\min_{x\in\mathbb{X}}\{J(x) \mid x \in C,\ \varphi(x) \geq p\} \tag{3}$$

Here, the space of decisions $\mathbb{X}$ is one of the following Lebesgue or Sobolev spaces with $q \in [1, \infty)$

$$\mathcal{X} := \mathbb{R} \times L^q(\mathbb{R}) \times L^q(\mathbb{R}^2) \times \cdots \times L^q(\mathbb{R}^{T-1})$$

$$\mathcal{X}^1 := \mathbb{R} \times W^{1,q}(\mathbb{R}) \times W^{1,q}(\mathbb{R}^2) \times \cdots \times W^{1,q}(\mathbb{R}^{T-1}).$$

The subset $C \subseteq \mathcal{X}$ (or $C \subseteq \mathcal{X}^1$) is meant to represent some abstract constraint on the decision, e.g., nonnegativity or bounds for the components. The focus of our attention will be on the inequality constraint $\varphi(x) \geq p$ which we will assume to represent a so called *joint dynamic chance constraint*. More precisely, $p \in (0, 1]$ is some given safety level and $\varphi : \mathcal{X} \to [0, 1]$ denotes a probability function defined for $x \in \mathcal{X}$ as follows

$$\begin{aligned} \varphi(x) := \mathbb{P}\left(h_i\left(x_1, x_2\left(\xi_1\right), \ldots, x_T\left(\xi_1, \ldots, \xi_{T-1}\right), \xi_1, \ldots, \xi_T\right)\right. \\ \left. \leq 0 \quad i = 1, \ldots, k\right), \end{aligned} \tag{4}$$

where $h_i : \mathbb{R}^T \times \mathbb{R}^T \to \mathbb{R}$ and $\xi := (\xi_1, \ldots, \xi_T)$ is a T-dimensional discrete time process on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Observe that with each component x_t of the decision x depending on past outcomes $(\xi_1, \ldots, \xi_{t-1})$ only, x represents an adapted decision process. We endow $\mathcal{X}$ and $\mathcal{X}^1$ with the maximum norm with respect to the usual norms in the coordinate spaces. Doing so, $\mathcal{X}$ and $\mathcal{X}^1$ are Banach spaces.

1.3 A motivating example

To illustrate applications for problem 3, we present a decision management optimization problem on a single water reservoir for hydroelectricity generation. Given a set of future time intervals $1, 2, \ldots, T$, the problem of the operator is to decide on an optimal release policy $(x_1, \ldots, x_T)$ of water, considering technical, economical and environmental aspects. By $\xi = (\xi_1, \ldots, \xi_T)$, we denote the random vector indicating the stochastic water inflow (e.g. precipitation, snow melt) to the reservoirs at corresponding time intervals . The main role of the reservoir is to generate electricity. At the same time, lower and upper limits l_*, l^* for the water level have to be satisfied in the reservoir, say for flood protection or for ecological reasons. By the random nature of the inflows, the time dependent water level $l_t(x, \xi)$ induced from the controlled water release x is a random variable too. Hence, the mentioned limits cannot be satisfied in a deterministic way. Rather, it is reasonable to impose them in a probabilistic way:

$$\mathbb{P}\left(l_* \leq l_t(x, \xi) \leq l^* \quad (t = 1, \ldots, T)\right) \geq p. \tag{5}$$

Here, $p \in [0, 1]$ denotes a probability level at which the random constraints are supposed to hold true. The current water level after time interval t is clearly given as the initial level plus the cumulated inflow minus the cumulated release so far:

$$l_t(x, \xi) = l_0 + \xi_1 + \cdots + \xi_t - x_1 - \cdots - x_t \quad (t = 1, \ldots, T).$$

Sometimes, one decides on the future water release in complete ignorance of future water inflow. This is the case, for instance, in day ahead markets, when energy production (water release) for each hour of the next day is fixed one day ahead. Then, decisions are just scalars for each time intervals and the probabilistic constraint (5)

becomes

$$\mathbb{P}\left(l_* \le l_0 + \sum_{\tau=1}^{t} \xi_\tau - \sum_{\tau=1}^{t} x_\tau \le l^* \quad (t = 1, \ldots, T)\right) \ge p \quad \left(x \in \mathbb{R}^T\right). \tag{6}$$

Such a static model does not take into account the temporal gain of information while the random inflow process unfolds. In longer term planning problems one therefore admits from the beginning that future decisions on water release are functions of past observations of the random inflow. Hence, rather than deciding on scalars $x_1, \ldots, x_T$, one is looking for functions $x_1, x_2(\cdot), x_3(\cdot, \cdot)$, so-called *policies*. In this dynamic setting better solutions of the underlying optimization problem can be expected (the static model being included as a special case with constant policies, e.g., $x_2(\cdot) \equiv x_2$ etc.). Hence, we adjust our static chance constraint above to a dynamic one, where $\left(x \in \mathcal{X}, \mathcal{X}^1\right)$:

$$\mathbb{P}\left(l_* \le l_0 + \sum_{\tau=1}^{t} \xi_\tau - \sum_{\tau=1}^{t} x_\tau\left(\xi_1, \ldots, \xi_{\tau-1}\right) \le l^* \quad (t = 1, \ldots, T)\right) \ge p.$$

A possible objective in a corresponding optimization problem might consist in the maximization of the expected overall water release (representing the amount of energy produced):

$$J(x) := -\mathbb{E} \sum_{t=1}^{T} x_t\left(\xi_1, \ldots, \xi_{\tau-1}\right).$$

Then, the optimization problem is of the form (3) with the probability function φ defined in (4) via the constraint mapping $h : \mathbb{R}^T \times \mathbb{R}^T \to \mathbb{R}^{2T}$. The latter has $k := 2T$ components

$$h_t(u, v) := l_0 - l^* + \sum_{\tau=1}^{t} v_\tau - \sum_{\tau=1}^{t} u_\tau \quad (t = 1, \ldots, T)$$

$$h_{T+t}(u, v) := l_* - l_0 + \sum_{\tau=1}^{t} u_\tau - \sum_{\tau=1}^{t} v_\tau \quad (t = 1, \ldots, T).$$

2 Basic structural properties of the general model

In this section we are going to collect some basic structural properties of the chance constraint $\varphi(x) \ge p$ in (3) and the involved probability function φ in (4). For convenience, we introduce the notation $u_{[i]} := (u_1, \ldots, u_i)$ for vectors $u \in \mathbb{R}^n$ and $1 \le i \le n$. With the policy $x \in \mathcal{X}$ we associate the joint policy (whose components have a common domain) as a mapping $[x] : \mathbb{R}^T \to \mathbb{R}^T$ defined by

$$[x](z) := \left(x_t(z_{[t-1]})\right)_{t=1,\dots,T} \quad \left(z \in \mathbb{R}^T\right), \tag{7}$$

with the convention $x_1(z_{[0]}) = x_1$. Finally, we introduce the maximum function related to the mapping h:

$$h^{\max} := \max_{i=1,\dots,k} h_i. \tag{8}$$

Then, the probability function in (4) can be compactly written as

$$\varphi(x) = \mathbb{P}\left(h^{\max}([x](\xi), \xi) \leq 0\right). \tag{9}$$

We first check, that this expression is well-defined. In order to ensure this, we make the following basic assumptions in (4) throughout this paper:

$$\left.\begin{array}{l} \xi \text{ possesses a density} \\ h \text{ is Borel measurable} \end{array}\right\}. \tag{BA}$$

Observe first that for given $x \in \mathcal{X}$ each component $x_t : \mathbb{R}^{t-1} \to \mathbb{R}$ is Borel measurable, whence the mapping $[x]$ is Borel measurable. Then, $h^{\max}([x](z), z)$ is a Borel measurable function of z because each h_i is so thanks to (BA). This implies that

$$\left\{\omega \in \Omega \,|\, h^{\max}([x](\xi(\omega)), \xi(\omega)) \leq 0\right\} \in \mathcal{A},$$

so that it is justified to speak of the probability of this event appearing in the definition of (9). It remains to show that this probability is independent of the representative of $x \in \mathcal{X}$. To see this, let $x^{(1)}, x^{(2)} \in \mathcal{X}$ such that $x_1^{(1)} = x_1^{(2)}$ and be such that

$$x_{t+1}^{(1)}(u) = x_{t+1}^{(2)}(u) \quad \forall u \in B_t \ \forall t = 1, \dots, T-1,$$

where $B_t \subseteq \mathbb{R}^t$ are Lebesgue measurable subsets with $\lambda_t\left(\mathbb{R}^t \backslash B_t\right) = 0$ (λ_t is the Lebesgue measure in $\mathbb{R}^t$). Define

$$C := \bigcup_{t=1}^{T-1} C_t, \text{ where } C_t := \left(\mathbb{R}^t \backslash B_t\right) \times \mathbb{R}^{T-t-1} \subseteq \mathbb{R}^{T-1} \quad \forall t = 1, \dots, T-1.$$

Then, $\lambda_{T-1}(C) = 0$ and

$$\left[x^{(1)}\right](z) = \left[x^{(2)}\right](z) \quad \forall z \in \left(\mathbb{R}^{T-1} \backslash C\right) \times \mathbb{R}.$$

Since ξ possesses a density f_ξ, it follows from (9) that

$$\varphi(x^{(1)}) = \int\limits_{\{z \,|\, h^{\max}([x^{(1)}](z), z) \leq 0\}} f_\xi(z)\, dz$$

$$
\begin{aligned}
&= \int_{\{z|h^{\max}([x^{(1)}](z),z)\leq 0\}\cap\{(\mathbb{R}^{T-1}\setminus C)\times\mathbb{R}\}} f_\xi(z)\,dz \\
&= \int_{\{z|h^{\max}([x^{(2)}](z),z)\leq 0\}\cap\{(\mathbb{R}^{T-1}\setminus C)\times\mathbb{R}\}} f_\xi(z)\,dz \\
&= \int_{\{z|h^{\max}([x^{(2)}](z),z)\leq 0\}} f_\xi(z)\,dz = \varphi(x^{(2)}).
\end{aligned}
$$

This shows, that the value of φ does not depend on the representative of $x \in \mathcal{X}$.

We will commence our analysis with some (lower-) semicontinuity properties and then derive consequences later on. The following Proposition turns out to be a crucial technical tool in this context:

Proposition 1 *In addition to the basic assumptions (BA), suppose that h in (4) has components h_i which are lower semicontinuous in their first argument vector (related with x). Consider a sequence $x^{(n)}$ in $\mathcal{X}$ which converges componentwise almost everywhere to some $x \in \mathcal{X}$. Then,*

$$\limsup_{n\to\infty} \varphi\left(x^{(n)}\right) \leq \varphi(x). \tag{10}$$

Moreover, if h has components h_i which are upper semicontinuous in their first argument vector and in addition

$$\lambda_T\left(\left\{z \in \mathbb{R}^T | h_i([x](z), z) = 0\right\}\right) = 0 \quad i = 1, \ldots, k, \tag{11}$$

then

$$\liminf_{n\to\infty} \varphi\left(x^{(n)}\right) \geq \varphi(x). \tag{12}$$

Proof We start with the first assertion (10). The function $h^{\max}$ in (8) is lower semicontinuous in its first argument vector because the h_i are assumed to be so. By assumption, we have that $x_1^{(n)} \to_n x_1$ and

$$x_{t+1}^{(n)}(u) \to_n x_{t+1}(u) \quad \forall u \in B_t \ \forall t = 1, \ldots, T-1,$$

for some Lebesgue measurable subsets $B_t \subseteq \mathbb{R}^t$ with $\lambda_t\left(\mathbb{R}^t \setminus B_t\right) = 0$. Without loss of generality (by passing to a superset whose difference with B_t has Lebesgue measure zero), we may assume that the B_t are Borel measurable. Repeating the construction from the beginning of this section, we find a subset $C \subseteq \mathbb{R}^{T-1}$ which now is Borel measurable and is such that $\lambda_{T-1}(C) = 0$ and

$$\left[x^{(n)}\right](z) \to_n [x](z) \quad \forall z \in \left(\mathbb{R}^{T-1}\setminus C\right) \times \mathbb{R}.$$

