**ORIGINAL PAPER** 



# Buffered and Reduced Multidimensional Distribution Functions and Their Application in Optimization

Bogdan Grechuk<sup>1</sup> · Michael Zabarankin<sup>2</sup> · Alexander Mafusalov<sup>3</sup> · Stan Uryasev<sup>4</sup>

Received: 7 October 2022 / Accepted: 9 July 2023 / Published online: 9 October 2023 © The Author(s) 2023

# Abstract

For a random variable, superdistribution has emerged as a valuable probability concept. Similar to cumulative distribution function (CDF), it uniquely defines the random variable and can be evaluated with a simple one-dimensional minimization formula. This work leverages the structure of that formula to introduce buffered CDF (bCDF) and reduced CDF (rCDF) for random vectors. bCDF and rCDF are shown to be the minimal Schur-convex upper bound and the maximal Schur-concave lower bound of the multivariate CDF, respectively. Special structure of bCDF and rCDF is used to construct an algorithm for solving optimization problems with bCDF and rCDF in objective or constraints. The efficiency of the algorithm is demonstrated in a case study on optimization of a collateralized debt obligation with bCDF functions in constraints.

Keywords Probability of exceedance  $\cdot$  Buffered probability of exceedance  $\cdot$  BPOE  $\cdot$  Conditional value-at-risk  $\cdot$  Cumulative distribution function  $\cdot$  CDF

# 1 Introduction

The cumulative distribution function (CDF) of a random variable (r.v.) is a fundamental notion in probability theory and plays a central role in stochastic optimization, risk management, statistics, reliability theory and various

Bogdan Grechuk bg83@leicester.ac.uk

<sup>&</sup>lt;sup>1</sup> School of Computing and Mathematic Sciences, University of Leicester, Leicester LE1 7RH, UK

<sup>&</sup>lt;sup>2</sup> Departement of Mathematical Sciences, Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ, USA

<sup>&</sup>lt;sup>3</sup> Independent researcher, Memphis, TN, USA

<sup>&</sup>lt;sup>4</sup> Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, USA

applications. For instance, engineered systems, e.g., electrical grids and gas/ oil pipelines, must be designed to comply with various safety and reliability regulations formulated in terms of the probability of failure calculated with CDFs. Also, smart electrical grids should be robust to rare events including power station faults and electromagnetic pulses, whereas gas pipeline operations should meet the consumer demand while being resilient to unforeseen production disruptions. Design of engineered structures and industrial systems involve multiple uncertain characteristics that are difficult to combine into a single loss function. For example, each watertight compartment of a submarine should be designed to maximize its survivability in case of an accident. Suppose that a submarine has eight compartments. Chances of failure of all compartments because of high outside pressure can be described by an eightdimensional CDF accounting for failures of individual compartments (which are not independent because of common-cause factors: similar compartment design, outside water pressure, etc.). Another example is to ensure an uninterrupted power supply from a system of redundant emergency generators in case of a natural disaster. All generators may be impacted by the same factors and failures of individual generators are not independent. Yet another example is to evaluate the probability that several escape roads will be blocked simultaneously because of common factors such as a hurricane. In all of these examples, the probability of an event should be assessed for a random vector rather than an r.v. From optimization perspective, one of the technical difficulties that engineers face when using a multivariate CDF is that for any fixed  $\mathbf{x} \in \mathbb{R}^n$ , the CDF  $F_{\mathbf{x}}(\mathbf{x})$  of a random vector **X** is not a convex function of **X**. Moreover, CDFs based on observations (i.e., discrete distributions) are discontinuous piecewise constant functions of  $\mathbf{x}$ . In this case, stochastic optimization problems result in mixed-integer optimization problems, which are notoriously hard to solve for a large number of scenarios. Also, the use of CDF in decision problems has two conceptual shortcomings. First one is that at any fixed point x, the CDF of an r.v. X does not "capture" the extent of X below x, even if a tail has a small weight, the average value of X in the tail could be quite significant. As a result, reliance just on CDF may yield solutions with undesirable tail outcomes occurring with small probabilities. The other conceptual shortcoming is that in general, a safety first principle based on the CDF of X does not agree with a common perception that aggregation reduces volatility.<sup>1</sup> This perception can be "captured" by the notion of *convex order*. Informally, an r.v. X dominates an r.v. Y in convex order, and we write  $X \ge_{cx} Y$ , if the distribution of Y can be obtained from the distribution of X by aggregation. A formal definition is that  $X \ge_{cx} Y$  if  $\mathbb{E}[g(X)] \ge \mathbb{E}[g(Y)]$  for any convex function  $g : \mathbb{R} \to \mathbb{R}$ .  $X \ge_{cx} Y$  implies that  $\mathbb{E}[X] = \mathbb{E}[Y]$  and  $\sigma(X) \ge \sigma(Y)$ , where  $\sigma(X)$  is standard deviation,—this supports the perception that Y, being aggregated from

<sup>&</sup>lt;sup>1</sup> Here, "safety first principle based on the CDF" means that for a given threshold  $x \in \mathbb{R}$ , an r.v. *X* is preferred to an r.v. *Y*, if and only if  $F_X(x) = \mathbb{P}[X \le x]$  does not exceed  $F_Y(x) = \mathbb{P}[Y \le x]$ . For example, let  $\mathbb{P}[X = -1] = 1/2$  and  $\mathbb{P}[X = 1] = 1/2$ . Then for x = 0,  $\mathbb{P}[X \le 0] = \mathbb{P}[-X \le 0] = 1/2$ , whereas  $\mathbb{P}[(X + (-X))/2 \le 0] = \mathbb{P}[0 \le 0] = 1 > 1/2$ . Consequently, both *X* and -X are (strictly) preferred to (X + (-X))/2, contrary to the perception that aggregation reduces risk.

*X* and having same expected value as *X*, is less volatile than *X*. A functional *f* is called *Schur-convex* if  $X \ge_{cx} Y$  implies that  $f(X) \ge f(Y)$ . In other words, a Schur-convex functional does not increase after data aggregation. Examples of such functionals include  $f(X) = \sigma(X)$  and  $f(X) = -\mathbb{E}[u(X)]$  for a concave function *u*. However,  $f(X) = F_X(x) = \mathbb{P}[X \le x]$  with fixed *x* is not Schur-convex. As a result, if  $X^* \in \arg\min_{X \in \mathcal{X}} F_X(x)$  for a feasible set  $\mathcal{X}$  and  $x \in \mathbb{R}$ , there can exist  $Y \in \mathcal{X}$  such that  $X^* \ge_{cx} Y$  and  $F_{X^*}(x) < F_Y(x)$ . In this case,  $\mathbb{E}[Y] = \mathbb{E}[X^*]$  and  $\sigma(Y) \le \sigma(X^*)$ , but *Y* is "less safe" than  $X^*$ .

In the one dimensional case, i.e., for an r.v. *X*, all these CDF deficiencies were addressed with the functions of *superquantile* and *buffered probability of exceedance* (*bPOE*) [6]. Superquantile, which is also known as *expected shortfall* and *conditional value-at-risk* (*CVaR*) [14, 15] of *X* with confidence level  $\alpha$ , is the expectation of the right  $(1 - \alpha)$ —tail of the distribution of *X*, i.e., it is the average of the largest outcomes with total probability  $1 - \alpha$ . Superquantile can be defined as follows

$$\overline{q}_X(\alpha) = \frac{1}{1-\alpha} \int_{\alpha}^{1} q_X(s) ds \equiv \min_{c \in \mathbb{R}} \left( c + \frac{1}{1-\alpha} \mathbb{E}[X-c]^+ \right), \qquad \alpha \in [0,1),$$

where  $q_X(s) = \inf\{x \in \mathbb{R} \mid F_X(x) > s\}$  is the quantile function and  $z^+$  denotes  $\max\{0, z\}$ . bPOE is an extension of the *buffered probability of failure* [13] and is equal to 1 minus the inverse of superquantile, see [6],

$$\overline{p}_X(x) = 1 - \overline{q}_X^{-1}(x)$$
 for  $x \neq \sup X$ .

It can also be evaluated by

$$\overline{p}_X(x) = \min_{a \ge 0} \mathbb{E}[a(X - x) + 1]^+ \text{ for } x \neq \sup X,$$
(1)

see [6, 7]. Considered as a function of r.v. *X* for fixed *x*, bPOE is the minimal quasiconcave upper bound (see [6]) of *probability of exceedance (POE)*,

$$p_X(x) = \mathbb{P}[X \ge x]$$
 for  $x \ne \sup X$ .