Denote $\Gamma := \xi^{-1}\left(\left(\mathbb{R}^{T-1}\backslash C\right) \times \mathbb{R}\right) \in \mathcal{A}$ and observe that

$$\mathbb{P}(\Gamma) = \int\limits_{\left(\mathbb{R}^{T-1}\backslash C\right)\times\mathbb{R}} f_{\xi}(z)\, dz = 1, \quad \left[x^{(n)}\right](\xi(\omega)) \to_n [x](\xi(\omega)) \quad \forall \omega \in \Gamma.$$

Consider the event sets

$$A_n := \left\{\omega \in \Omega \,|\, h^{\max}\left(\left[x^{(n)}\right](\xi(\omega)), \xi(\omega)\right) \leq 0\right\} \quad (n \in \mathbb{N})$$
$$A := \left\{\omega \in \Omega \,|\, h^{\max}([x](\xi(\omega)), \xi(\omega)) \leq 0\right\}.$$

Fix an arbitrary $\omega \in (\Omega\backslash A) \cap \Gamma$. Then, the lower semicontinuity of $h^{\max}$ in its first argument vector yields that

$$\liminf_{n\to\infty} h^{\max}\left(\left[x^{(n)}\right](\xi(\omega)), \xi(\omega)\right) \geq h^{\max}([x](\xi(\omega)), \xi(\omega)) > 0.$$

Consequently, for any $\omega \in (\Omega\backslash A) \cap \Gamma$, there exists some $n_0(\omega) \in \mathbb{N}$ such that

$$h^{\max}\left(\left[x^{(n)}\right](\xi(\omega)), \xi(\omega)\right) > 0 \quad \forall n \geq n_0(\omega) \tag{13}$$

Denote by χ_Q the characteristic function of a set Q. Now, (13) entails that $\chi_{A_n}(\omega) \to_n 0$ for all $\omega \in (\Omega\backslash A) \cap \Gamma$. In other words, since $\mathbb{P}(\Gamma) = 1$, χ_{A_n} converges pointwise $\mathbb{P}$ -almost surely to χ_A on the set $\Omega\backslash A$. Since $\chi_{A_n} \leq 1$, the dominated convergence theorem provides that

$$\int\limits_{\Omega\backslash A} \chi_{A_n} d\mathbb{P} \to_n 0.$$

Now, let $x^{(n_l)}$ be a subsequence realizing the limsup in (10) as a limit. Then, in view of the relation above, we arrive at (10):

$$\begin{aligned}
\limsup_{n\to\infty} \varphi\left(x^{(n)}\right) &= \lim_{l\to\infty} \varphi\left(x^{(n_l)}\right) = \lim_{l\to\infty} \mathbb{P}\left(h^{\max}\left(\left[x^{(n_l)}\right](\xi), \xi\right) \leq 0\right) \\
&= \lim_{l\to\infty} \mathbb{P}\left(A_{n_l}\right) = \lim_{l\to\infty} \int\limits_{\Omega} \chi_{A_{n_l}} d\mathbb{P} \\
&\leq \limsup_{l\to\infty} \int\limits_{\Omega\backslash A} \chi_{A_{n_l}} d\mathbb{P} + \limsup_{l\to\infty} \int\limits_{A} \chi_{A_{n_l}} d\mathbb{P} \\
&= \limsup_{l\to\infty} \int\limits_{A} \chi_{A_{n_l}} d\mathbb{P} \leq \limsup_{l\to\infty} \int\limits_{A} d\mathbb{P} = \mathbb{P}(A) \\
&= \mathbb{P}\left(h^{\max}([x](\xi), \xi) \leq 0\right) = \varphi(x).
\end{aligned}$$

As for (12), observe first that with the components h_i being upper semicontinuous in their first argument vector, the components $(-h_i)$ of $-h$ are lower semicontinuous

in their first argument vector. Denote by $\tilde{\varphi}$ the probability function in (4) or (9), respectively, associated with $-h$ rather than with h. Then, by the just proven relation (10), we have that

$$\limsup_{n\to\infty} \mathbb{P}\left(-h^{\max}\left(\left[x^{(n)}\right](\xi),\xi\right)\le 0\right) = \limsup_{n\to\infty} \tilde{\varphi}\left(x^{(n)}\right) \le \tilde{\varphi}(x) = \mathbb{P}\left(-h^{\max}([x](\xi),\xi)\le 0\right).$$

It now follows that

$$\begin{aligned}
\liminf_{n\to\infty} \varphi\left(x^{(n)}\right) &= \liminf_{n\to\infty} \mathbb{P}\left(h^{\max}\left(\left[x^{(n)}\right](\xi),\xi\right)\le 0\right)\\
&\ge \liminf_{n\to\infty} \mathbb{P}\left(-h^{\max}\left(\left[x^{(n)}\right](\xi),\xi\right)> 0\right)\\
&= -\limsup_{n\to\infty} -\mathbb{P}\left(-h^{\max}\left(\left[x^{(n)}\right](\xi),\xi\right)> 0\right)\\
&= -\limsup_{n\to\infty} \left(\mathbb{P}\left(-h^{\max}\left(\left[x^{(n)}\right](\xi),\xi\right)\le 0\right)-1\right)\\
&= 1-\limsup_{n\to\infty} \mathbb{P}\left(-h^{\max}\left(\left[x^{(n)}\right](\xi),\xi\right)\le 0\right)\\
&\ge 1-\mathbb{P}\left(-h^{\max}([x](\xi),\xi)\le 0\right) = \mathbb{P}\left(h^{\max}([x](\xi),\xi)< 0\right).
\end{aligned}$$

From (11) and the basic assumption (BA) that ξ possesses a density, we infer that

$$\begin{aligned}
\mathbb{P}\left(h^{\max}([x](\xi),\xi)=0\right) &= \lambda_T\left(\left\{z\in\mathbb{R}^T \,|\, h^{\max}([x](z),z)=0\right\}\right)\\
&\le \sum_{i=1}^{k} \lambda_T\left(\left\{z\in\mathbb{R}^T \,|\, h_i([x](z),z)=0\right\}\right)=0.
\end{aligned}$$

Hence, we may continue the previous chain of (in-)equalities, in order to arrive at (12):

$$\liminf_{n\to\infty} \varphi\left(x^{(n)}\right) \ge \mathbb{P}\left(h^{\max}([x](\xi),\xi)\le 0\right) = \varphi(x).$$

□

The following Lemma will allow us to derive from Proposition 1 the announced (semi-) continuity properties for φ. We do not claim that this Lemma is new but are not able to provide a reference.

Lemma 1 *Consider a sequence $\left\{x^{(n)}\right\} \subseteq \mathcal{X}^1$ which converges weakly to $x \in \mathcal{X}^1$. Then, there exists a subsequence $\left\{x^{(n_k)}\right\}$ which converges almost everywhere to x.*

Proof Consider $\{x^{(n)}\} \subseteq \mathcal{X}^1$ which converges weakly to $x \in \mathcal{X}^1$. Since our space $\mathcal{X}^1$ is a product spaces, it is enough to prove that each coordinates has a subsequence with the desired property.

Let us fix $i \in \{2, \ldots, T\}$ (the case $i = 1$ is trivial). For simplicity of notation let us denote $f_n := x_i^{(n)}$, $f := x^{(n)}$. Since f_n converges weakly to f we have that f_n is bounded in $W^{1,q}(\mathbb{R}^{i-1})$.

Consider $r \in \mathbb{N}\backslash\{0\}$, define the domain $U_r := \mathbb{B}_r \subseteq \mathbb{R}^{i-1}$, the Euclidean ball centered at zero with radius r. We have that the restriction of f_n and f belongs to $W^{1,1}(U_r)$, and since U_r is bounded we have that f_n and f belong to $W^{1,q}(U_r)$. Now, by Rellich–Kondrachov's Theorem (see, e.g., [1, Theorem 6.3, Part I] and [1, p. 84]) we can extract a subsequence f_{n_k} which converges in norm and almost everywhere to $z \in L^1(U_r)$. Moreover, since f_{n_k} also converges weakly to f we have that $z = f$ almost everywhere on U_r. Finally, using induction and a diagonal argument we are done. □

Theorem 1 *In addition to the basic assumptions (BA), suppose that h in (4) has components h_i which are lower semicontinuous in their first argument vector (related with x). Then, $\varphi : \mathcal{X} \to [0, 1]$ defined in (4) is upper semicontinuous in the norm topology of $\mathcal{X}$. Its restriction $\varphi|_{\mathcal{X}^1} : \mathcal{X}^1 \to [0, 1]$ is sequentially upper semicontinuous with respect to the weak topology of $\mathcal{X}^1$. If, h in (4) has components h_i which are upper semicontinuous in their first argument vector and condition (11) is satisfied, then $\varphi : \mathcal{X} \to [0, 1]$ is lower semicontinuous in the norm topology of $\mathcal{X}$ and its restriction $\varphi|_{\mathcal{X}^1} : \mathcal{X}^1 \to [0, 1]$ is sequentially lower semicontinuous with respect to the weak topology of $\mathcal{X}^1$.*

Proof Let $\left\{x^{(n)}\right\} \subseteq \mathcal{X}$ be a sequence strongly converging to some $x \in \mathcal{X}$. Consider a subsequence $\left\{x^{(n_k)}\right\}$ such that

$$\limsup_{n\to\infty} \varphi\left(x^{(n)}\right) = \lim_{k\to\infty} \varphi\left(x^{(n_k)}\right).$$

It is well known that there exists a further subsequence $\left\{x^{(n_{k_l})}\right\}$ converging almost everywhere to x (see, e.g., [3, Theorem 13.6]). Then, by (10),

$$\limsup_{n\to\infty} \varphi\left(x^{(n)}\right) = \lim_{l\to\infty} \varphi\left(x^{(n_{k_l})}\right) \leq \varphi(x) \tag{14}$$

which shows the upper semicontinuity of φ in the norm topology of $\mathcal{X}$.

Next, let $\left\{x^{(n)}\right\} \subseteq \mathcal{X}^1$ be a sequence weakly converging to some $x \in \mathcal{X}^1$. Then, repeating the previous argument—this time justifying almost everywhere convergence of a subsequence on the basis of Lemma 1—we derive in the same way inequality (14), thus proving the sequential upper semicontinuity of $\varphi|_{\mathcal{X}^1}$ with respect to the weak topology of $\mathcal{X}^1$.

Under the additional assumption (11), the same argumentation as above can be repeated along with (12), in order to derive the remaining assertions. □

Corollary 1 *Denote by*

$$M(p) := \{x \in \mathcal{X} | \varphi(x) \geq p\}; \quad M^1(p) := \left\{x \in \mathcal{X}^1 | \varphi(x) \geq p\right\}$$

the sets of feasible decisions in problem (3) *defined by the dynamic probabilistic constraint. In addition to the basic assumptions (BA), suppose that h in* (4) *has components h_i which are lower semicontinuous in their first argument vector (related with x). Then, $M(p)$ is strongly closed in $\mathcal{X}$ and $M^1(p)$ is weakly sequentially closed in $\mathcal{X}^1$.*

Corollary 2 *In addition to the basic assumptions (BA) and to condition* (11) *suppose that h in* (4) *has components h_i which are continuous in their first argument vector (related with x). Then, $\varphi : \mathcal{X} \to [0,1]$ defined in* (4) *is continuous in the norm topology of $\mathcal{X}$. Its restriction $\varphi|_{\mathcal{X}^1} : \mathcal{X}^1 \to [0,1]$ is sequentially continuous with respect to the weak topology of $\mathcal{X}^1$.*

We are now in a position to prove with standard arguments the existence of solutions to problem (3) related with the space $\mathcal{X}^1$ of decisions:

Theorem 2 *Consider the optimization problem* (3) *with $\mathcal{X}^1$ as the space of decisions. In addition to the basic assumptions (BA), we suppose that*

1. *The index q in the definition of the space $\mathcal{X}^1$ satisfies $1 < q < \infty$.*
2. *The abstract constraint set $C \subseteq \mathcal{X}^1$ is norm closed, bounded and convex.*
3. *The objective function J is weakly sequentially lower semicontinuos.*
4. *The mapping h in* (4) *has components h_i which are lower semicontinuous in their first argument vector (related with x).*
5. *The set of feasible decisions of problem* (3) *is nonempty.*

Then, (3) *admits a solution.*

Proof As a consequence of 1., $\mathcal{X}^1$ is a reflexive Banach space. Therefore, 2. implies that C is weakly sequentially compact. By 4. and Corollary 1, the set $M^1(p) = \left\{x \in \mathcal{X}^1 | \varphi(x) \geq p\right\}$ is weakly sequentially closed. Hence, with 5., the feasible set $C \cap M^1(p)$ of (3) is nonempty and weakly sequentially compact. Now, with 3., the Weierstrass Theorem guarantees the existence of a solution to (3). □