Superquantile and bPOE constraints are equivalent, see [6]:

$$\overline{p}_X(x) \leq 1 - \alpha \quad \Longleftrightarrow \quad \overline{q}_X(\alpha) \leq x \text{ for } x \neq \sup X, \ \alpha \in [0, 1).$$
 (2)

The aim of this work is to construct extensions of multivariate CDF, defined for the lower tail rather than the upper tail. The analogue of superquantile for the lower tail will be called subquantile:

$$\underline{q}_{X}(\alpha) = \frac{1}{\alpha} \int_{0}^{\alpha} q_{X}(s) ds \equiv \max_{c \in \mathbb{R}} \left( c - \frac{1}{\alpha} \mathbb{E}[c - X]^{+} \right), \quad \alpha \in (0, 1].$$
(3)

To define an analog of bPOE for the lower tail, let an r.v. X with the CDF  $F_X : \mathbb{R} \to [0, 1]$  represent a reward of some kind, e.g., money, energy, satisfied demand, etc. Then for a threshold  $\alpha \in (0, 1]$ , *buffered CDF* (*bCDF*) is defined as an inverse of subquantile, i.e.,  $\overline{F}_X(x) = \alpha$ , where  $\alpha$  is a solution of the equation  $x = q_v(\alpha)$ . In other words,

$$\overline{F}_X(x) = \alpha \quad \iff \quad x = \underline{q}_X(\alpha) \text{ for } x \neq \inf X, \ \alpha \in (0, 1].$$
 (4)

For every fixed x, bCDF  $\overline{F}_X(x)$  is a quasi-convex function of r.v. X and is the minimal quasi-convex upper bound of  $F_X(x)$ . This follows from the fact that bPOE is the minimal quasi-convex upper bound of POE [6]; see also [18, Eq's (3.2.6)–(3.2.8)] and [2, Eq (30)]. bCDF admits a representation similar to (1):

$$\overline{F}_{X}(x) = \overline{p}_{-X}(-x) = \min_{a \ge 0} \mathbb{E}[a(x-X)+1]^{+}.$$
(5)

Similar to the constraint equivalence (2) for the upper tail, (4) yields constraint equivalence for the lower tail:

$$\overline{F}_X(x) \leq \alpha \quad \Longleftrightarrow \quad \underline{q}_X(\alpha) \geq x$$
$$\iff \quad \max_{c \in \mathbb{R}} \left( c - \frac{1}{\alpha} \mathbb{E}[c - X]^+ \right) \geq x \text{ for } x \neq \inf X, \ \alpha \in (0, 1].$$

Under certain conditions, the maximum in the last constraint can be dropped, and the constraint takes the form  $c - \alpha^{-1} \mathbb{E}[c - X]^+ \ge x$ , which can be reformulated as linear constraints for discrete distributions. However, the main advantage of bCDF compared to CDF is that bCDF describes the distribution tails. For instance, when r.v. X represents a portfolio value and depends on the portfolio weights (decision variables), minimization of  $\overline{F}_X(x)$  yields an optimal solution  $X^*$  with  $\overline{F}_{X^*}(x) = \alpha^*$ such that the *average value* of  $X^*$  in  $\alpha^* \cdot 100\%$  of the worst outcomes is at least x. Under some mild conditions, optimization of one-dimensional bCDF can be reduced to convex and linear programming (for discrete distributions). For example, let  $\mathbf{X} = (X_1, \dots, X_m)$  be a known random vector,  $x \in \mathbb{R}$  a given threshold,  $\mathbf{w} = (w_1, \dots, w_m) \in \mathbb{R}^m$  an unknown vector, and  $X = \mathbf{w}^T \mathbf{X} = w_1 X_1 + \dots + w_m X_m$ . Then, the problem of minimizing  $\overline{F}_X(x)$  with respect to  $\mathbf{w}$  can be formulated with (5) as

$$\min_{\mathbf{w}} \overline{F}_X(x) = \min_{\mathbf{w}} \min_{a \ge 0} \mathbb{E}[a(x - \mathbf{w}^{\mathsf{T}} \mathbf{X}) + 1]^+ = \min_{\mathbf{c}, a \ge 0} \mathbb{E}[ax - \mathbf{c}^{\mathsf{T}} \mathbf{X} + 1]^+$$

where  $\mathbf{c} = a \mathbf{w}$  is a new decision vector. The last optimization problem can be reduced to convex and linear programming for a discrete distribution as in [6]. This also follows from a simple observation that

$$F_X(x) = \overline{p}_{x - \mathbf{w}^{\mathsf{T}} \mathbf{X}}(0).$$
(6)

Rockafellar and Royset [12] introduced *super CDF* (*sCDF*) as the inverse of superquantile, i.e.,  $\underline{F}_{\mathbf{X}}(\mathbf{x}) = \overline{q}_{\mathbf{X}}^{-1}(x) = 1 - \overline{p}_{\mathbf{X}}(x)$ , which can be shown to be the maximal quasi-concave lower bound of CDF. It follows from (1) that

$$\underline{F}_{X}(x) = 1 - \overline{p}_{X}(x) = 1 - \min_{a \ge 0} \mathbb{E}[a(X - x) + 1]^{+} = \max_{a \ge 0} \mathbb{E}[\min\{a(x - X), 1\}].$$
 (7)

We will call super CDF by *reduced CDF* (*rCDF*) to emphasize that this function is a lower bound of CDF, versus buffered CDF which is an upper bound of CDF.

It follows from (5) and (7) that bCDF and rCDF are related by

$$\underline{F}_{X}(x) = 1 - \overline{F}_{-X}(-x), \qquad \overline{F}_{X}(x) = 1 - \underline{F}_{-X}(-x).$$
(8)

Generalizing bCDF and rCDF for random vectors is not straightforward. The difficulty is that in the multivariate case, the notions of quantile and its inverse are not well-defined. A potential insight can be offered through exploring the properties of (5) and (7). This work generalizes bCDF and rCDF given by (5) and (7) to random vectors. In contrast to the one-dimensional case, the multivariate bCDF and rCDF are not quasi-convex and quasi-concave functions of a random vector. We show that they are the minimal Schur-convex upper bound and the maximal Schur-concave lower bound of the multivariate CDF, respectively, and that their special structure allows us constructing an algorithm, which can quickly find their local extrema. We demonstrate efficiency of the algorithm in a case study on optimization of a collateralized debt obligation with bCDF functions in constraints.

The paper is organized into five sections and five appendices. Section 2 introduces bCDF and rCDF and shows that they are upper and lower bounds for a multivariate CDF. Section 3 considers bCDF and rCDF in optimization problems. Section 4 presents applications of bCDF and rCDF. Section 5 discusses the case study. Appendices A, B and C present proofs of three propositions and Appendices D and E summarize algorithms for solving optimization problems with bCDF and rCDF in objective function and constraints.

## 2 The lower and upper bounds for multivariate CDF

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, where  $\Omega$  is an arbitrary non-empty set,  $\mathcal{F}$  is the  $\sigma$ -algebra of subsets of  $\Omega$ , and  $\mathbb{P}$  is a probability measure on  $(\Omega, \mathcal{F})$ . Sets in  $\mathcal{F}$  are called events. A probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  is called atomless if there exists an r.v. on it with a continuous CDF.

An *n*-dimensional random vector  $\mathbf{X} = (X_1, ..., X_n)$  is a function  $\mathbf{X} : \Omega \to \mathbb{R}^n$  such that for every  $\mathbf{x} \in \mathbb{R}^n$  set  $\{\omega \in \Omega \mid \mathbf{X}(\omega) \leq \mathbf{x}\}$  is an event,<sup>2</sup> and  $X_1, ..., X_n$  are called r.v.'s. On an atomless probability space, there exists a collection of random vectors with any given joint distribution. Let  $L^{1,n}(\Omega)$  and  $L^{\infty,n}(\Omega)$  denote the sets of *n*-dimensional random vectors  $\mathbf{X}$  for which  $\mathbb{E}[\mathbf{X}] = \int_{\Omega} \mathbf{X}(\omega) d\mathbb{P}$  and  $\max_i \sup |X_i|$  exist and are finite, respectively.

**Definition 1** A random vector  $\mathbf{X} \in L^{1,n}(\Omega)$  dominates random vector  $\mathbf{Y} \in L^{1,n}(\Omega)$  in convex order, and we write  $\mathbf{X} \ge_{cx} \mathbf{Y}$ , if  $\mathbb{E}[g(\mathbf{X})] \ge \mathbb{E}[g(\mathbf{Y})]$  for any convex function  $g : \mathbb{R}^n \to \mathbb{R}$ .

The convexity of *g* implies existence of  $\mathbf{a} \in \mathbb{R}^n$  and  $b \in \mathbb{R}$  such that  $g(x) \ge \mathbf{a}^T x + b$  for all  $\mathbf{x} \in \mathbb{R}^n$ , so that  $\mathbb{E}[g(\mathbf{X})] \ge \mathbb{E}[\mathbf{a}^T \mathbf{X} + b] = \mathbf{a}^T \mathbb{E}[\mathbf{X}] + b > -\infty$  is well-defined for every  $\mathbf{X} \in L^{1,n}(\Omega)$ , although it can be  $+\infty$ .

<sup>&</sup>lt;sup>2</sup> Inequalities between vectors hold component-wise.

**Definition 2** A function  $f : L^{1,n}(\Omega) \to \mathbb{R}$  is Schur-convex if  $\mathbf{X} \ge_{cx} \mathbf{Y}$  implies  $f(\mathbf{X}) \ge f(\mathbf{Y})$ . A function *f* is Schur-concave if -f is Schur-convex.

Definition 2 implies that any function in the form  $f(\mathbf{X}) = \mathbb{E}[g(\mathbf{X})]$  is Schur-convex and Schur-concave if g is convex and concave, respectively. Any function f has the unique minimal Schur-convex upper bound given by

$$f_u(\mathbf{X}) = \sup_{\mathbf{Y} \leqslant_{cx} \mathbf{X}} f(\mathbf{Y}),$$

see [6], as well as the the unique maximal Schur-concave lower bound

$$f_l(\mathbf{X}) = \inf_{\mathbf{Y} \leqslant_{cx} \mathbf{X}} f(\mathbf{Y}).$$

Let

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}[\mathbf{X} \leqslant \mathbf{x}]$$

denote multivariate CDF of X.

**Definition 3** For any  $\mathbf{X} \in L^{1,n}(\Omega)$ , *buffered CDF (bCDF)* and *reduced CDF (rCDF)* are defined by

$$\overline{F}_{\mathbf{X}}(\mathbf{x}) = \inf_{\mathbf{a} \in \mathbb{R}^n_+} \mathbb{E} \left[ \mathbf{a}^\top (\mathbf{x} - \mathbf{X}) + 1 \right]^+, \tag{9}$$

$$\underline{F}_{\mathbf{X}}(\mathbf{x}) = \sup_{\mathbf{a} \in \mathbb{R}^n_+} \mathbb{E}[\min\{1, a_1(x_1 - X_1), \dots, a_n(x_n - X_n)\}],\tag{10}$$

respectively, where  $\mathbf{a} = (a_1, \dots, a_n), \mathbf{a}^\mathsf{T} \mathbf{X} = \sum_{i=1}^n a_i X_i$ , and  $\mathbb{R}^n_+ = \{ \mathbf{a} \in \mathbb{R}^n, \mathbf{a} \ge 0 \}.$ 

**Remark** With an auxiliary variable  $a_0 \ge 0$  and the following set (simplex)

$$\mathbf{A} = \left\{ \mathbf{a} \in \mathbb{R}^{n}_{+} \, \middle| \, \sum_{i=1}^{n} a_{i} = n \right\},\tag{11}$$

bCDF and rCDF, defined by (9) and (10), respectively, can be recast as

$$\overline{F}_{\mathbf{X}}(\mathbf{x}) = \inf_{\mathbf{a} \in \mathbf{A}, \ a_0 \in \mathbb{R}_+} \mathbb{E} \left[ a_0 \ \mathbf{a}^\top (\mathbf{x} - \mathbf{X}) + 1 \right]^+ = \inf_{\mathbf{a} \in \mathbf{A}} \overline{p}_{\mathbf{a}^\top (\mathbf{x} - \mathbf{X})}(0), \tag{12}$$

$$\underline{F}_{\mathbf{X}}(\mathbf{x}) = 1 - \inf_{\mathbf{a} \in \mathbb{R}^{n}_{+}} \mathbb{E} \Big[ 1 + \max\{a_{1}(X_{1} - x_{1}), \dots, a_{n}(X_{n} - x_{n})\} \Big]^{+} \\ = 1 - \inf_{a_{0} \ge 0} \inf_{\mathbf{a} \in \mathbf{A}} \mathbb{E} \Big[ 1 + a_{0} \max\{a_{1}(X_{1} - x_{1}), \dots, a_{n}(X_{n} - x_{n})\} \Big]^{+}$$
(13)
$$= 1 - \inf_{\mathbf{a} \in \mathbf{A}} \overline{p}_{\max\{a_{1}(X_{1} - x_{1}), \dots, a_{n}(X_{n} - x_{n})\}} (0),$$

provided that infimums in (12) and (13) with respect to  $a_0 \ge 0$  are attained. Representations (12) and (13) show that bCDF and rCDF of a random vector **X** are, in

fact, bPOE of corresponding r.v.'s and minimized with respect to the vector **a** over the simplex **A**. These representations are important in devising efficient algorithms for solving optimization problems with bCDF and rCDF. If **X** depends on decision variables, optimal decision problems involving bCDF and rCDF in either objective function or constraints are *bilevel optimization*, see Sect. 3.

First we show that bCDF and rCDF are the upper and lower bounds for CDF, respectively.

**Proposition 2.1** *For any*  $\mathbf{X} \in L^{1,n}(\Omega)$  *and any*  $\mathbf{x} \in \mathbb{R}^n$ *,* 

$$\underline{F}_{\mathbf{X}}(\mathbf{x}) \leqslant \mathbb{P}[\mathbf{X} < \mathbf{x}] \leqslant \mathbb{P}[\mathbf{X} \leqslant \mathbf{x}] = F_{\mathbf{X}}(\mathbf{x}) \leqslant F_{\mathbf{X}}(\mathbf{x}).$$
(14)

**Proof** Let  $\mathbf{a} \in \mathbb{R}^n_+$  be arbitrary, and let

$$A = \{ w \in \Omega : \mathbf{X}(\omega) < \mathbf{x} \}$$

be the event that  $\mathbf{X}(\omega) < \mathbf{x}$ . Then for any  $\omega \notin A$  we have  $X_i(\omega) \ge x_i$  for some *i*, so that

$$\min\{1, a_1(x_1 - X_1(\omega)), \dots, a_n(x_n - X_n(\omega))\} \leq a_i(x_i - X_i(\omega)) \leq 0 = I_A(\omega),$$

where  $I_A$  is the indicator indicator function<sup>3</sup> of the event A. Also, for  $\omega \in A$ ,

$$\min\{1, a_1(x_1 - X_1(\omega)), \dots, a_n(x_n - X_n(\omega))\} \le 1 = I_A(\omega),$$

and consequently,

$$\mathbb{E}[\min\{1, a_1(x_1 - X_1), \dots, a_n(x_n - X_n)\}] \leq \mathbb{E}[I_A] = \mathbb{P}[A] = \mathbb{P}[\mathbf{X} < \mathbf{x}].$$

Since  $a \in \mathbb{R}^n_+$  is arbitrary, the last inequality and (10) yield the first inequality in (14). Similarly, if

 $B = \{ w \in \Omega : \mathbf{X}(\omega) \leq \mathbf{x} \}$ 

is the event that  $\mathbf{X}(\omega) \leq \mathbf{x}$ , then for any  $\omega \in B$ ,

$$[\mathbf{a}^{\top}(\mathbf{x} - \mathbf{X}(\omega)) + 1]^{+} \ge [0 + 1]^{+} = 1 = I_{B}(\omega),$$

and for any  $\omega \notin B$ ,

$$[\mathbf{a}^{\top}(\mathbf{x} - \mathbf{X}(\omega)) + 1]^+ \ge 0 = I_B(\omega),$$

so that

$$\mathbb{E}[\mathbf{a}^{\mathsf{T}}(\mathbf{x} - \mathbf{X}) + 1]^{+} \ge E[I_{B}] = \mathbb{P}[B] = \mathbb{P}[\mathbf{X} \le \mathbf{x}] = F_{\mathbf{X}}(\mathbf{x}),$$

which along with (9) yields the last inequality in (14).

<sup>&</sup>lt;sup>3</sup> As usual, the indicator function of the event A is the random variable  $I_A$  such that  $I_A(w) = 1$  for  $w \in A$  and  $I_A(w) = 0$  for  $w \notin A$ .

It will be shown later that bCDF and rCDF are the minimal Schur-convex upper bounds of  $F_{\mathbf{X}}(\mathbf{x})$  and the maximal Schur-concave lower bounds of  $\mathbb{P}[\mathbf{X} < \mathbf{x}]$ , respectively.

For an r.v. *X*,

$$\overline{F}_X(x) = \overline{p}_{-X}(-x)$$
 and  $\underline{F}_X(x) = 1 - \overline{p}_X(x)$ ,

where  $\overline{p}_X(x)$  is the (one-dimensional) bPOE defined in (1).

Also, note that

$$1 - \underline{F}_{\mathbf{X}}(\mathbf{x}) = 1 - \sup_{\mathbf{a} \in \mathbb{R}^n_+} \mathbb{E}[\min\{1, a_1(x_1 - X_1), \dots, a_n(x_n - X_n)\}]$$
(15a)

$$= \inf_{\mathbf{a} \in \mathbb{R}^{n}_{+}} \mathbb{E}[\max\{a_{1}(X_{1} - x_{1}), \dots, a_{n}(X_{n} - x_{n})\} + 1]^{+}$$
(15b)

$$\leq \inf_{a_0 \ge 0} \mathbb{E}[a_0 \max\{X_1 - x_1, \dots, X_n - x_n\} + 1]^+$$
(15c)

$$= \overline{p}_{\max\{X_1 - x_1, \dots, X_n - x_n\}}(0).$$
(15d)

Note that  $F_{\mathbf{X}}(\mathbf{x})$  can be represented by

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}[\mathbf{X} \leq \mathbf{x}] = \sup_{W \in \{0,1\}} \mathbb{E}[W] \quad \text{subject to} \quad X_i W \leq x_i W, \quad i = 1, \dots, n,$$
(16)

where " $W \in \{0, 1\}$ " indicates that the supremum is over all random variables such that  $W(\omega) \in \{0, 1\}$  for every  $\omega \in \Omega$ , while the constraint  $X_i W \leq x_i W$  means that  $X_i(\omega)W(\omega) \leq x_i W(\omega)$  for all  $\omega \in \Omega$ . The supremum in (16) is attained when *W* is the indicator function of the event  $\mathbf{X} \leq \mathbf{x}$ . Similarly, bCDF has the following interpretation through dual characterization.