The following example illustrates that, under the assumptions of Corollary 1, $M(p)$ cannot be expected to be weakly sequentially closed in $\mathcal{X}$ (in contrast with $M^1(p)$ and $\mathcal{X}^1$) and therefore existence of solutions as in Theorem 2 cannot be expected in the space $\mathcal{X}$:

Example 1 Let $T = 2, k = 2, q = 2, p = 0.5 + (2\pi)^{-1}$ and let ξ have a uniform distribution on the rectangle $[0, 4\pi] \times [0,1]$. Define the mapping h by $h_i(x_1, x_2, z_1, z_2) = z_i - x_i$ for $i = 1, 2$. Then, the h_i are continuous such that our basic assumptions (BA) are satisfied and Corollary 1 guarantees that $M(p)$ is strongly closed in $\mathcal{X}$. Now, define the sequence

$$x^{(n)} := \left(x_1^{(n)}, x_2^{(n)}\right) \in \mathcal{X} = \mathbb{R} \times L^2(\mathbb{R}) \quad \text{by}$$

$$x_1^{(n)} := 4\pi; \quad x_2^{(n)}(t) := \begin{cases} 0 & t \in (-\infty, 0] \cup (4\pi, \infty) \\ \sin(nt) & t \in (0, 2\pi] \\ 1 & t \in (2\pi, 4\pi] \end{cases}.$$

Then, $x^{(n)}$ weakly converges to $x := \left(4\pi, \chi_{[2\pi,4\pi]}\right)$. Moreover, by definition of h and ξ and by (4), it holds that

$$\begin{aligned}\varphi\left(x^{(n)}\right) &= \mathbb{P}\left(\xi_1 \leq x_1^{(n)}, \xi_2 \leq x_2^{(n)}(\xi_1)\right)\\ &= \mathbb{P}\left(0 \leq \xi_1 < 2\pi, 0 \leq \xi_2 \leq \sin(n\xi_1)\right) + \mathbb{P}\left(2\pi \leq \xi_1 \leq 4\pi, 0 \leq \xi_2 \leq 1\right)\\ &= (2\pi)^{-1} + 0.5 = p.\end{aligned}$$

Therefore, $x^{(n)} \in M(p)$. On the other hand,

$$\begin{aligned}\varphi(x) &= \mathbb{P}\left(\xi_1 \leq 4\pi, \xi_2 \leq \chi_{[2\pi,4\pi]}(\xi_1)\right) = \mathbb{P}\left(2\pi \leq \xi_1 \leq 4\pi, 0 \leq \xi_2 \leq 1\right)\\ &= 0.5 < p.\end{aligned}$$

It follows that $x \notin M(p)$, whence $M(p)$ fails to be weakly sequentially closed.

We finish this Section by briefly addressing the issue of convexity of the feasible set defined by the probabilistic constraint $\varphi(x) \geq p$ in (3). Assume first that we would deal with a joint static probabilistic constraint, which means that the decision policies x are supposed to be constants: $x[z] \equiv x \in \mathbb{R}^T$ in (7). Assume further, that ξ has a logconcave density (e.g., multivariate Gaussian) and that the mapping h is affine linear: $h(x, z) = Ax + Bz + b$. This is the cae, for instance, for the reservoir problem with static probabilistic constraint (6). Then, thanks to a result by Prékopa [15, Th. 10.2.1.], the inequality $\varphi(x) \geq p$ defines a convex set of feasible decisions x for any right-hand side probability level p. Unfortunately, a similar convexity result gets lost in the dynamic setting. Indeed, we may revisit Example 1, where the density of the given uniform distribution is constant on the rectangle and zero outside, hence logconcave (in the extended-valued meaning). Moreover, the mapping $h(x, z) = z - x$ is linear. As for the feasible set $M(p) := \{x \in \mathcal{X} | \varphi(x) \geq p\}$, we have seen in Example 1 that it is strongly closed but fails to be weakly sequentially closed. If it was convex, then closedness would imply weak closedness, hence weak sequential closedness, which is a contradiction.

3 Properties of the probability function in a simple two-stage model

In this section, we are going to investigate analytical properties (continuity, Lipschitz continuity, differentiability including explicit derivatives) of the probability function φ in (4) in the framework of the simplest meaningful dynamic setting. More precisely, we consider a two-stage model ($T = 2$) of the following joint and separated probabilistic constraint:

$$\varphi(x) := \mathbb{P}\left(\xi_1 \leq x_1, \xi_2 \leq x_2(\xi_1)\right) \geq p. \tag{15}$$

This corresponds to the choice of the mapping $h : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ defined by $h(x, z) = z - x$ in (4). We will choose $\mathcal{X}$ with index $q = 2$ to be the base space of decisions, which means that $x_2 \in L^2(\mathbb{R})$. In all results hereafter, we shall explicitly work with a given density of ξ. By continuity of h, our basic assumptions (BA) will be automatically satisfied then.

3.1 Continuity and lipschitz continuity

Proposition 2 *If ξ has a density, then the probability function $\varphi : \mathbb{R} \times L^2(\mathbb{R}) \to \mathbb{R}$ is continuous.*

Proof Since h is continuous, it suffices by Corollary 2 to check condition (11) at an arbitrary $x \in \mathcal{X}$. For the first component h_1 of h it reads as

$$\lambda_2\left(\left\{z \in \mathbb{R}^2 | z_1 = x_1\right\}\right) = 0$$

which is evidently true. For the second component we observe that

$$\begin{aligned}&\lambda_2\left(\left\{z \in \mathbb{R}^2 | z_2 = x_2(z_1)\right\}\right) = 0\\ &\quad\Longleftrightarrow \lambda_1\left(\{z_2 \in \mathbb{R} | z_2 = x_2(z_1)\}\right) = 0 \quad \text{a.e. } z_1 \in \mathbb{R}\end{aligned}$$

and that the right-hand side is evidently true. □

Before extending the previous result on continuity to the stronger Lipschitz continuity, we introduce the following two assumptions on the density g_ξ of a two-dimensional random vector ξ:

$$\exists C \geq 0 : g_{\xi_1}(r)\left(= \int_{\mathbb{R}} g_\xi(r, s)\, ds\right) \leq C \quad \text{a.e. } r \in \mathbb{R} \tag{16}$$

$$\sup_{s \in \mathbb{R}} g_\xi(\cdot, s) \in L^2(\mathbb{R}) \tag{17}$$

Note that (16) means that the first marginal density of ξ (which is the density g_{ξ_1} of the first component of ξ) is bounded.

Proposition 3 *Let the density g_ξ of ξ satisfy (16) and (17). Then, φ is Lipschitz continuous.*

Proof Consider an arbitrary couple $x, y \in \mathcal{X}$. We start with the obvious estimate

$$|\varphi(x) - \varphi(y)| \leq |\varphi(x_1, x_2) - \varphi(y_1, x_2)| + |\varphi(y_1, x_2) - \varphi(y_1, y_2)|. \tag{18}$$

Without loss of generality, assume that $x_1 \leq y_1$. Now, by (15), and taking into account assumption (16), we have that

$$\begin{aligned}|\varphi(x_1, x_2) - \varphi(y_1, x_2)| &= |\mathbb{P}(\xi_1 \leq x_1, \xi_2 \leq x_2(\xi_1)) - \mathbb{P}(\xi_1 \leq y_1, \xi_2 \leq x_2(\xi_1))|\\ &= \mathbb{P}(x_1 < \xi_1 \leq y_1, \xi_2 \leq x_2(\xi_1))\\ &= \int_{x_1}^{y_1}\int_{-\infty}^{x_2(r)} g_\xi(r, s)\, dsdr \leq \int_{x_1}^{y_1}\int_{-\infty}^{\infty} g_\xi(r, s)\, dsdr\\ &\leq C\,|y_1 - x_1|.\end{aligned}$$

Likewise, exploiting (17), the fact that $x_2, y_2 \in L^2(\mathbb{R})$ and the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned}
|\varphi(y_1, x_2) - \varphi(y_1, y_2)| &= |\mathbb{P}(\xi_1 \leq y_1, \xi_2 \leq x_2(\xi_1)) - \mathbb{P}(\xi_1 \leq y_1, \xi_2 \leq y_2(\xi_1))| \\
&= \left| \int_{-\infty}^{y_1} \left(\int_{-\infty}^{x_2(r)} g_\xi(r,s)\, ds - \int_{-\infty}^{y_2(r)} g_\xi(r,s)\, ds \right) dr \right| \\
&\leq \int_{-\infty}^{y_1} \left| \int_{-\infty}^{x_2(r)} g_\xi(r,s)\, ds - \int_{-\infty}^{y_2(r)} g_\xi(r,s)\, ds \right| dr \\
&= \int_{-\infty}^{y_1} \int_{\min\{x_2(r), y_2(r)\}}^{\max\{x_2(r), y_2(r)\}} g_\xi(r,s)\, ds dr \\
&\leq \int_{-\infty}^{\infty} \sup_{s \in \mathbb{R}} g_\xi(r,s)\, |x_2(r) - y_2(r)|\, dr \\
&\leq \left\| \sup_{s \in \mathbb{R}} g_\xi(r,s) \right\|_{L^2(\mathbb{R})} \|x_2 - y_2\|_{L^2(\mathbb{R})} = \tilde{C} \|x_2 - y_2\|_{L^2(\mathbb{R})}.
\end{aligned}$$

Along with (18), we conclude that

$$|\varphi(x) - \varphi(y)| \leq (C + \tilde{C}) \|x - y\|_{\mathcal{X}}.$$

□

The following example shows, that the assumptions of Proposition 3 are not strong enough to guarantee the differentiability of φ:

Example 2 Let $\xi \sim \mathcal{N}(0, I_2)$ have a bivariate standard Gaussian distribution (uncorrelated components with mean zero and unit variance). By Proposition 6, the assumptions (16) and (17) of Proposition 3 are satisfied and, hence, φ is Lipschitz continuous. On the other hand, φ fails to be differentiable. To see this, we fix $\hat{x}_2 := \chi_{[0,1]} \in L^2(\mathbb{R})$ and observe that the partial real function $\tilde{\varphi}(x_1) := \varphi(x_1, \hat{x}_2)$ fails to be differentiable. Indeed, the following explicit representation can be immediately verified, where Φ refers to the cumulative distribution function of the one dimensional standard Gaussian distribution:

$$\tilde{\varphi}(x_1) = \begin{cases} \Phi(0)\, \Phi(x_1) & x_1 \leq 0 \\ \Phi^2(0) + \Phi(1)\, (\Phi(x_1) - \Phi(0)) & x_1 \in [0,1] \\ \Phi^2(0) + \Phi(1)\, (\Phi(1) - \Phi(0)) + (\Phi(0)\, (\Phi(x_1) - \Phi(1))) & x_1 \geq 0 \end{cases}$$

The graph of this function is shown in Fig. 1. Clearly, $\tilde{\varphi}$ is Lipschitz continuous because φ is so. On the other hand, it fails to be differentiable at $x_1 = 0$ and $x_1 = 1$. This can be seen for $x_1 = 0$, for instance, by deriving the first two expressions above at 0. With f denoting the density of the standard Gaussian distribution, the derivative of the first expression-yielding the left directional derivative of $\tilde{\varphi}$ at 0-gives $\Phi(0)\, f(0)$. On the other hand, the derivative of the second expression-yielding the right directional

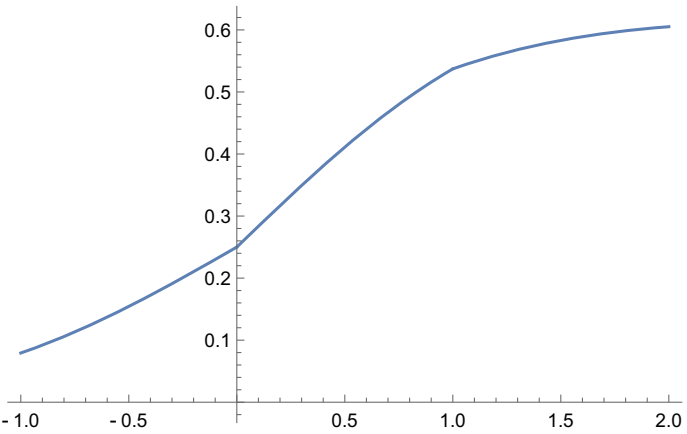

Fig. 1 Plot of function $\tilde{\varphi}$ from Example 2

derivative of $\tilde{\varphi}$ at 0-gives $\Phi(1) f(0)$. Since $\Phi(1) > \Phi(0)$ and $f(0) > 0$, both values are different, hence $\tilde{\varphi}$ fails to be differentiable at 0.