**Proposition 2.2** *For any*  $\mathbf{X} \in L^{1,n}(\Omega)$  *and any*  $\mathbf{x} \in \mathbb{R}^{n}$ *,* 

$$\overline{F}_{\mathbf{X}}(\mathbf{x}) = \sup_{W \in L^{1}(\Omega)} \mathbb{E}[W] \text{ subject to } \mathbb{E}[X_{i}W] \leq x_{i}\mathbb{E}[W], \quad i = 1, \dots, n, \quad W \in [0, 1],$$

where  $W \in [0, 1]$  means that  $0 \leq W(\omega) \leq 1$  for all  $\omega \in \Omega$ .

**Proof** See Appendix A.

**Proposition 2.3** If the underlying probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  is atomless, then  $\overline{F}_{\mathbf{X}}(\mathbf{x})$  is the minimal Schur-convex upper bound for  $F_{\mathbf{X}}(\mathbf{x})$ .

**Proof** See Appendix **B**.

When n = 1,  $\overline{F}_X(x)$  is also a unique minimal quasi-convex upper bound for  $F_X(x)$ , see [6]. This is not true when  $n \ge 2$ .

**Proposition 2.4** In general (when  $n \ge 2$ ),  $\overline{F}_{\mathbf{X}}(\mathbf{x})$  is not quasi-convex in  $\mathbf{X}$ .

**Proof** Let  $n \ge 2$ ,  $\mathbf{0} = (0, 0, \dots, 0)$ ,  $\mathbf{x}_0 = (1, -1, 0, \dots, 0) \in \mathbb{R}^n$  and  $\mathbf{X}$  be such that  $\mathbb{P}[\mathbf{X} = \mathbf{x}_0] = 1$ . Then  $\overline{F}_{\mathbf{X}}(\mathbf{0}) = \min_{\mathbf{a} \in \mathbb{R}^n_+} \mathbb{E}[\mathbf{a}^\top (\mathbf{0} - \mathbf{X}) + 1]^+ = \min_{\mathbf{a} \in \mathbb{R}^n_+} [-\mathbf{a}^\top \mathbf{x}_0 + 1]^+ = 0$ , where the last equality follows from  $[-\mathbf{a}^\top \mathbf{x}_0 + 1]^+ \ge 0$  and that  $[-\mathbf{a}^\top \mathbf{x}_0 + 1]^+ = 0$  for  $\mathbf{a} = (1, 0, \dots, 0)$ . Similarly, for  $\mathbf{a} = (0, 1, 0, \dots, 0)$ , it is valid  $[\mathbf{a}^\top \mathbf{x}_0 + 1]^+ = 0$ , which yields  $\overline{F}_{-\mathbf{X}}(\mathbf{0}) = 0$ . However,  $\overline{F}_{(\mathbf{X}+(-\mathbf{X}))/2}(0) = \overline{F}_{\mathbf{0}}(\mathbf{0}) = 1 > 0$ . Consequently,  $\overline{F}_{\mathbf{X}}(\mathbf{x})$  is not quasi-convex.

**Remark** The proof of Proposition 2.4 uses constant random vectors. Suppose that n = 2 and that **X** is a constant r.v., i.e.,  $\mathbb{P}[\mathbf{X} = (x_1, x_2)] = 1$ . Let

$$g(x_1, x_2) = \overline{F}_{\mathbf{X}}(\mathbf{0}) = \inf_{\substack{(a_1, a_2) \in \mathbb{R}^2_+}} \mathbb{E}[(a_1, a_2)^\top (-\mathbf{X}) + 1]^+$$
$$= \inf_{\substack{(a_1, a_2) \in \mathbb{R}^2_+}} [a_1(-x_1) + a_2(-x_2) + 1]^+$$

If  $x_1 \leq 0$  and  $x_2 \leq 0$ , then  $a_1(-x_1) + a_2(-x_2) + 1 \geq 1$ , so that  $g(x_1, x_2) = 1$ . If  $x_1 > 0$  then for  $a_1 = 1/x_1$  and  $a_2 = 0$ , we have  $a_1(-x_1) + a_2(-x_2) + 1 = 0$ , and consequently,  $g(x_1, x_2) = 0$ . Similarly, if  $x_2 > 0$  then  $g(x_1, x_2) = 0$ . Thus,  $g(x_1, x_2) = 1$  when  $x_1 \leq 0$  and  $x_2 \leq 0$ , and  $g(x_1, x_2) = 0$  otherwise. This function is not quasiconvex and cannot be made quasi-convex by altering it values at a finite number of points.

Now let  $I = \{ (a'_1, a'_2) | a'_1 \ge 0, a'_2 \ge 0, a'_1 + a'_2 = 1 \}$  be the interval with endpoints (1, 0) and (0, 1). Then  $g(x_1, x_2)$  can be represented as

$$g(x_1, x_2) = \inf_{(a'_1, a'_2) \in I} \inf_{a_0 \ge 0} [a_0(-a'_1x_1 - a'_2x_2) + 1]^+ = \inf_{(a'_1, a'_2) \in I} h_{a'_1, a'_2}(x_1, x_2),$$

where

$$h_{a_1',a_2'}(x_1,x_2) = \inf_{a_0 \ge 0} [a_0(-a_1'x_1 - a_2'x_2) + 1]^+.$$

Observe that  $h_{a'_1,a'_2}(x_1,x_2) = 1$  when  $a'_1x_1 + a'_2x_2 \le 0$  and  $h_{a'_1,a'_2}(x_1,x_2) = 0$  otherwise, and that  $h_{a'_1,a'_2}(x_1,x_2)$  is quasi-convex. However, the infimum of such functions over  $(a'_1,a'_2) \in I$  is not quasi-convex. As  $(a'_1,a'_2)$  moves along *I* from (1, 0) to (0, 1), the line  $a'_1x_1 + a'_2x_2$  rotates from the *x*-axis to the *y*-axis. As a result,  $g(x_1,x_2) = 1$  if and only if the point  $(x_1,x_2)$  remains "under" the rotating line, which occurs if and only if  $x_1 \le 0$  and  $x_2 \le 0$ .

Note that  $g(x_1, x_2)$  is the infimum of quasi-convex functions  $h_{a'_1,a'_2}(x_1, x_2)$ , but  $g(x_1, x_2)$  is not "closely approximated" by any of these functions. Also, every function  $h_{a'_1,a'_2}(x_1, x_2)$  is a quasi-convex upper bound for  $g(x_1, x_2)$ . However, none of these quasi-convex upper bounds is "best" or "minimal". So,  $g(x_1, x_2)$  has many quasi-convex upper bounds, but the unique "minimal" quasi-convex upper bound does not exist.

In fact,  $\underline{F}_{\mathbf{X}}(\mathbf{x})$  is also not quasi-concave in **X**, even for n = 2, see [9]. It can be interpreted through dual characterization in a way similar to that of bCDF. Indeed,

$$1 - F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}[X_1 > x_1 \text{ or } \dots \text{ or } X_n > x_n]$$
  
= 
$$\sup_{W_i \in \{0,1\}} \sum_{i=1}^n \mathbb{E}[W_i] \text{ subject to } X_i W_i > x_i W_i, \quad i = 1, \dots, n, \quad \sum_{i=1}^n W_i \le 1.$$
(17)

Proposition 2.5 Define

$$\underline{F}_{\mathbf{X}}^{D}(\mathbf{x}) = \inf_{\mathbf{W} \in L^{\infty,n}(\Omega)} \left( 1 - \sum_{i=1}^{n} \mathbb{E}[W_i] \right)$$
  
subject to  $\mathbb{E}[X_i W_i] \ge x_i \mathbb{E}[W_i], W_i \in [0, 1], i = 1, ..., n, \sum_{i=1}^{n} W_i \in [0, 1].$ 

Also, let

$$\underline{F}_{\mathbf{X}}^{L}(\mathbf{x}) = \inf_{Y \leq_{cx} X} \mathbb{P}[\mathbf{X} < \mathbf{x}]$$

*be the maximal Schur-concave lower bound for*  $\mathbb{P}[\mathbf{X} < \mathbf{x}]$ *. Then on an atomless probability space,* 

$$\underline{F}_{\mathbf{X}}(\mathbf{x}) = \underline{F}_{\mathbf{X}}^{D}(\mathbf{x}) = \underline{F}_{\mathbf{X}}^{L}(\mathbf{x}).$$

**Proof** See Appendix C.

For illustration of these results, let r.v.'s  $X_1$  and  $X_2$  model profits from two different investments and let an investor evaluate the probability of both  $X_1$  and  $X_2$ to be non-positive. Let  $\mathbf{X} = (X_1, X_2)$ ,  $\mathbf{0} = (0, 0)$ , and let  $F_{\mathbf{X}}(\mathbf{0}) = \mathbb{P}[\mathbf{X} \leq \mathbf{0}]$ . Let  $\mathbf{X}$ assume values (-1, -1), (1, 1) with probabilities 1/3 and 2/3, respectively. Then  $F_{\mathbf{X}}(\mathbf{0}) = 1/3$ . Suppose X takes values (-1, -1), (1, 1), and (1, 1) in the events  $\omega_1$ ,  $\omega_2$  and  $\omega_3$  with probabilities  $\mathbb{P}[\omega_1] = \mathbb{P}[\omega_2] = \mathbb{P}[\omega_3] = 1/3$ . Let  $\mathcal{F} = \{A, B\}$  be a partition of the probability space by the events  $A = \{\omega_1, \omega_2\}$  and  $B = \{\omega_3\}$ . Then  $\mathbf{Y} = \mathbb{E}[\mathbf{X}|\mathcal{F}]$  is a random vector equal to (0, 0) and (1, 1) with probabilities 2/3 and 1/3, respectively. Note that Y is "averaged" version of X. However, we have  $F_{\mathbf{Y}}(\mathbf{0}) = 2/3 > 1/3 = F_{\mathbf{Y}}(\mathbf{0})$ . This means that  $F_{\mathbf{X}}(\mathbf{0})$  is not Schur-convex as a function of X. Instead, we suggest to use its Schur-convex upper bound given in Proposition 2.5—such bound decreases when X is replaced by its average-out version.