We shall see in the next section that the reason for the failure of differentiability of φ in Example 2 is the discontinuity of the second stage policy $x_2 := \chi_{[0,1]}$ at which the derivative is considered. More precisely, this circumstance concerns just the partial differentiability of φ with respect to its first argument x_1, whereas the partial differentiability of φ with respect to x_2 remains unaffected by a possible discontinuity of x_2.

3.2 Differentiability

Before verifying the partial differentiability of φ with respect to its first argument, we shall prove the following

Lemma 2 *Let a bivariate probability density g satisfy the following technical (uniform calmness) condition:*

$$\begin{gathered} \forall \bar{r} \in \mathbb{R}\ \exists l \in L^1(\mathbb{R}), \varepsilon > 0: \\ |g(r,s) - g(\bar{r},s)| \leq l(s)\,|r - \bar{r}| \quad \forall r \in (\bar{r} - \varepsilon, \bar{r} + \varepsilon)\ \ a.e.\ s \in \mathbb{R}. \end{gathered} \tag{19}$$

Assume further, that $f : \mathbb{R} \to \mathbb{R}$ is continuous. Then, the function

$$\alpha(r) := \int_{-\infty}^{f(r)} g(r,s)\,ds \quad (r \in \mathbb{R}) \tag{20}$$

is finite-valued and continuous.

Proof Fix an arbitrary $\bar{r} \in \mathbb{R}$ and consider an arbitrary sequence $r_n \to \bar{r}$. We are going to show that $\alpha(r_n) \to \alpha(\bar{r})$. We observe first that

$$g(r_n, s)\, \chi_{(-\infty, f(r_n)]}(s) \to g(\bar{r}, s)\, \chi_{(-\infty, f(\bar{r})]}(s) \quad \text{a.e. } s \in \mathbb{R}. \tag{21}$$

Indeed, if $s < f(\bar{r})$, then, $s < f(r_n)$ and if $s > f(\bar{r})$, then, $s > f(r_n)$ for n large enough by continuity of f. Hence, for each $s \neq f(\bar{r})$, one has that $\chi_{(-\infty, f(r_n)]}(s) = \chi_{(-\infty, f(\bar{r})]}(s)$ for n large enough. By (19), there exists a subset $A \subseteq \mathbb{R}$ such that $\lambda_1(A) = 0$ and $g(r_n, s) \to g(\bar{r}, s)$ for all $s \in \mathbb{R} \backslash A$. Consequently, (21) holds true for all $s \in \mathbb{R} \backslash (A \cup \{f(\bar{r})\})$, where $\lambda_1(A \cup \{f(\bar{r})\}) = 0$.

From (19) we conclude that, for n large enough and almost every $s \in \mathbb{R}$,

$$|g(r_n, s) - g(\bar{r}, s)| \leq l(s)\, |r_n - \bar{r}| \leq l(s).$$

Therefore, for n large enough and almost every $s \in \mathbb{R}$,

$$g(r_n, s)\, \chi_{(-\infty, f(r_n)]}(s) \leq g(r_n, s) \leq l(s) + g(\bar{r}, s)\,. \tag{22}$$

We show that $g(\bar{r}, \cdot) \in L^1(\mathbb{R})$: Indeed, as $g \in L^1\left(\mathbb{R}^2\right)$ (as a probability density), Fubini's Theorem yields that $g(r, \cdot) \in L^1(\mathbb{R})$ for almost every $r \in \mathbb{R}$. Hence, there exists some $\tilde{r} \in (\bar{r} - \varepsilon, \bar{r} + \varepsilon)$ with ε from (19) such that $g(\tilde{r}, \cdot) \in L^1(\mathbb{R})$ and

$$g(\bar{r}, s) \leq l(s)\, |\tilde{r} - \bar{r}| + g(\tilde{r}, s) \quad \text{a.e. } s \in \mathbb{R}.$$

Since $l \in L^1(\mathbb{R})$ and $|\tilde{r} - \bar{r}| \leq \varepsilon$, it follows that $g(\bar{r}, \cdot) \in L^1(\mathbb{R})$. Hence, by (22), $l + g(\bar{r}, \cdot)$ is an integrable majorant for the sequence of functions $g(r_n, \cdot)\, \chi_{(-\infty, f(r_n)]}$, which by (21) converges pointwise almost everywhere to the function $g(\bar{r}, \cdot)\, \chi_{(-\infty, f(\bar{r})]}$. Therefore, by Lebesgue's dominated convergence theorem, the value

$$\alpha(\bar{r}) = \int g(\bar{r}, s)\, \chi_{(-\infty, f(\bar{r})]}(s) ds < \infty$$

in (20) is finite and it holds that

$$\alpha(r_n) = \int g(r_n, s)\, \chi_{(-\infty, f(r_n)]}(s) ds \to_n \alpha(\bar{r})\,.$$

Since $\bar{r}$ was chosen arbitrarily, we have shown that α is finite-valued and continuous. □

The preceding Lemma allows us to formulate the desired result on partial differentiability of φ with respect to its first argument:

Proposition 4 *Let the density g_ξ of ξ satisfies (19) and fix $\bar{x}_2 \in L^2(\mathbb{R})$ such that $\bar{x}_2$ is continuous. Then, the partial derivative of φ w.r.t. x_1 exists at any $(\bar{x}_1, \bar{x}_2)$, it equals*

$$\frac{\partial \varphi}{\partial x_1}(\bar{x}_1, \bar{x}_2) = \int_{-\infty}^{\bar{x}_2(\bar{x}_1)} g_\xi(\bar{x}_1, s)\, ds.$$

Moreover, it depends continuously on x_1.

Proof Let $\bar{x}_1$ be arbitrary. By (15), we have that

$$\varphi(\bar{x}_1, \bar{x}_2) = \int_{-\infty}^{\bar{x}_1} \int_{-\infty}^{\bar{x}_2(r)} g_\xi(r, s)\, ds dr = \int_{-\infty}^{\bar{x}_1} \alpha(r)\, dr$$

with α defined in Lemma 2 upon setting $f(r) := \bar{x}_2(r)$ and $g := g_\xi$. Since $\bar{x}_2$ is supposed to be continuous, the assumptions of Lemma 2 are satisfied. Thus, by taking into account that α is continuous according to 2, we arrive at

$$\frac{\partial \varphi}{\partial x_1}(\bar{x}_1, \bar{x}_2) = \lim_{h \to 0} \frac{\varphi(\bar{x}_1 + h, \bar{x}_2) - \varphi(\bar{x}_1, \bar{x}_2)}{h} = \alpha(\bar{x}_1) = \int_{-\infty}^{\bar{x}_2(\bar{x}_1)} g(\bar{x}_1, s)\, ds.$$

Continuity of $\frac{\partial \varphi}{\partial x_1}(\cdot, \bar{x}_2) = \alpha$ follows once more from the continuity of α. □

Observe, that a full continuity result (with respect to x_1 and x_2 simultaneously) cannot be expected for the partial derivative $\frac{\partial \varphi}{\partial x_1}$ because, by virtue of Example 2, it may not even be defined for discontinuous x_2 approaching the continuous policy $\bar{x}_2$. In contrast to the partial derivative w.r.t. x_1, the partial derivative of φ with respect to x_2 does not require any assumptions on the fixed second-stage policy $\bar{x}_2$ but rather some additional assumptions on the density g_ξ:

Proposition 5 *Let the density* g_ξ *of* ξ *satisfies assumption* (17) *as well as the assumption of being Lipschitz continuous in the second argument uniformly in the first argument:*

$$\exists C > 0 : \left| g_\xi(r, s) - g_\xi(r, t) \right| \leq C\, |s - t| \quad \forall r, s, t \in \mathbb{R}. \tag{23}$$

Fix an arbitrary $(\bar{x}_1, \bar{x}_2) \in \mathcal{X} = \mathbb{R} \times L^2(\mathbb{R})$. *Then, the partial derivative* $\nabla_{x_2} \varphi$ *exists at* $(\bar{x}_1, \bar{x}_2)$, *it is given by the expression*

$$\nabla_{x_2} \varphi(\bar{x}_1, \bar{x}_2) = g_\xi(\cdot, \bar{x}_2(\cdot))\, \chi_{(-\infty, \bar{x}_1]}. \tag{24}$$

and it is continuous in (x_1, x_2).

Proof We put $\gamma(x_2) := \varphi(\bar{x}_1, x_2)$ for all $x_2 \in L^2(\mathbb{R})$ and show that this function is Fréchet differentiable at $\bar{x}_2$. Define the linear function

$$A(h) := \int_{-\infty}^{\bar{x}_1} g_\xi(r, \bar{x}_2(r))\, h(r) dr \quad \left(h \in L^2(\mathbb{R}) \right). \tag{25}$$

From (17) we infer that $g_\xi(\cdot, \bar{x}_2(\cdot)) \in L^2(\mathbb{R})$, whence by the Cauchy-Schwarz inequality,

$$|A(h)| \leq \int_{-\infty}^{\infty} g_\xi(r, \bar{x}_2(r))\, |h(r)|\, dr \leq \left\| g_\xi(\cdot, \bar{x}_2(\cdot)) \right\|_{L^2} \|h\|_{L^2}.$$

Consequently, A is a continuous linear functional. Hence, the Fréchet differentiability of γ at $\bar{x}_2$ will be proven, once we can show that

$$\lim_{\|h\|_{L^2}\to 0} \|h\|_{L^2}^{-1} \left(\gamma\left(\bar{x}_2+h\right)-\gamma\left(\bar{x}_2\right)-A(h)\right)=0. \tag{26}$$

Indeed, the definition of γ and (15) entail that

$$\begin{aligned}
&\gamma\left(\bar{x}_2+h\right)-\gamma\left(\bar{x}_2\right)-A(h)\\
&=\int_{-\infty}^{\bar{x}_1}\int_{-\infty}^{\bar{x}_2(r)+h(r)} g_\xi\left(r,s\right)dsdr-\int_{-\infty}^{\bar{x}_1}\int_{-\infty}^{\bar{x}_2(r)} g_\xi\left(r,s\right)dsdr\\
&\quad-\int_{-\infty}^{\bar{x}_1} g_\xi\left(r,\bar{x}_2(r)\right)h(r)dr\\
&=\int_{-\infty}^{\bar{x}_1}\left(\left(\operatorname{sgn} h\left(r\right)\right)\int_{\min\{\bar{x}_2(r),\bar{x}_2(r)+h(r)\}}^{\max\{\bar{x}_2(r),\bar{x}_2(r)+h(r)\}} g_\xi\left(r,s\right)ds-g_\xi\left(r,\bar{x}_2(r)\right)h(r)\right)dr\\
&=\int_{-\infty}^{\bar{x}_1}\left(\operatorname{sgn} h\left(r\right)\right)\int_{\min\{\bar{x}_2(r),\bar{x}_2(r)+h(r)\}}^{\max\{\bar{x}_2(r),\bar{x}_2(r)+h(r)\}}\left(g_\xi\left(r,s\right)-g_\xi\left(r,\bar{x}_2(r)\right)\right)dsdr.
\end{aligned}$$

By (23), we have that

$$\begin{aligned}
&\left|g_\xi\left(r,s\right)-g_\xi\left(r,\bar{x}_2(r)\right)\right|\le C\left|h\left(r\right)\right|\\
&\quad\forall r\in\mathbb{R}\ \forall s\in\left[\min\{\bar{x}_2\left(r\right),\bar{x}_2(r)+h\left(r\right)\},\max\{\bar{x}_2\left(r\right),\bar{x}_2(r)+h\left(r\right)\}\right].
\end{aligned}$$

Consequently, we derive the following relation implying (26).

$$\left|\gamma\left(\bar{x}_2+h\right)-\gamma\left(\bar{x}_2\right)-A(h)\right|\le C\int_{-\infty}^{\bar{x}_1}\left|h\left(r\right)\right|^2dr=C\left\|h\right\|_{L^2}^2.$$