This example also demonstrates convenience of modeling the probability space as atomless even for discrete random variables. If the original random vector **X** were defined on a discrete probability space  $\Omega = \{\omega_1, \omega_2\}$  with  $\mathbb{P}[\omega_1] = 1/3$  and  $\mathbb{P}[\omega_2] = 2/3$ , the "averaged-out" version **Y** of **X** would not exist. For this  $\Omega$ ,

$$\sup_{Y\leqslant_{cx}X}F_{\mathbf{Y}}(\mathbf{0})=\frac{1}{3}\neq\frac{2}{3},$$

🖄 Springer

which illustrates that Proposition 2.3 fails on discrete probability spaces.

# 3 Buffered and reduced CDFs in optimization problems

Let **X** be a random vector,  $\mathbf{x} \in \mathbb{R}^n$ , and let  $\mathbf{f}(\mathbf{w};\mathbf{X}) = (f_1(\mathbf{w};\mathbf{X}), \dots, f_n(\mathbf{w};\mathbf{X}))$  be a vector of concave functions of **w** and  $\mathbf{h}(\mathbf{w};\mathbf{X}) = (h_1(\mathbf{w};\mathbf{X}), \dots, h_n(\mathbf{w};\mathbf{X}))$  be a vector of convex functions of **w**, defined on a convex set  $\mathbf{W} \subseteq \mathbb{R}^k$ . Let also  $\mathbb{E}[|f_i(\mathbf{w};\mathbf{X})|] < \infty$ ,  $\mathbb{E}[|h_i(\mathbf{w};\mathbf{X})|] < \infty$  for all  $\mathbf{w} \in \mathbf{W}$  and  $i \in \{1, \dots, n\}$ .

If all  $f_i$  in the definition of  $\mathbf{f}(\mathbf{w}; \mathbf{X})$  are convex, and  $\mathcal{W} = \mathcal{W}_1 \times \cdots \times \mathcal{W}_n$  with all  $\mathcal{W}_1, \ldots, \mathcal{W}_n$  being convex, then the problem  $\min_{\mathbf{w} \in \mathcal{W}} \overline{F}_{\mathbf{f}(\mathbf{w}; \mathbf{X})}(\mathbf{x})$  is convex.

### 3.1 bCDF and rCDFs in objective

This section considers the following two optimization problems

$$\min_{\mathbf{w}\in\mathbf{W}}\overline{F}_{\mathbf{f}(\mathbf{w};\mathbf{X})}(\mathbf{x}) \tag{18}$$

and

$$\max_{\mathbf{w}\in\mathbf{W}} \underline{F}_{\mathbf{h}(\mathbf{w};\mathbf{X})}(\mathbf{x}).$$
(19)

With functions  $G_F(\mathbf{a}, \mathbf{w})$  and  $G_H(\mathbf{a}, \mathbf{w})$  defined by

$$G_F(\mathbf{a}, \mathbf{w}) = \mathbb{E}[\mathbf{a}^\top (\mathbf{x} - \mathbf{f}(\mathbf{w}; \mathbf{X})) + 1]^+$$

and

$$G_{H}(\mathbf{a}, \mathbf{w}) = -\mathbb{E}[\min\{1, a_{1}(x_{1} - h_{1}(\mathbf{w}; \mathbf{X})), \dots, a_{n}(x_{n} - h_{n}(\mathbf{w}; \mathbf{X}))\}],$$

bCDF and rCDF can be recast as

$$\overline{F}_{\mathbf{f}(\mathbf{w};\mathbf{X})}(\mathbf{x}) = \inf_{\mathbf{a}\in\mathbb{R}^n_+} G_F(\mathbf{a},\mathbf{w})$$
(20)

and

$$\underline{F}_{\mathbf{h}(\mathbf{w};\mathbf{X})}(\mathbf{x}) = \sup_{\mathbf{a}\in\mathbb{R}^n_+} -G_H(\mathbf{a},\mathbf{w}).$$
(21)

It is assumed that in (20) and (21), the infimum and supremum with respect to **a** can be replaced by the minimum and maximum, respectively.<sup>4</sup> With this assumption, optimization problems (18) and (19) along with (20) and (21) take the form

$$\min_{\mathbf{w}\in\mathbf{W}}F_{\mathbf{f}(\mathbf{w};\mathbf{X})}(\mathbf{x}) = \min_{\mathbf{a}\in\mathbb{R}^{n}_{+},\ \mathbf{w}\in\mathbf{W}}G_{F}(\mathbf{a},\mathbf{w})$$
(22)

<sup>&</sup>lt;sup>4</sup> See, e.g., [11, Theorem 27.1] for sufficient conditions for a convex function to attain its minimum.

and

1

414

$$\max_{\mathbf{w}\in\mathbf{W}}\underline{F}_{\mathbf{h}(\mathbf{w};\mathbf{X})}(\mathbf{x}) = \max_{\mathbf{a}\in\mathbb{R}_{+}^{n}, \ \mathbf{w}\in\mathbf{W}} -G_{H}(\mathbf{a},\mathbf{w}) = -\min_{\mathbf{a}\in\mathbb{R}_{+}^{n}, \ \mathbf{w}\in\mathbf{W}} G_{H}(\mathbf{a},\mathbf{w}),$$
(23)

respectively, and (22) and (23) can be written as

$$\min_{\mathbf{a}\in\mathbb{R}^n_+, \ \mathbf{w}\in\mathbf{W}} G_I(\mathbf{a},\mathbf{w}), \quad I\in\{F,H\}.$$
(24)

Further, with an auxiliary variable  $a_0 \ge 0$  and with the set **A** defined by (11), problem (24) can be equivalently reformulated in the form

$$\min_{\mathbf{a}\in\mathbf{A},\ a_0\in\mathbb{R}_+,\ \mathbf{w}\in\mathbf{W}} G_l(a_0\,\mathbf{a},\mathbf{w}), \quad I\in\{F,H\}.$$
(25)

The idea is to solve (25) iteratively by alternating variables  $(a_0, \mathbf{w})$  and  $\mathbf{a}$ : solve (25) with respect to  $(a_0, \mathbf{w})$ , then with respect to  $\mathbf{a}$ , and again with respect to  $(a_0, \mathbf{w})$ , and then again with respect to  $\mathbf{a}$ , and so on. This method is known as coordinate descent, or, more generally, Block Coordinate Descent (BCD), in which variables are divided into *m* blocks, and optimization is done iteratively for one block at a time. BCD may not converge to stationary points of a non-convex objective function, which is convex in each block coordinate [10]. However, global convergence was studied under additional assumptions: two-block (m = 2) or strict quasiconvexity for m - 2 blocks [3, 4] and uniqueness of minimizer per block [1, Sect. 2.7]. See also [5] for convergence of BCD for nonconvex problems and for relevant references.

Note that in our case the optimization problems

$$\min_{\mathbf{a}\in\mathbf{A}} G_I(a_0 \, \mathbf{a}, \mathbf{w}), \quad I \in \{F, H\},$$

are convex with respect to  $\mathbf{a}$ , since the objective is the maximum of linear functions with respect to components of the vector  $\mathbf{a}$ . On the other hand, the optimization problems

$$\min_{u_0 \in \mathbb{R}, \mathbf{w} \in \mathbf{W}} G_I(a_0 \mathbf{a}, \mathbf{w}) = \min_{\mathbf{w} \in \mathbf{W}} \min_{a_0 \in \mathbb{R}} G_I(a_0 \mathbf{a}, \mathbf{w}), \quad I \in \{F, H\},$$

can be reduced to convex programming in **w**, since they are bPOE minimization problems. Indeed,

$$\min_{a_0 \in \mathbb{R}, \mathbf{w} \in \mathbf{W}} G_F(a_0 \mathbf{a}, \mathbf{w}) = \min_{\mathbf{w} \in \mathbf{W}} \min_{a_0 \in \mathbb{R}} G_F(a_0 \mathbf{a}, \mathbf{w}) = \min_{\mathbf{w} \in \mathbf{W}} \min_{a_0 \in \mathbb{R}} \mathbb{E}[a_0 \mathbf{a}^\top (\mathbf{x} - \mathbf{f}(\mathbf{w}; \mathbf{X})) + 1]^+$$
$$= \min_{\mathbf{w} \in \mathbf{W}} \overline{p}_{\mathbf{a}^\top (\mathbf{x} - \mathbf{f}(\mathbf{w}; \mathbf{X}))}(0).$$
(26)