It follows that $\nabla_{x_2}\varphi\left(\bar{x}_1,\bar{x}_2\right)=\nabla\gamma\left(\bar{x}_2\right)=A$. Since A in (25) has been shown to be a continuous linear functional on $L^2\left(\mathbb{R}\right)$, it can be identified with the function $g_\xi\left(\cdot,\bar{x}_2(\cdot)\right)\chi_{(-\infty,\bar{x}_1]}\in L^2\left(\mathbb{R}\right)$. This entails the asserted formula (24). It remains to show that the expression given there depends continuously on (x_1,x_2). To this aim, consider a sequence $(x_1^{(n)},x_2^{(n)})$ in $\mathcal{X}$ strongly converging to $(\bar{x}_1,\bar{x}_2)\in\mathcal{X}$. We have to show that

$$\nabla_{x_2}\varphi(x_1^{(n)},x_2^{(n)})\to_n\nabla_{x_2}\varphi\left(\bar{x}_1,\bar{x}_2\right)$$

in $L^2\left(\mathbb{R}\right)$. We will do this by showing the equivalent fact that every subsequence $(x_1^{(n_k)},x_2^{(n_k)})$ of $(x_1^{(n)},x_2^{(n)})$ has again a subsequence $(x_1^{(n_{k_l})},x_2^{(n_{k_l})})$ such that

$$\nabla_{x_2}\varphi(x_1^{(n_{k_l})},x_2^{(n_{k_l})})\to_l\nabla_{x_2}\varphi\left(\bar{x}_1,\bar{x}_2\right) \tag{27}$$

in $L^2\left(\mathbb{R}\right)$. So, let $(x_1^{(n_k)},x_2^{(n_k)})$ be such an arbitrary subsequence. Observe first that the strong convergence $(x_1^{(n_k)},x_2^{(n_k)})\to_l(\bar{x}_1,\bar{x}_2)$ in $\mathbb{R}\times L^2\left(\mathbb{R}\right)$ implies the almost

everywhere pointwise convergence for a subsequence:

$$(x_1^{(n_{k_l})}(r), x_2^{(n_{k_l})}(r)) \to_l (\bar{x}_1(r), \bar{x}_2(r)) \quad \text{a.e. } r \in \mathbb{R}. \tag{28}$$

As g_ξ is continuous in its second argument by (23), it follows from (28) that

$$g_\xi(r, x_2^{(n_{k_l})}(r)) \to_l g_\xi(r, \bar{x}_2(r)) \quad \text{a.e. } r \in \mathbb{R}.$$

Moreover,

$$\chi_{(-\infty, x_1^{(n_{k_l})}]}(r) \to_l \chi_{(-\infty,\bar{x}_1]}(r) \quad \forall r \in \mathbb{R} \setminus \{\bar{x}_1\}.$$

We conclude from (24) that

$$\begin{aligned}\nabla_{x_2}\varphi(x_1^{(n_{k_l})}, x_2^{(n_{k_l})})(r) &= g_\xi(r, x_2^{(n_{k_l})}(r))\chi_{(-\infty, x_1^{(n_{k_l})}]}(r) \\ &\to_l g_\xi(r, \bar{x}_2(r))\chi_{(-\infty,\bar{x}_1]}(r) = \nabla_{x_2}\varphi(\bar{x}_1, \bar{x}_2).\end{aligned}$$

for almost every $r \in \mathbb{R}$. On the other hand, by (17)

$$\nabla_{x_2}\varphi(x_1^{(n_{k_l})}, x_2^{(n_{k_l})})(\cdot) = g_\xi(\cdot, x_2^{(n_{k_l})}(\cdot))\chi_{(-\infty, x_1^{(n_{k_l})}]}(\cdot) \le \sup_{s\in\mathbb{R}} g_\xi(\cdot, s) \in L^2(\mathbb{R}).$$

Therefore, Lebesgue's Dominated Convergence Theorem (for $L^2(\mathbb{R})$) yields the asserted convergence (27) in $L^2(\mathbb{R})$. □

3.3 Distributions satisfying the assumptions

In this Section we are going to specify the results of the preceding sections to the special case of a bivariate Gaussian distribution and a uniform distribution on a rectangle. First, we verify that all relevant assumptions are satisfied in the Gaussian case:

Proposition 6 *Let ξ be a bivariate random vector distributed according to $\xi \sim \mathcal{N}(\mu, \Sigma)$ with regular Σ. Then its density g_ξ satisfies the assumptions* (16)*,* (17)*,* (19) *and* (23)*.*

Proof The first marginal density of g_ξ is the density g_{ξ_1} of its first component $\xi_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$ which is bounded of course. Hence, (16) holds true. To show (17), recall that

$$g_\xi(r, s) = C \exp\left(-\frac{1}{2}(r-\mu_1, s-\mu_2)\,\Sigma^{-1}\begin{pmatrix} r-\mu_1 \\ s-\mu_2\end{pmatrix}\right), \tag{29}$$

where C is some normalizing factor. With $C_2 > 0$ denoting the smallest eigenvalue of Σ^{-1}, we infer that, for all $s \in \mathbb{R}$,

$$g_\xi(r, s) \le C \exp\left(-\frac{C_2}{2}(r-\mu_1)^2 - \frac{C_2}{2}(s-\mu_2)^2\right)$$

$$\leq C \exp\left(-\frac{C_2}{2}(r-\mu_1)^2\right) \in L^2(\mathbb{R}) \tag{30}$$

which implies (17). In order to verify (19) and (23), we first calculate the gradient of g_ξ:

$$\nabla g_\xi(r,s) = -g_\xi(r,s)\,\Sigma^{-1}\begin{pmatrix} r-\mu_1 \\ s-\mu_2 \end{pmatrix} \quad \forall r,s \in \mathbb{R}.$$

Denoting $h_i(r) := \exp\left(-\frac{C_2}{2}(r-\mu_i)^2\right)$ for $i=1,2$, it follows from (30) that

$$\left\|\nabla g_\xi(r,s)\right\| \leq \tilde{C} h_1(r) h_2(s) \left\|(r-\mu_1, s-\mu_2)\right\| \quad \forall r,s \in \mathbb{R},$$

where $\tilde{C} := C\left\|\Sigma^{-1}\right\|$. Since the function $h_1(r)h_2(s)\left\|(r-\mu_1, s-\mu_2)\right\|^2$ is bounded from above, we have that, for some $C_3 > 0$,

$$\sqrt{h_1(r)h_2(s)}\left\|(r-\mu_1, s-\mu_2)\right\| \leq C_3 \quad \forall r,s \in \mathbb{R}.$$

Hence, thanks to $h_1, h_2 \leq 1$, we get that

$$\left\|\nabla g_\xi(r,s)\right\| \leq \tilde{C}C_3\sqrt{h_1(r)h_2(s)} \leq \tilde{C}C_3\sqrt{h_2(s)} \leq \tilde{C}C_3 \quad \forall r,s \in \mathbb{R}.$$

Then, (23) follows from the Mean Value Theorem:

$$\left|g_\xi(r,s) - g_\xi(r,t)\right| \leq \tilde{C}C_3\,|s-t| \quad \forall r,s,t \in \mathbb{R}.$$

Similarly, for arbitrarily fixed $\bar{r} \in \mathbb{R}$,

$$\left|g_\xi(r,s) - g_\xi(\bar{r},s)\right| \leq \tilde{C}C_3\sqrt{h_2(s)}\,|r-\bar{r}| \quad \forall r,s \in \mathbb{R},$$

where $\sqrt{h_2} \in L^1(\mathbb{R})$. This proves (19). □

Theorem 3 *Let ξ be a bivariate random vector distributed according to $\mathcal{N}(\mu,\Sigma)$ with regular Σ . Then, the probability function φ in (15) is Lipschitz continuous and has a second partial derivative at an arbitrary $(\bar{x}_1,\bar{x}_2) \in \mathcal{X} = \mathbb{R} \times L^2(\mathbb{R})$ which is given by the explicit formula*

$$\nabla_{x_2}\varphi(\bar{x}_1,\bar{x}_2)(r) = \begin{cases} \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}\left(\begin{smallmatrix} r-\mu_1 \\ \bar{x}_2(r)-\mu_2 \end{smallmatrix}\right)^\top \Sigma^{-1}\left(\begin{smallmatrix} r-\mu_1 \\ \bar{x}_2(r)-\mu_2 \end{smallmatrix}\right)\right) & \text{if } r \leq \bar{x}_1 \\ 0 & \text{if } r > \bar{x}_1 \end{cases}. \tag{31}$$

Here, $\nabla_{x_2}\varphi$ depends continuously (in the norm of $\mathcal{X}$) on $x = (x_1,x_2)$. Moreover, φ has a first partial derivative at an arbitrary $(\bar{x}_1,\bar{x}_2) \in \mathcal{X} = \mathbb{R} \times L^2(\mathbb{R})$ with continuous $\bar{x}_2$ which is given by the explicit formula

$$\frac{\partial \varphi}{\partial x_1}(\bar{x}_1, \bar{x}_2) = \frac{1}{\sqrt{2\pi\,\Sigma_{11}}} \exp\left(-\frac{1}{2\Sigma_{11}}(r-\mu_1)^2\right) \Phi\left(\frac{\bar{x}_2(\bar{x}_1) - \mu_2 - \Sigma_{11}^{-1}\Sigma_{12}(\bar{x}_1-\mu_1)}{\sqrt{\Sigma_{22} - \Sigma_{11}^{-1}\Sigma_{12}^2}}\right), \tag{32}$$

where $\Phi(t) := (2\pi)^{-1/2} \int_{-\infty}^{t} e^{-s^2/2} ds$ *refers to the cumulative distribution function of the one-dimensional standard Gaussian distribution* $\mathcal{N}(0,1)$ *. Here* $\frac{\partial \varphi}{\partial x_1}(\cdot, \bar{x}_2)$ *is continuous.*

Proof The Lipschitz continuity, the existence of partial derivatives and the corresponding continuity statements follow from Propositions 3, 4 and 5 via Proposition 6. Relation (31) is obtained by specifying (24) for the density of $\mathcal{N}(\mu, \Sigma)$ (see (29) with $C := \left(2\pi\sqrt{\det \Sigma}\right)^{-1}$). Concerning (32), we recall the formula derived in Proposition 4:

$$\begin{aligned}\frac{\partial \varphi}{\partial x_1}(\bar{x}_1, \bar{x}_2) &= \int_{-\infty}^{\bar{x}_2(\bar{x}_1)} g_\xi(\bar{x}_1, s)\, ds = g_{\xi_1}(\bar{x}_1) \int_{-\infty}^{\bar{x}_2(\bar{x}_1)} \frac{g_\xi(\bar{x}_1, s)}{g_{\xi_1}(\bar{x}_1)} ds \\ &= g_{\xi_1}(\bar{x}_1) \int_{-\infty}^{\bar{x}_2(\bar{x}_1)} g_{\xi_2|\xi_1=\bar{x}_1}(s)\, ds \\ &= g_{\xi_1}(\bar{x}_1)\, G_{\xi_2|\xi_1\bar{x}_1}(\bar{x}_2(\bar{x}_1)),\end{aligned} \tag{33}$$

where $g_{\xi_2|\xi_1=\bar{x}_1}$ and $G_{\xi_2|\xi_1=\bar{x}_1}$ refer to the conditional density and cumulative distribution function, respectively, of ξ_2 given $\xi_1 = \bar{x}_1$. As it is well known for the Gaussian case assumed here, one has that the conditioned random variable $\xi_2|\xi_1 = \bar{x}_1$ has a one-dimensional Gaussian distribution with

$$(\xi_2|\xi_1 = \bar{x}_1) \sim \mathcal{N}\left(\mu_2 + \Sigma_{11}^{-1}\Sigma_{12}(\bar{x}_1-\mu_1), \Sigma_{22} - \Sigma_{11}^{-1}\Sigma_{12}^2\right).$$

After normalization, we get that

$$\eta := \frac{(\xi_2|\xi_1 = \bar{x}_1) - \mu_2 - \Sigma_{11}^{-1}\Sigma_{12}(\bar{x}_1-\mu_1)}{\sqrt{\Sigma_{22} - \Sigma_{11}^{-1}\Sigma_{12}^2}} \sim \mathcal{N}(0,1).$$

Now, the definition of $G_{\xi_2|\xi_1=\bar{x}_1}$ yields that

$$\begin{aligned}G_{\xi_2|\xi_1=\bar{x}_1}(\bar{x}_2(\bar{x}_1)) &= \mathbb{P}\left((\xi_2|\xi_1 = \bar{x}_1) \le \bar{x}_2(\bar{x}_1)\right) \\ &= \mathbb{P}\left(\eta \le \frac{\bar{x}_2(\bar{x}_1) - \mu_2 - \Sigma_{11}^{-1}\Sigma_{12}(\bar{x}_1-\mu_1)}{\sqrt{\Sigma_{22} - \Sigma_{11}^{-1}\Sigma_{12}^2}}\right)\end{aligned}$$

$$= \Phi\left(\frac{\bar{x}_2(\bar{x}_1) - \mu_2 - \Sigma_{11}^{-1}\Sigma_{12}\,(\bar{x}_1 - \mu_1)}{\sqrt{\Sigma_{22} - \Sigma_{11}^{-1}\Sigma_{12}^2}}\right),$$

where Φ is the cumulative distribution function of $\mathcal{N}(0, 1)$. Now the asserted formula (32) follows form (33) upon plugging in the formula for the first marginal density g_{ξ_1} of g_ξ having distribution $\mathcal{N}(\mu_1, \Sigma_{11})$. □

Corresponding results can be expected for many other bivariate distributions having continuous density. As a contrast, we briefly refer to uniform distributions over rectangles for which no differentiability results for φ but at least Lipschitz continuity can be expected:

Proposition 7 *Let ξ be a bivariate random vector having a uniform distribution over the rectangle $[a, b] \times [c, d]$. Then, the probability function φ in (15) is Lipschitz continuous.*

Proof The density g_ξ satisfies the assumptions (16) and (17) thanks to the following easy to verify relations

$$g_{\xi_1} = \frac{1}{b-a}\chi_{[a,b]}; \quad \sup_{s\in\mathbb{R}} g_\xi(\cdot, s) = \frac{1}{(b-a)(d-c)}\chi_{[a,b]}.$$

Now, the assertion follows from Proposition 3. □

Note, that a uniform distribution as in the previous Proposition cannot satisfy relations (19) and (23) because of the discontinuity of its density. Therefore, no differentiability results as in Propositions 4 and 5 can be expected and counter examples are easily constructed.

3.4 Application to an optimization problem

In the following, we consider the simple dynamic probabilistic constraint as a part of the following two-stage optimization problem (15):

$$\min_{x\in\mathcal{X}} \left\{c_1 x_1 + c_2 \mathbb{E}x_2(\xi_1)\,\chi_{(-\infty,x_1]}(\xi_1)\,|\varphi(x) \geq p\right\}, \tag{34}$$

where ξ (occuring in the definition of φ) is a bivariate random vector distributed according to $\mathcal{N}(\mu, \Sigma)$. The objective is linear in the decisions, it could represent, for instance, linear costs. Since the second stage decision is random, its costs are represented as an expected value. Note, however, that considering the full expected value $\mathbb{E}x_2(\xi_1)$ would not make much sense: Indeed, since function values of $x_2(\xi_1)$ for arguments $\xi_1 > x_1$ do not affect the probability $\varphi(x)$ (see (15)), one could drive the expected value $\mathbb{E}x_2(\xi_1)$ to $-\infty$ while keeping the decision x feasible. Therefore, we measure the costs of x_2 by ignoring in the objective its values beyond x_1 and rather considering the expected value of $x_2\chi_{(-\infty,x_1]}$.

In a first step, one might be interested in deriving some information from necessary optimality conditions for this problem. Here, one has to take into account that φ is not continuously differentiable [see Example (2)]. However, φ is continuously partially differentiable with respect to x_2 thanks to Proposition 5. This suggests to consider the decomposed version of problem (34):

$$\min_{x_1 \in \mathbb{R}} \left\{ c_1 x_1 + \min_{x_2 \in L^2(\mathbb{R})} \left\{ c_2 \mathbb{E} x_2 \left(\xi_1\right) \chi_{(-\infty, x_1]} \left(\xi_1\right) | \varphi(x_1, x_2) \geq p \right\} \right\}. \tag{35}$$

Here, the one-dimensional outer minimization over x_1 can be realized by elementary numerical approaches. Therefore, our interest will focus on the inner minimization problem over $x_2 \in L^2\left(\mathbb{R}\right)$ for some fixed $\bar{x}_1 \in \mathbb{R}$:

$$\min_{x_2 \in L^2(\mathbb{R})} \left\{ c_2 \mathbb{E} x_2 \left(\xi_1\right) \chi_{(-\infty, \bar{x}_1]} \left(\xi_1\right) | \varphi(\bar{x}_1, x_2) \geq p \right\}. \tag{36}$$

For this inner optimization problem, the data (objective and constraint) are continuously differentiable and one can formulate necessary optimality conditions at some fixed $\bar{x}_2 \in L^2\left(\mathbb{R}\right)$ provided that $\nabla_{x_2} \varphi(\bar{x}_1, \bar{x}_2) \neq 0$. This, however, is an immediate consequence of (31). Hence, one may formulate the following necessary optimality condition:

Proposition 8 *Let $\bar{x}_2 \in L^2\left(\mathbb{R}\right)$ be a solution of the optimization problem (36) (with some fixed $\bar{x}_1 \in \mathbb{R}$). Then, $\bar{x}_2$ is affine linear on the set $(-\infty, x_1]$ with the explicit value of Σ_{12}/Σ_{11} for its slope.*

Proof Without loss of generality, we may assume that $c_2 = 1$ in (36) because the solution of the problem is not affected by the value of c_2. The gradient of the objective evaluated at $\bar{x}_2$ has to be a multiple of the gradient $\nabla \varphi(\bar{x}_1, \bar{x}_2)$ to the constraint in (36) also evaluated at $\bar{x}_2$. Clearly, the objective

$$\mathbb{E} x_2 \left(\xi_1\right) = \int x_2 \left(r\right) \chi_{(-\infty, x_1]} \left(r\right) g_{\xi_1}(r) dr$$

(with g_{ξ_1} referring to the density of ξ_1) has a gradient which is given by the function $\chi_{(-\infty, x_1]} g_{\xi_1}$. Hence, there exists some multiplier λ such that

$$g_{\xi_1}(r) = \lambda \nabla_{x_2} \varphi(\bar{x}_1, \bar{x}_2) \left(r\right) \quad \text{a.e. } r \leq \bar{x}_1.$$

Since the one-dimensional Gaussian density g_{ξ_1} is strictly positive, we infer that $\lambda > 0$. Given the explicit formula for g_{ξ_1} as well as for $\nabla_{x_2} \varphi(\bar{x}_1, \bar{x}_2)$ in (31), we derive the existence of constants $K_1, K_2 > 0$ (where the latter already incorporates the multiplier λ) such that for almost every $r \leq \bar{x}_1$:

$$K_1 \exp\left(-\frac{1}{2}\frac{(r-\mu_1)^2}{\Sigma_{11}}\right)$$
$$= K_2 \exp\left(-\frac{1}{2}(r-\mu_1, \bar{x}_2(r)-\mu_2)\,\Sigma^{-1}\begin{pmatrix} r-\mu_1 \\ \bar{x}_2(r)-\mu_2 \end{pmatrix}\right) \tag{37}$$

We fix an arbitrary r for which (37) holds true. Using the correlation ρ between the two components ξ_1 and ξ_2, the inverse covariance matrix can be written as

$$\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\Sigma_{11}} & -\frac{\rho}{\sqrt{\Sigma_{11}\Sigma_{22}}} \\ -\frac{\rho}{\sqrt{\Sigma_{11}\Sigma_{22}}} & \frac{1}{\Sigma_{22}} \end{pmatrix}; \quad \left(\rho := \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}\right).$$

Taking the log in (37) and rearranging terms, one arrives at

$$\log\frac{K_1}{K_2} = \frac{1}{1-\rho^2}\left(\frac{-\rho^2}{2\Sigma_{11}}(r-\mu_1)^2 + \frac{\rho}{\sqrt{\Sigma_{11}\Sigma_{22}}}(r-\mu_1)(\bar{x}_2(r)-\mu_2) - \frac{1}{2\Sigma_{22}}(\bar{x}_2(r)-\mu_2)^2\right).$$

Putting

$$\alpha := \frac{\bar{x}_2(r)-\mu_2}{\sqrt{\Sigma_{22}}}; \quad \beta := 2\left(1-\rho^2\right)\log\frac{K_1}{K_2},$$

the last identity can be rewritten as

$$\alpha^2 - \frac{2\rho}{\sqrt{\Sigma_{11}}}(r-\mu_1)\,\alpha + \frac{\rho^2}{\Sigma_{11}}(r-\mu_1)^2 + \beta = 0.$$

Resolution for α yields that

$$\alpha = \frac{\rho}{\sqrt{\Sigma_{11}}}(r-\mu_1) \pm \sqrt{-\beta}.$$

Resubstituting for α and β gives our assertion on the structure of $\bar{x}_2$:

$$\bar{x}_2(r) = \frac{\sqrt{\Sigma_{22}}\rho}{\sqrt{\Sigma_{11}}}r + \mu_2 - \frac{\sqrt{\Sigma_{22}}\rho}{\sqrt{\Sigma_{11}}}\mu_1 \pm \sqrt{\Sigma_{22}}\sqrt{2\left(1-\rho^2\right)\log\frac{K_2}{K_1}}$$

□

Unfortunately, since an affine linear function cannot belong to $L^2(\mathbb{R})$ unless it is identically zero, we draw the following negative conclusion of Proposition 8:

Corollary 3 *If the components ξ_1 and ξ_2 of ξ are not independent, then problem (36) has no local, much less global solution.*

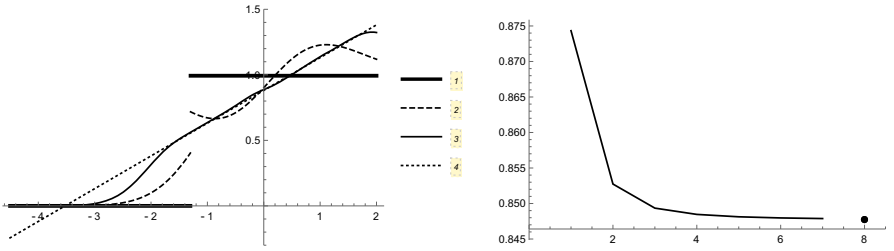

Fig. 2 Plot of several iterates for a projected gradient algorithm applied to problem (36) (left) and of associated values of the objective (right)

Proof The independence assumption implies that $\Sigma_{12} \neq 0$. Hence, if (36) had a local solution $\bar{x}_2 \in L^2(\mathbb{R})$, then this solution would be a linear function (in the range from $-\infty$ to $\bar{x}_1$) with nonzero slope by Proposition 8. Therefore it does not belong to $L^2(\mathbb{R})$, a contradiction.

Before deriving a remedy to the outcome of Corollary 3, we want to illustrate the use of the gradient information collected in (31) in a numerical context. We consider problem (36) with the following data:

$$c_2 = 1; \quad \bar{x}_1 = 2; \quad p = 0.8; \quad \xi \sim \mathcal{N}\left((0,0), \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}\right).$$

Using the explicit representation of the gradients for the objective and the constraint (see proof of Proposition 8), we apply a simple projected gradient algorithm in order to improve the second stage decision x_2 in (36).

The left diagram of Fig. 2 shows some iterates of this algorithm. All plotted policies realize exactly the desired probability $p = 0.8$ in the definition of the chance constraint in (36). The starting point for x_2 was chosen as a simple step function "1", which after the first iteration turned into a nonlinear—still discontinuous—policy "2". After some further iterations, the policy becomes continuous. Interestingly, after seven iterations "3", the policy is affine linear on a certain subdomain. It turns out that on this subdomain, the policy perfectly coincides with the affine linear policy "4" satisfying the necessary optimality condition in Proposition 8. The latter is easily identified by its slope, which according to Proposition 8 calculates as $\Sigma_{12}/\Sigma_{1,1} = 0.25$ and by its intercept which has to be chosen in order to match the probability level $p = 0.8$. Observe, that all iterates decay to zero on the left end of the negative axis in order to belong to $L^2(\mathbb{R})$. The right diagram of Fig. 2 plots the objective for the first seven iterates and for the the affine linear policy from Proposition 8 (isolated point). Evidently, the necessary optimality condition from Proposition 8 still carries some information on the optimality condition though not belonging to the L^2 space.