Since all components of the vector function  $\mathbf{f}(\mathbf{w};\mathbf{X})$  are concave in  $\mathbf{w}$ , the function  $\mathbf{a}^{\mathsf{T}}(\mathbf{x} - \mathbf{f}(\mathbf{w};\mathbf{X}))$  is convex in  $\mathbf{w}$ . Consequently, minimization of bPOE with respect to  $\mathbf{w}$  for this function can be reduced to convex programming, see [6]. Also, (15a) and (15b) imply that

$$\min_{a_{0} \in \mathbb{R}, \mathbf{w} \in \mathbf{W}} G_{H}(a_{0} \mathbf{a}, \mathbf{w}) 
= \min_{\mathbf{w} \in \mathbf{W}} \min_{a_{0} \in \mathbb{R}} -\mathbb{E}[\min\{1, a_{0}a_{1}(x_{1} - h_{1}(\mathbf{w}; \mathbf{X})), \dots, a_{0}a_{n}(x_{n} - h_{n}(\mathbf{w}; \mathbf{X}))\}] 
= \min_{\mathbf{w} \in \mathbf{W}} \left( -\max_{a_{0} \in \mathbb{R}} \mathbb{E}[\min\{1, a_{0}a_{1}(x_{1} - h_{1}(\mathbf{w}; \mathbf{X})), \dots, a_{0}a_{n}(x_{n} - h_{n}(\mathbf{w}; \mathbf{X}))\}] \right) 
= \min_{\mathbf{w} \in \mathbf{W}} \left( \min_{a_{0} \in \mathbb{R}} \mathbb{E}[\max\{a_{0}a_{1}(h_{1}(\mathbf{w}; \mathbf{X}) - x_{1}), \dots, a_{0}a_{n}(h_{n}(\mathbf{w}; \mathbf{X}) - x_{n})\} + 1]^{+} - 1 \right) 
= -1 + \min_{\mathbf{w} \in \mathbf{W}} \overline{p}_{\max\{a_{1}(h_{1}(\mathbf{w}; \mathbf{X}) - x_{1}), \dots, a_{n}(h_{n}(\mathbf{w}; \mathbf{X}) - x_{n})\}} (0).$$
(27)

Since all components of the vector function h(w;X) are convex in w, and components of the vector **a** are nonnegative, the function

$$\max\{a_1(h_1(\mathbf{w};\mathbf{X}) - x_1), \dots, a_n(h_n(\mathbf{w};\mathbf{X}) - x_n)\}\$$

is convex in **w**. Consequently, minimization of bPOE for this function can be reduced to convex programming, see [6]. Finally, with bPOE functions, problem (25) can be formulated as follows

$$\min_{\mathbf{a}\in\mathbf{A},\,\mathbf{w}\in\mathbf{W}}\overline{p}_{g_{I}(\mathbf{a},\mathbf{w})}(0),\quad I\in\{F,H\},$$
(28)

where

$$g_F(\mathbf{a}, \mathbf{w}) = \mathbf{a}^{\mathsf{T}}(\mathbf{x} - \mathbf{f}(\mathbf{w}; \mathbf{X})),$$
  

$$g_H(\mathbf{a}, \mathbf{w}) = \max\{a_1(h_1(\mathbf{w}; \mathbf{X}) - x_1), \dots, a_n(h_n(\mathbf{w}; \mathbf{X}) - x_n)\}.$$

The algorithm for solving problem (28) is summarized in Appendix D.

### 3.2 bCDF and rCDF in constraints

This section considers minimization of an objective function  $V(\mathbf{w})$  with respect to variables  $\mathbf{w} \in \mathbb{R}^k$ :

$$\min_{\mathbf{w}\in\mathbf{W}} V(\mathbf{w}) \tag{29}$$

subject to constraints on buffered and reduced CDFs:

$$F_{\mathbf{f}(\mathbf{w};\mathbf{X})}(\mathbf{x}) \leqslant \overline{\gamma},\tag{30}$$

$$\underline{F}_{\mathbf{h}(\mathbf{w};\mathbf{X})}(\mathbf{x}) \ge \underline{\gamma}. \tag{31}$$

With the functions  $G_F(a_F \mathbf{a}_F, \mathbf{w})$  and  $G_H(a_H \mathbf{a}_H, \mathbf{w})$  introduced in Sect. 3.1 and under the assumption that in (20) and (21), the infimum and supremum with respect to **a** can be replaced by the minimum and maximum, respectively, problem (29)–(31) can be reformulated as follows

$$\min_{\mathbf{w}\in\mathbf{W},\ a_{I}\in\mathbb{R}_{+},\ \mathbf{a}_{I}\in\mathbf{A},\ I\in\{F,H\}}V(\mathbf{w})$$
(32)

subject to

$$G_F(a_F \,\mathbf{a}_F, \mathbf{w}) \leqslant \overline{\gamma},\tag{33}$$

$$G_{H}(a_{H}\,\mathbf{a}_{H},\mathbf{w})\leqslant-\gamma.\tag{34}$$

Relationship (26) implies that the constraint (33) with respect to variables  $(a_F, \mathbf{a}_F, \mathbf{w})$  can be replaced by

$$\overline{p}_{\mathbf{a}_{t}^{\mathsf{T}}(\mathbf{x}-\mathbf{f}(\mathbf{w};\mathbf{X}))}(0) \leqslant \overline{\gamma}$$

with respect to variables ( $\mathbf{a}_F$ ,  $\mathbf{w}$ ). By (2), the last constraint is equivalent to the superquantile constraint:

$$\overline{q}_{\mathbf{a}_{F}^{\mathsf{T}}(\mathbf{x}-\mathbf{f}(\mathbf{w};\mathbf{X}))}(1-\overline{\gamma}) \leqslant 0.$$
(35)

Similarly, (27) implies that the constraint (34) with respect to variables  $(a_H, \mathbf{a}_H, \mathbf{w})$  can be replaced by

$$-1 + \overline{p}_{\max\{a_{H_1}(h_1(\mathbf{w};\mathbf{X})-x_1),\dots,a_{H_n}(h_n(\mathbf{w};\mathbf{X})-x_n)\}}(0) \leq -\underline{\gamma}$$

with respect to variables  $(\mathbf{a}_H, \mathbf{w})$ . By (2), the last constraint is equivalent to the superquantile constraint

$$q_{\max\{a_{H1}(h_1(\mathbf{w};\mathbf{X})-x_1),\ldots,a_{Hn}(h_n(\mathbf{w};\mathbf{X})-x_n)}(\underline{\gamma}) \leqslant 0.$$
(36)

Finally, problem (32)–(34) can be recast in the form

$$\min_{\mathbf{w}\in\mathbf{W}, \mathbf{a}_F\in\mathbf{A}, \mathbf{a}_H\in\mathbf{A}} V(\mathbf{w}) \text{ subject to (35) and (36).}$$
(37)

Problem (37) can be solved iteratively as follows: solve (37) with respect to variables **w** with fixed  $\mathbf{a}_F$  and  $\mathbf{a}_H$ , then minimize the subquantile functions in (35) and (36) with respect to  $\mathbf{a}_F$  and  $\mathbf{a}_H$ , respectively, i.e., solve

$$\min_{\mathbf{a}_{F}\in\mathbf{A}}\bar{q}_{\mathbf{a}_{F}^{\mathsf{T}}(\mathbf{x}-\mathbf{f}(\mathbf{w};\mathbf{X}))}(1-\bar{\gamma}),\tag{38}$$

$$\min_{\mathbf{a}_{H} \in \mathbf{A}} \overline{q}_{\max\{a_{H1}(h_{1}(\mathbf{w};\mathbf{X})-x_{1}),\dots,a_{Hn}(h_{n}(\mathbf{w};\mathbf{X})-x_{n})(\underline{\gamma})},$$
(39)

then solve (37) again with respect to **w** with previously found  $\mathbf{a}_F$  and  $\mathbf{a}_H$ , then solve (38) and (39), and so on.

The described algorithm for solving problem (37) is summarized in Appendix E.

## 4 Applications

### 4.1 Two-asset option

A two-asset option is an exotic option whose payoff depends on prices  $p_A(t)$  and  $p_B(t)$  of assets A and B at time t. For example, it may have nonzero payoff if  $p_A(T)$  and  $p_B(T)$  simultaneously exceed some strike prices  $C_A$  and  $C_B$ , respectively, at some future time T. In this case, if  $p_A(T)$  and  $p_B(T)$  are assumed to be random variables, the option pays with the probability

$$p = \mathbb{P}[p_A(T) \ge C_A, p_B(T) \ge C_B]. \tag{40}$$

Since in general,  $p_A(T)$  and  $p_B(T)$  are not independent random variables,  $\mathbb{P}[p_A(T) \ge C_A, p_B(T) \ge C_B] \neq \mathbb{P}[p_A(T) \ge C_A] \cdot \mathbb{P}[p_B(T) \ge C_B]$ . If  $p_A(0)$  and  $p_B(0)$  are (known) prices of assets A and B at the current time t = 0, then (40) can be recast in terms of adjusted rates of returns  $r_A = (p_A(T) - C_A)/p_A(0)$  and  $r_B = (p_B(T) - C_B)/p_B(0)$ :

$$p = \mathbb{P}[r_A \ge 0, r_B \ge 0].$$

The joint distribution of  $r_A$  and  $r_B$  can be estimated from the historical data. Let  $r_{A,i}$  and  $r_{B,i}$ , i = 1, ..., T, be the adjusted rates of return of assets A and B over the last T periods. Then p can be estimated as the number of times  $(r_{A,i}, r_{B,i}) \ge 0$  divided by T. However, this estimate is sensitive to small perturbations in the data. Alternatively, we can define  $X = (-r_A, -r_B)$  and estimate the smallest Schur-convex upper bound (9) for p:

$$\begin{split} \overline{F}_{-X}(0) &= \inf_{a \ge 0, b \ge 0} \mathbb{E}[a r_A + b r_B + 1]^+ \\ &= \inf_{a \ge 0, b \ge 0} \frac{1}{T} \sum_{i=1}^T [a r_{A,i} + b r_{B,i} + 1]^+, \\ &= \min_{\substack{a \ge 0, b \ge 0, \\ z_1, \dots, z_T}} \frac{1}{T} \sum_{i=1}^T z_i \\ &\text{subject to} \quad z_i \ge a r_{A,i} + b r_{B,i} + 1, \quad z_i \ge 0, \quad i = 1, \dots, T, \end{split}$$

which is a linear program.