3.5 Derivation of a global solution

Motivated by the negative result of Corollary 3, we consider now the optimization problem

$$\min_{(x_1,x_2)\in\mathbb{R}\times\mathcal{X}^*} \left\{c_1x_1 + c_2\mathbb{E}x_2\left(\xi_1\right)\chi_{(-\infty,x_1]}\left(\xi_1\right)|\varphi\left(x_1,x_2\right)\geq p;\ x_2\in M\right\}, \tag{38}$$

where $\mathcal{X}^*$ is the set of all Borel measurable functions $f:\mathbb{R}\to\mathbb{R}$, $c_1,c_2>0$ are some positive cost coefficients, $\xi\sim\mathcal{N}\left(\mu,\Sigma\right)$ is a bivariate Gaussian random vector, φ is the probability function defined in (15), $p\in(0,1]$ is some probability level and the additional constraint set is given by

$$M:=\left\{x_2\in\mathcal{X}^*|\,x_2\left(r\right)\geq\mu_2+\frac{\Sigma_{12}}{\Sigma_{11}}\left(r-\mu_1\right)\right\}.$$

This problem differs from (34) first in that the space of second stage decisions is much larger than the space $L^2\left(\mathbb{R}\right)$ considered before, so that it also includes affine linear functions. At the same time we add the technical constraint $x_2\in M$ in order to identify a global solution by a direct argument rather than by necessary optimality conditions. However, we shall make use of the result obtained before in Proposition 8 to get the right guess for a candidate of an optimal second stage solution (affine linear function with slope $\frac{\Sigma_{12}}{\Sigma_{11}}$). Note that we do not require the expectation of $x_2\chi_{(-\infty,x_1]}$ in (38) to be finite. It turns out that the solution of (38) can be reduced to a one-dimensional optimization. Before stating the result, we introduce the real functions

$$\alpha\left(t\right) := \Phi^{-1}\left(\frac{p}{\Phi\left(\frac{t-\mu_1}{\sqrt{\Sigma_{11}}}\right)}\right)\sqrt{\Sigma_{22}-\frac{\Sigma_{12}^2}{\Sigma_{11}}}+\mu_2-\frac{\Sigma_{12}}{\Sigma_{11}}\mu_1 \tag{39}$$

$$\left(t>\sqrt{\Sigma_{11}}\Phi^{-1}\left(p\right)+\mu_1\right), \tag{40}$$

$$\beta\left(t\right) := \mu_1-\sqrt{\Sigma_{11}}\frac{\phi\left(\frac{t-\mu_1}{\sqrt{\Sigma_{11}}}\right)}{\Phi\left(\frac{t-\mu_1}{\sqrt{\Sigma_{11}}}\right)} \tag{41}$$

where μ_i and Σ_{ij} refer to the corresponding components of μ and Σ, respectively and Φ denotes as before the cumulative distribution function of the one-dimensional standard Gaussian distribution $\mathcal{N}\left(0,1\right)$. Recall that Φ is invertible with inverse Φ^{-1} called the quantile function of $\mathcal{N}\left(0,1\right)$. Since Φ^{-1} is defined only on the open interval $(0,1)$, (39) is defined correctly only if $p<1$ (which we shall impose in the Theorem below) and if

$$\Phi\left(\frac{t-\mu_1}{\sqrt{\Sigma_{11}}}\right)>p,$$

which leads to the constrained domain of definition in (39). Finally, the function

$$\phi(t) := \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

appearing in (41), is the density of the one-dimensional standard Gaussian distribution.

Theorem 4 *Let $p \in \left[\frac{1}{2}, 1\right)$ be given. Then, a global solution of problem (38) is given by $\left(x_1^*, x_2^*\right) \in \mathbb{R} \times \mathcal{X}^*$, where x_1^* is a minimizer over $\mathbb{R}$ of the real function $c_1 t + c_2 f(t)$ over the open interval $\left(\sqrt{\Sigma_{11}}\Phi^{-1}(p) + \mu_1, \infty\right)$, with*

$$f(t) := \left[\frac{\Sigma_{12}}{\Sigma_{11}}\beta(t) + \alpha(t)\,\Phi\left(\frac{t-\mu_1}{\sqrt{\Sigma_{11}}}\right)\right] \quad (t \in \mathbb{R})$$

and

$$x_2^*(r) := \frac{\Sigma_{12}}{\Sigma_{11}} r + \alpha\left(x_1^*\right) \quad (r \in \mathbb{R}).$$

Proof We start our proof with an intermediary result. To this aim, fix an arbitrary

$$x_1 > \sqrt{\Sigma_{11}}\Phi^{-1}(p) + \mu_1, \tag{42}$$

in order to make the value $\alpha(x_1)$ well defined in (39). We claim that the second stage policy defined by

$$y(r) := \frac{\Sigma_{12}}{\Sigma_{11}} r + \alpha(x_1) \quad (r \in \mathbb{R}) \tag{43}$$

is a global solution to the problem

$$\min_{x_2 \in \mathcal{X}^*} \left\{c_2 \mathbb{E} x_2(\xi_1)\, \chi_{(-\infty, x_1]}(\xi_1) \,|\varphi(x_1, x_2) \geq p;\ x_2 \in M\right\}. \tag{44}$$

In a first step we check the feasibility of y in this problem. Introducing the linear transformation

$$\eta_1 := \xi_1; \quad \eta_2 := \xi_2 - \frac{\Sigma_{12}}{\Sigma_{11}}\xi_1 \tag{45}$$

of random variables, we observe that according to well-known rules the random vector $\eta := (\eta_1, \eta_2)$ obeys a bivariate Gaussian law $\mathcal{N}\left(\tilde{\mu}, \tilde{\Sigma}\right)$ with parameters

$$\tilde{\mu} = \left(\mu_1, \mu_2 - \frac{\Sigma_{12}}{\Sigma_{11}}\mu_1\right), \quad \tilde{\Sigma} = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \frac{\Sigma_{12}^2}{\Sigma_{11}} \end{pmatrix}.$$

In particular, the components η_1, η_2—having zero covariance—are independent. It follows from the definition (15) of φ and from (43) that

$$\varphi(x_1, y) = \mathbb{P}(\xi_1 \leq x_1, \xi_2 \leq y(\xi_1)) = \mathbb{P}(\eta_1 \leq x_1, \eta_2 \leq \alpha(x_1))$$

$$= \mathbb{P}(\eta_1 \leq x_1)\, \mathbb{P}(\eta_2 \leq \alpha(x_1)),$$

where the last equality follows from the independence of η_1 and η_2. Again, by the well-known transformation laws of Gaussian distributions as well as by (39), it holds that

$$\begin{aligned}
\mathbb{P}(\eta_1 \leq x_1) &= \Phi\left(\frac{x_1 - \tilde{\mu}_1}{\sqrt{\tilde{\Sigma}_{11}}}\right) = \Phi\left(\frac{x_1 - \mu_1}{\sqrt{\Sigma_{11}}}\right); \\
\mathbb{P}(\eta_2 \leq \alpha(x_1)) &= \Phi\left(\frac{\alpha(x_1) - \tilde{\mu}_2}{\sqrt{\tilde{\Sigma}_{22}}}\right) = \Phi\left(\frac{\alpha(x_1) - \mu_2 + \frac{\Sigma_{12}}{\Sigma_{11}}\mu_1}{\sqrt{\Sigma_{22} - \frac{\Sigma_{12}^2}{\Sigma_{11}}}}\right) \\
&= \frac{p}{\Phi\left(\frac{x_1 - \mu_1}{\sqrt{\Sigma_{11}}}\right)}. \qquad (46)
\end{aligned}$$

Consequently, we arrive at $\varphi(x_1, y) = p$. Hence, $x_2 := y$ is feasible with respect to the constraint $\varphi(x_1, x_2) \geq p$. Next, we verify that $y \in M$. By definition of y, M and α, it suffices to show that

$$\alpha(x_1) \geq \mu_2 - \frac{\Sigma_{12}}{\Sigma_{11}}\mu_1 = \tilde{\mu}_2. \qquad (47)$$

Indeed, the assumption that $\alpha(x_1) < \tilde{\mu}_2$ would lead—via the fact that the values of Φ are strictly smaller than one—to the contradiction

$$p < \frac{p}{\Phi\left(\frac{x_1 - \mu_1}{\sqrt{\Sigma_{11}}}\right)} = \mathbb{P}(\eta_2 \leq \alpha(x_1)) < \mathbb{P}(\eta_2 \leq \tilde{\mu}_2) \leq \frac{1}{2}$$

with our assumption that $p \geq \frac{1}{2}$. Summarizing, $y \in \mathcal{X}^*$ defined by (43) is a feasible second stage policy in problem (44).

In the last step of our initial claim, we show that there is no other feasible second stage decision in (44) that would yield a strictly smaller objective value than y. Indeed, assume to the contrary, that some function $\tilde{y} \in M$ with $\varphi(x_1, \tilde{y}) \geq p$ would realize in (44) a strictly smaller objective value than y. Then, since $c_2 > 0$,

$$\mathbb{E}\tilde{y}(\xi_1)\, \chi_{(-\infty, x_1]}(\xi_1) < \mathbb{E}y(\xi_1)\, \chi_{(-\infty, x_1]}(\xi_1). \qquad (48)$$

Now, by (45) and with g_η denoting the density of η,

$$\begin{aligned}
0 \leq \varphi(x_1, \tilde{y}) - \varphi(x_1, y) &= \mathbb{P}(\xi_1 \leq x_1, \xi_2 \leq \tilde{y}(\xi_1)) - \mathbb{P}(\xi_1 \leq x_1, \xi_2 \leq y(\xi_1)) \\
&= \mathbb{P}\left(\eta_1 \leq x_1, \eta_2 \leq \tilde{y}(\eta_1) - \frac{\Sigma_{12}}{\Sigma_{11}}\eta_1\right) - \mathbb{P}\left(\eta_1 \leq x_1, \eta_2 \leq y(\eta_1) - \frac{\Sigma_{12}}{\Sigma_{11}}\eta_1\right) \\
&= \int_{-\infty}^{x_1} \int_{-\infty}^{\tilde{y}(r) - \frac{\Sigma_{12}}{\Sigma_{11}}r} g_\eta(s, r)\, dsdr - \int_{-\infty}^{x_1} \int_{-\infty}^{y(r) - \frac{\Sigma_{12}}{\Sigma_{11}}r} g_\eta(s, r)\, dsdr.
\end{aligned}$$

Recalling that, by independence of the components η_1 and η_2 , we may write $g_\eta(s,r) = g_{\eta_1}(r)\, g_{\eta_2}(s)$, where g_{η_1},g_{η_2} refer to the one-dimensional densities of η_1 and η_2, we may—taking into account (43)—continue as

$$\begin{aligned} 0 &\le \int_{-\infty}^{x_1} g_{\eta_1}(r) \left(\int_{-\infty}^{\tilde{y}(r)-\frac{\Sigma_{12}}{\Sigma_{11}}r} g_{\eta_2}(s)\, ds \right) dr \\ &\quad - \int_{-\infty}^{x_1} g_{\eta_1}(r) \left(\int_{-\infty}^{\alpha(x_1)} g_{\eta_2}(s)\, ds \right) dr \\ &= \int_{-\infty}^{x_1} g_{\eta_1}(r)\, F_{\eta_2}\left(\tilde{y}(r) - \frac{\Sigma_{12}}{\Sigma_{11}}r \right) dr - \int_{-\infty}^{x_1} g_{\eta_1}(r)\, F_{\eta_2}(\alpha(x_1))\, dr \\ &= \int_{-\infty}^{x_1} g_{\eta_1}(r) \left[F_{\eta_2}\left(\tilde{y}(r) - \frac{\Sigma_{12}}{\Sigma_{11}}r \right) - F_{\eta_2}(\alpha(x_1)) \right] dr, \end{aligned} \tag{49}$$

where F_{η_2} refers to the cumulative distribution function of η_2. Since one-dimensional Gaussian distribution functions are concave right of their mean (their second derivative coincides with the first derivative of the density, and so is negative right of the mean), we have the relation

$$F_{\eta_2}(t) \le F_{\eta_2}(s) + F'_{\eta_2}(s)(t-s) \quad \forall s, t \ge \tilde{\mu}_2.$$

Now, $\alpha(x_1) \ge \tilde{\mu}_2$ by (47) and also, because of $\tilde{y} \in M$,

$$\tilde{y}(r) - \frac{\Sigma_{12}}{\Sigma_{11}}r \ge \mu_2 - \frac{\Sigma_{12}}{\Sigma_{11}}\mu_1 = \tilde{\mu}_2.$$

Therefore, we may conclude that

$$F_{\eta_2}\left(\tilde{y}(r) - \frac{\Sigma_{12}}{\Sigma_{11}}r \right) \le F_{\eta_2}(\alpha(x_1)) + \Delta \left(\tilde{y}(r) - \frac{\Sigma_{12}}{\Sigma_{11}}r - \alpha(x_1) \right),$$

where, by positivity of Gaussian densities,

$$\Delta := F'_{\eta_2}(\alpha(x_1)) = g_{\eta_2}(\alpha(x_1)) > 0.$$

This allows us, along with (43) and (48), to proceed lead (49) to the contradiction

$$\begin{aligned} 0 &\le \Delta \int_{-\infty}^{x_1} g_{\eta_1}(r) \left(\tilde{y}(r) - \frac{\Sigma_{12}}{\Sigma_{11}}r - \alpha(x_1) \right) dr = \Delta \int_{-\infty}^{x_1} g_{\eta_1}(r) \left[\tilde{y}(r) - y(r) \right] dr \\ &= \Delta \left[\mathbb{E}\tilde{y}(\xi_1)\, \chi_{(-\infty,x_1]}(\xi_1) - \mathbb{E}y(\xi_1)\, \chi_{(-\infty,x_1]}(\xi_1) \right] < 0. \end{aligned}$$

This proves our initial claim that y in (43) is a global solution to (44). Accordingly, for each $x_1 \in \mathbb{R}$ satisfying (42), we have that

$$\min_{x_2 \in \mathcal{X}^*} \left\{ c_2 \mathbb{E}x_2(\xi_1)\, \chi_{(-\infty,x_1]}(\xi_1) \,|\, \varphi(x_1,x_2) \ge p;\ x_2 \in M \right\}$$

$$= c_2 \mathbb{E} y\left(\xi_1\right) \chi_{(-\infty, x_1]}\left(\xi_1\right) = c_2 \mathbb{E}\left[\frac{\Sigma_{12}}{\Sigma_{11}} \xi_1 + \alpha\left(x_1\right)\right] \chi_{(-\infty, x_1]}\left(\xi_1\right)$$
$$= c_2\left[\frac{\Sigma_{12}}{\Sigma_{11}} \mathbb{E} \xi_1|_{\xi_1 \leq x_1} + \alpha\left(x_1\right) \mathbb{E} \chi_{(-\infty, x_1]}\left(\xi_1\right)\right].$$

It is well-known that for the Gaussian random variable ξ_1 with mean μ_1 and standard deviation $\sqrt{\Sigma_{11}}$ the mean conditioned to $\xi_1 \leq x_1$ calculates as $\mathbb{E} \xi_1|_{\xi_1 \leq x_1} = \beta\left(x_1\right)$, where β is defined in (41). Since, moreover, by (46),

$$\mathbb{E} \chi_{(-\infty, x_1]}\left(\xi_1\right) = \mathbb{P}\left(\xi_1 \leq x_1\right) = \mathbb{P}\left(\eta_1 \leq x_1\right) = \Phi\left(\frac{x_1 - \mu_1}{\sqrt{\Sigma_{11}}}\right),$$

we may conclude that, with the function f as introduced in the statement of this Theorem,

$$\min_{x_2 \in \mathcal{X}^*}\left\{c_2 \mathbb{E} x_2\left(\xi_1\right) \chi_{(-\infty, x_1]}\left(\xi_1\right) | \varphi\left(x_1, x_2\right) \geq p;\ x_2 \in M\right\} = c_2 f\left(x_1\right). \tag{50}$$

Now, finally, we turn to our given problem (38) and decompose it pretty much the same way we did in (35):

$$\min_{x_1 \in \mathbb{R}}\left\{c_1 x_1 + \min_{x_2 \in \mathcal{X}^*}\left\{c_2 \mathbb{E} x_2\left(\xi_1\right) \chi_{(-\infty, x_1]}\left(\xi_1\right) | \varphi\left(x_1, x_2\right) \geq p;\ x_2 \in M\right\}\right\}.$$

Observe first that for arbitrary x_1 with $x_1 \leq \sqrt{\Sigma_{11}} \Phi^{-1}(p) + \mu_1$ the feasible set of the inner problem above is empty. Indeed, if there existed some $x_2 \in \mathcal{X}^*$ with $\varphi\left(x_1, x_2\right) \geq p$, then, repeating an earlier argumentation on top of (49), we could establish the contradiction

$$p \leq \varphi\left(x_1, x_2\right) = \mathbb{P}\left(\xi_1 \leq x_1, \xi_2 \leq x_2\left(\xi_1\right)\right) = \int_{-\infty}^{x_1} g_{\eta_1}(r) F_{\eta_2}\left(\tilde{y}(r) - \frac{\Sigma_{12}}{\Sigma_{11}} r\right) dr$$
$$< \int_{-\infty}^{x_1} g_{\eta_1}(r)\, dr = \mathbb{P}\left(\eta_1 \leq x_1\right) = \mathbb{P}\left(\xi_1 \leq x_1\right)$$
$$= \mathbb{P}\left(\frac{\xi_1 - \mu_1}{\sqrt{\Sigma_{11}}} \leq \frac{x_1 - \mu_1}{\sqrt{\Sigma_{11}}}\right) = \Phi\left(\frac{x_1 - \mu_1}{\sqrt{\Sigma_{11}}}\right) \leq \Phi\left(\Phi^{-1}(p)\right) = p.$$

Here, the strict inequality follows from the fact that $g_{\eta_1} > 0$ and $F_{\eta_2} < 1$, while the last inequality is a consequence of Φ being nondecreasing. Hence, for $x_1 \leq \sqrt{\Sigma_{11}} \Phi^{-1}(p) + \mu_1$, the minimum value of the objective over the empty feasible set is equal to infinity. Therefore, such x_1 can be ignored in the outer minimization and we can write our problem, thanks to (50) as

$$\min_{x_1 > \sqrt{\Sigma_{11}} \Phi^{-1}(p) + \mu_1}\left\{c_1 x_1 + \min_{x_2 \in \mathcal{X}^*}\left\{c_2 \mathbb{E} x_2\left(\xi_1\right) \chi_{(-\infty, x_1]}\left(\xi_1\right) | \varphi\left(x_1, x_2\right) \geq p;\ x_2 \in M\right\}\right\}$$
$$= \min_{x_1 > \sqrt{\Sigma_{11}} \Phi^{-1}(p) + \mu_1}\left\{c_1 x_1 + c_2 f\left(x_1\right)\right\}.$$

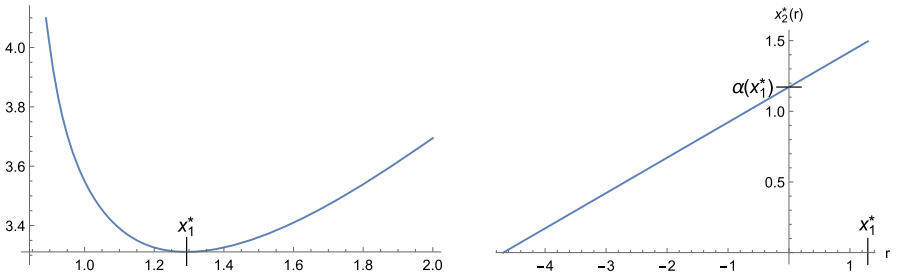

Fig. 3 Illustration of a solution to problem 38: Optimal first stage decision x_1^* as minimizer of the function $c_1 t + c_2 f(t)$ (left) and optimal second stage decision $x_2^*(r)$ as affine linear function with slope and intercept as indicated in Theorem 4 (right)

This proves our assertion on an optimal solution x_1^*. As shown in the first part in this proof, the optimal second-stage decision in (44) associated with the first-stage decision x_1^* is defined in (43) and yields the asserted formula for x_2^* in the statement of this proof. □

Figure 3 illustrates the solution of problem (38) for the data

$$c_1 = 1;\ c_2 = 2;\ p = 0.8;\ \mu = (0,0);\ \Sigma = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$$

Acknowledgements The authors would like to express their gratitude to an anonymous referee for his comments motivating them to prove the result of Theorem 4. The first author would like to thank the MATH+ Berlin Mathematical graduate school for their support and the Stiftung der Deutsche Wirtschaft for their funding. The second author gratefully acknowledges support by DFG in the Collaborative Research Centre CRC/Transregio 154, Project B04 and by the FMJH Program Gaspard Monge in optimization and operations research including support to this program by EDF. The third author was partially supported by Grants: Fondecyt Regular 1190110, Fondecyt Regular 1200283 and Programa Regional Mathamsud 20-Math-08 CODE: MATH190003.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Adams, R.A., Fournier, J.F.: Sobolev Spaces, volume 140 of Pure and Applied Mathematics (Amsterdam), 2nd edn. Elsevier/Academic Press, Amsterdam (2003)
2. Alais, J.C., Carpentier, P., De Lara, M.: Multi-usage hydropower single dam management: chance-constrained optimization and stochastic viability. Energy Syst. **8**, 7–30 (2017)

3. Aliprantis, C.D., Border, K.C.: Infinite Dimensional Analysis, 3rd edn. Springer, Berlin (2006). A hitchhiker's guide
4. Andrieu, L., Henrion, R., Römisch, W.: A model for dynamic chance constraints in hydro power reservoir management. Eur. J. Oper. Res. **207**, 579–589 (2010)
5. Charnes, A., Cooper, W.W., Symonds, G.H.: Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. Manag. Sci. **4**, 235–263 (1958)
6. Farshbaf-Shaker, M.H., Henrion, R., Hömberg, D.: Properties of chance constraints in infinite dimensions with an application to PDE constrained optimization. Set-Valued Var. Anal. **26**, 821–841 (2018)
7. Farshbaf-Shaker, M.H., Gugat, M., Heitsch, H., Henrion, R.: Optimal Neumann boundary control of a vibrating string with uncertain initial data and probabilistic terminal constraints. SIAM J. Control Optim. **58**, 2288–2311 (2020)
8. Gassmann, H.I., Prékopa, A.: On stages and consistency checks in stochastic programming. Oper. Res. Lett. **33**, 171–175 (2005)
9. Geletu, A., Hoffmann, A., Schmidt, P., Li, P.: Chance constrained optimization of elliptic PDE systems with a smoothing convex approximation. ESAIM Control Optim. Calc. Var. (2020). https://doi.org/10.1051/cocv/2019077
10. Guigues, V., Henrion, R.: Joint dynamic probabilistic constraints with projected linear decision rules. Optim. Methods Softw. **32**, 1006–1032 (2017)
11. Liu, X., Kücükyavuz, S., Luedtke, J.: Decomposition algorithms for two-stage chance-constrained programs. Math. Program. **157**, 219–243 (2016)
12. Luedtke, J., Ahmed, S.: A sample approximation approach for optimization with probabilistic constraints. SIAM J. Optim. **19**, 674–699 (2008)
13. Pagnoncelli, B., Ahmed, S., Shapiro, A.: Sample average approximation method for chance constrained programming: theory and applications. J. Optim. Theory Appl. **142**, 399–416 (2009)
14. Prékopa, A.: Contributions to the theory of stochastic programming. Math. Program. **4**, 202–221 (1973)
15. Prékopa, A.: Stochastic Programming. Kluwer, Dordrecht (1995)
16. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming, volume 9 of MOS-SIAM Series on Optimization, 2nd edn. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (2014)
17. van Ackooij, W., Henrion, R., Pérez-Aros, P.: Generalized gradients for probabilistic/robust (probust) constraints. Optimization **69**, 1451–1479 (2020)

Affiliations

T. González Grandón[1] · R. Henrion[2] · P. Pérez-Aros[3]

T. González Grandón
grandont@hu-berlin.de

P. Pérez-Aros
pedro.perez@uoh.cl

1 Humboldt University Berlin, Spandauer Straße 1, 10178 Berlin, Germany

2 Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstraße 39, 10117 Berlin, Germany

3 Instituto de Ciencias de la Ingenieria Universidad de O'Higgins, Libertador Bernardo O'Higgings 611, Rancagua, Chile