### 4.2 Credit rating

Optimizing a portfolio of credit default swaps (CDSs) with constraints on the default probabilities of tranches [8] is a non-convex optimization problem. Let *M* be the number of tranches, *T* the number of time periods, and  $x_m^t$ , m = 1, ..., M, t = 1, ..., T, the attachment point of tranche *m* at time *t*. Also, suppose there are *K* available CDSs. Let  $\theta_k^t$  be the cumulative loss of *k*-th CDS at time *t*, and let  $y_k$  be the weight of the *k*-th asset in the asset pool. Then the total loss of the CDS pool at time *t* is  $L(\theta_s^t \mathbf{y}) = \sum_{k=1}^{K} \theta_k^t y_k$ ,

where  $\theta^t = (\theta_1^t, \dots, \theta_K^t)$  and  $\mathbf{y} = (y_1, \dots, y_K)$ . Consider constraints on the default probabilities of tranches [8]:

$$1 - \mathbb{P}\left[L(\boldsymbol{\theta}^{t}, \mathbf{y}) \leq x_{m}^{1}, \dots, L(\boldsymbol{\theta}^{t}, \mathbf{y}) \leq x_{m}^{T}\right] \leq p_{m}, \quad m = 2, \dots, M.$$

$$(41)$$

With  $X_t = L(\theta^t, \mathbf{y}) - x_m^t, t = 1, \dots, T$ , (41) can be recast as

$$\mathbb{P}\left[\max\{X_1, \dots, X_T\} \ge 0\right] \le p_m, \quad m = 2, \dots, M.$$
(42)

Each constraint in (42) is non-convex and can be approximated by

$$\overline{p}_{\max\{X_1,\dots,X_T\}}(0) \leqslant \overline{p}_m,\tag{43}$$

where  $\overline{p}_m$  is some scaled bound on probability; see [8]. Also, (15b)–(15d) imply that

$$1 - \underline{F}_{\mathbf{X}}(\mathbf{0}) \leqslant \overline{p}_{\max\{X_1, \dots, X_T\}}(\mathbf{0}),$$

where  $\underline{F}_{\mathbf{X}}(\mathbf{x})$  is defined in (10). Consequently,  $1 - \underline{F}_{\mathbf{X}}(\mathbf{0})$  is a better upper bound for  $\mathbb{P}[\max(X_1, \dots, X_T) \ge 0]$  than  $\overline{p}_{\max\{X_1, \dots, X_T\}}(0)$ , and (43) can be replaced by

$$1 - \underline{F}_{\mathbf{X}}(\mathbf{0}) \leqslant \overline{p}_m. \tag{44}$$

In fact,  $1 - \underline{F}_{\mathbf{X}}(\mathbf{0})$  is *the best* Schur convex upper bound on this probability for an atomless probability space; see Proposition 2.5. The optimization problem considered in [8] is formulated by

$$\min_{\mathbf{x},\mathbf{y}} \sum_{i=1}^{T} \frac{1}{(1+r)^{t}} \sum_{m=1}^{M} \Delta s_{m} \mathbb{E}[[x_{m+1}^{t} - L(\boldsymbol{\theta}^{t}, \mathbf{y})]^{+}]$$
subject to  $1 - \mathbb{P}[L(\boldsymbol{\theta}^{t}, \mathbf{y}) \leq x_{m}^{1}, \dots, L(\boldsymbol{\theta}^{t}, \mathbf{y}) \leq x_{m}^{T}] \leq p_{m}, \quad m = 2, \dots, M,$ 

$$\sum_{k=1}^{K} c_{k} y_{k} \geq \eta, \quad \sum_{k=1}^{K} y_{k} = 1, \quad y_{k} \geq 0, \quad k = 1, \dots, K,$$

$$x_{m}^{t} \geq x_{m-1}^{t}, \quad m = 1, \dots, M, \quad t = 1, \dots, T,$$

$$0 \leq x_{m}^{t} \leq 1, m = 2, \dots, M, \quad t = 1, \dots, T,$$
(45)

where  $\mathbf{x} = \{x_m^t\}_{m=2,...,M}^{t=1,...,T}$ , *r* is the interest rate,  $\Delta s_m = s_m - s_{m-1}$ ,  $s_m$  is the spread payment for each tranche m = 1, ..., M,  $c_k$  is the annual income spread payment of the *k*-th CDS, k = 1, ..., K, and  $\eta$  is a real-valued parameter. After replacing constraints (41) by (44), the optimization problem (45) takes the form

$$\min_{\mathbf{x}, \mathbf{y}, \mathbf{a}} \sum_{i=1}^{T} \frac{1}{(1+r)^{t}} \sum_{m=1}^{M} \Delta s_{m} \mathbb{E}[[x_{m+1}^{t} - L(\boldsymbol{\theta}, \mathbf{y})]^{+}]$$
(46a)

subject to  $\mathbb{E}[\max\{a_1^m(L(\boldsymbol{\theta}^1, \mathbf{y}) - x_m^1), \dots, a_T^m(L(\boldsymbol{\theta}^T, \mathbf{y}) - x_m^T)\} + 1]^+ \leq \overline{p}_m,$ 

$$m = 2, \dots, M, \tag{46b}$$

$$a_t^m \ge 0, \quad m = 2, \dots, M, \quad t = 1, \dots, T,$$
 (46c)

$$\sum_{k=1}^{K} c_k y_k \ge \eta, \sum_{k=1}^{K} y_k = 1, \quad y_k \ge 0, \quad k = 1, \dots, K,$$
(46d)

$$\begin{aligned} x_m^t \geqslant x_{m-1}^t, \ m = 1, \dots, M, \ t = 1, \dots, T, \\ 0 \leqslant x_m^t \leqslant 1, m = 2, \dots, M, \ t = 1, \dots, T, \end{aligned}$$
 (46e)

where  $\mathbf{a} = \{a_m^t\}_{m=2,...,M}^{t=1,...,T}$ 

# 5 Case study

Veremyev et al. [17] solved the CDS portfolio problem (45) and found an optimal portfolio of CDSs along with optimal attachment/detachment points over a multiperiod horizon. Pertaia et al. [8] replaced multivariate POE constraints in (45) by one-dimensional bPOE constraints (43) and solved the corresponding optimization problem.<sup>5</sup>

Here we use Algorithm 2 to solve problem (46a)–(46e) for 53 CDSs over 5 time periods with 5 tranches for 300,000 simulated scenarios and for  $\eta = 120$  and  $\eta = 80$  in (46d).<sup>6</sup> The experiments were run on an Intel i7-2.6Ghz processor with 32GB DDR4-3200Mhz RAM and used Portfolio Safeguard (PSG),<sup>7</sup> which has precoded superquantile (CVaR) function as well as many other popular risk functions.

Table 1 presents objective value in Step 2 or 4, running time in Step 2 or 4, and subquantile sub-problem running time in Step 3 or 5 for each iteration in Algorithm 2 with  $\eta = 120$  and  $\epsilon = 10^{-7}$ . It shows that no improvements are made after the second iteration (Algorithm 2 stops after the third iteration, but Table 1 shows one more iteration to demonstrate that there are no further improvements in the objective value). Constraint (36) in problem (37), which is solved in Steps 2 and 4, is equivalent to constraints (46b) with m = 2, 3, 4, 5. Table 2 reports slack in constraints (46b) for m = 2, 3, 4, 5 as percentage of the constraint right-hand side, i.e.,

<sup>&</sup>lt;sup>5</sup> Solution of problem (45) and of the one with bPOE constraints are available at http://uryasev.ams. stonybrook.edu/index.php/research/testproblems/financial\_engineering/structuring-step-up-cdo/

<sup>&</sup>lt;sup>6</sup> The codes and solutions of this problem are available at http://uryasev.ams.stonybrook.edu/index.php/research/testproblems/financial\_engineering/structuring-step-up-cdo/.

<sup>&</sup>lt;sup>7</sup> Portfolio Safeguard software is a product of American Optimal Decisions, Inc. (http://aorda.com). To optimize convex nonsmooth functions, PSG solves a sequence of linear or quadratic programming problems. For large-scale optimization problems, PSG can call GUROBI solver (Gurobi Optimization, https://www.gurobi.com/).

$$\delta_m = 100\% \cdot \frac{\overline{p}_m - \mathbb{E}[\max\{a_1^m(L(\boldsymbol{\theta}^1, \mathbf{y}) - x_m^1), \dots, a_T^m(L(\boldsymbol{\theta}^T, \mathbf{y}) - x_m^T)\} + 1]^+}{\overline{p}_m}.$$
 (47)

At Step 2 in the first iteration, constraints (46b) for m = 3, 4, 5 are active and after that, all constraints in (46b) become inactive till the algorithm stops after the third iteration. Since problem (46a)-(46e) without constraints (46b) is convex and none of constraints (46b) is active at optimality, the found solution is globally optimal. Note that one-dimensional bPOE constraints (43) are active, as we see from Step 2 in the first iteration.

Tables 3 and 4 present results similar to those in Tables 1 and 2 but for  $\eta = 80$ . This time, the objective value stops improving after the second iteration even if constraints (46b) for m = 3, 4, 5 remain active in all iterations.

<b>Table 1</b> Objective value in Step 2 or 4, running time in Step 2 or 4, and subquantile sub-problem running time in Step 3 or 5 for each iteration in Algorithm 2 for $\eta = 120$ in constraint (46d)	Iteration	Objective value	Main problem running time (s)	Subquantile sub- problem running time (s)
	1	590.24	376.74	_
	2	577.83	433.05	1.27
	3	577.83	368.86	1.02
	4	577.83	286.11	1.07

<b>Table 2</b> Slack, $\delta_m$ (defined by	Ite
(47)), in constraints (46b) for	
m = 2, 3, 4, 5, in Algorithm 2	
for $\eta = 120$ in constraint (46d)	1

Iteration	Step	$\delta_2, \%$	$\delta_3, \%$	$\delta_4,\%$	$\delta_5, \%$
1	2	3.24	0.00	0.00	0.00
1	3	14.23	10.91	8.78	5.08
2	4	3.84	0.050	0.057	0.097
2	5	4.52	1.37	1.32	0.80
3	4	4.67	1.36	1.32	0.80
3	5	4.52	1.37	1.32	0.80
4	4	4.67	1.36	1.32	0.80
4	5	4.52	1.37	1.32	0.80

Table 3 Objective value in Step 2 or 4, running time in Step 2 or 4, and subquantile sub-problem running time in Step 3 or 5 for each iteration in Algorithm 2 for  $\eta = 80$  in constraint (46d)

Iteration	Objective value	Main problem running time (s)	Subquantile sub- problem running time (s)
1	545.22	363.62	_
2	544.41	419.76	3.63
3	544.41	291.45	3.19
4	544.41	273.03	3.26

$m = 2, 3, 4, 5, \text{ in Algorithm 2} for \eta = 120 \text{ in constraint (46d)} 1 \qquad 2 \qquad 3.24 \qquad 0.00 \qquad 0.00 \\ 1 \qquad 3 \qquad 3.66 \qquad 0.83 \qquad 1.00 \\ 2 \qquad 4 \qquad 2.95 \qquad 0 \qquad 0 \\ 2 \qquad 5 \qquad 2.86 \qquad 0 \qquad 0 \\ 1 \qquad 0 \qquad 0 \qquad 0 \\ 0 \qquad 0 \qquad 0 \\ 0 \qquad 0 \qquad 0 \\ 0 \qquad 0 \qquad$	$\delta_5, \%$
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	0.00
2       4       2.95       0       0         2       5       2.86       0       0	1.11
2 5 2.86 0 0	0
	0
3 4 2.95 0 0	0
3 5 2.86 0 0	0
4 4 2.95 0 0	0
4 5 2.86 0 0	0

The case study shows that optimization problems with bCDF functions can be efficiently optimized (even for large-scale optimization problems with 300,000 scenarios).

# Appendix

#### A. Proof of proposition 2.2

With notation

$$\overline{p}(\mathbf{X}) = \overline{F}_{-\mathbf{X}}(0) = \inf_{\mathbf{a} \in \mathbb{R}^n_+} \mathbb{E}[\mathbf{a}^\top \mathbf{X} + 1]^+$$
(48)

Proposition 2.2 can be reformulated as follows.

## **Proposition A.1** For any $\mathbf{X} \in L^{1,n}(\Omega)$ ,

$$\overline{p}(\mathbf{X}) = \sup_{W \in L^{1}(\Omega)} \mathbb{E}[W] \text{ subject to } \mathbb{E}[X_{i}W] \ge 0, \quad i = 1, \dots, n, \quad W \in [0, 1].$$
(49)

**Proof** Let  $\overline{p}'(\mathbf{X})$  be the right-hand side of (49). We first prove that  $\overline{p}'(\mathbf{X}) \leq \overline{p}(\mathbf{X})$ . Let  $\mathcal{W}$  be the feasible set in (49). For every  $\mathcal{W} \in \mathcal{W}$ , and every  $\mathbf{a} \in \mathbb{R}^n_+$ ,

$$\mathbb{E}[W] \leq \mathbb{E}[W] + \mathbf{a}^{\top} \mathbb{E}[\mathbf{X}_{i}W] + \mathbb{E}[(1-W)[\mathbf{a}^{\top}\mathbf{X}+1]^{+}].$$

Since the second and the third terms in the right-hand side are non-negative,

$$\mathbb{E}[W] \leq \mathbb{E}[W(1 + \mathbf{a}^{\mathsf{T}}\mathbf{X} - [\mathbf{a}^{\mathsf{T}}\mathbf{X} + 1]^{+})] + \mathbb{E}[\mathbf{a}^{\mathsf{T}}\mathbf{X} + 1]^{+} \leq \mathbb{E}[\mathbf{a}^{\mathsf{T}}\mathbf{X} + 1]^{+},$$

and since  $W \in \mathcal{W}$  and  $\mathbf{a} \in \mathbb{R}^n_+$  are arbitrary,  $\overline{p}'(\mathbf{X}) \leq \overline{p}(\mathbf{X})$  follows.

We next prove  $\overline{p}'(\mathbf{X}) \ge \overline{p}(\mathbf{X})$ . For n = 1, X is a random variable and  $\overline{p}'(X) \ge \overline{p}(X)$  was claimed in the proof of Proposition 2.5 in [6] (however, the proof was incomplete). In this case, let  $\underline{q}_X(\alpha)$  be defined by (3) and let  $\underline{q}_X^{-1}(x)$  be the inverse function for  $q_y(\alpha)$ , which is properly defined on  $x \in [\mathbb{E}X, \sup X)$ , see [6].

Then consider the following cases:

- If sup X < 0, then  $\mathcal{W} = \{0\}$  and  $\overline{p}'(X) = \mathbb{E}[0] = 0$ .
- If  $\sup X = 0$ , then  $I_{\{X=0\}} \in \mathcal{W}$ , hence  $\overline{p}'(X) \ge \mathbb{E}[I_{\{X=0\}}] = \mathbb{P}[X = 0] = \mathbb{P}[X = \sup X]$ .
- If E[X] < 0 < sup X, then let t = q<sub>X</sub>(q<sub>X</sub><sup>-1</sup>(0)) and let A<sup>−</sup>, A<sup>0</sup>, A<sup>+</sup> be the events X < t, X = t, and X > t, respectively. The definition of t implies that

$$1 - \mathbb{P}[A^+] - \mathbb{P}[A^0] \le \underline{q}_X^{-1}(0) \le 1 - \mathbb{P}[A^+].$$
(50)

By definition of  $q_{y}^{-1}(0)$ ,

$$0 = \int_{\underline{q}_{X}^{-1}(0)}^{1} H_{X}(\alpha) d\alpha = \int_{\underline{q}_{X}^{-1}(0)}^{1-\mathbb{P}[A^{+}]} t d\alpha + \int_{1-\mathbb{P}[A^{+}]}^{1} H_{X}(\alpha) d\alpha$$
  
=  $t \Big( 1 - \mathbb{P}[A^{+}] - \underline{q}_{X}^{-1}(0) \Big) + \int_{A^{+}} X(\omega) d\mathbb{P} = rt \mathbb{P}[A_{0}] + \int_{A^{+}} X(\omega) d\mathbb{P},$ 

where  $r = \frac{1 - r_1 A_1 - q_x}{\mathbb{P}[A_0]}$  if  $\mathbb{P}[A_0] > 0$  and r = 0 if  $\mathbb{P}[A_0] = 0$ . Note that  $0 \le r \le 1$  by (50).

Now, let *W* be an r.v. assuming values of 0, *r*, and 1 on  $A^-$ ,  $A^0$ , and  $A^+$ , respectively. Then

$$\mathbb{E}[XW] = 0 + r \int_{A^0} X(\omega) d\mathbb{P} + \int_{A^+} X(\omega) d\mathbb{P} = rt\mathbb{P}[A_0] + \int_{A^+} X(\omega) d\mathbb{P} = 0,$$

hence  $W \in \mathcal{W}$ . Thus,

$$\overline{p}'(X) \ge \mathbb{E}[W] = r\mathbb{P}[A_0] + \mathbb{P}[A_+] = 1 - \underline{q}_X^{-1}(0).$$

• If  $0 \leq \mathbb{E}[X]$  then  $W \equiv 1 \in \mathcal{W}$ , hence  $\overline{p}'(X) = \mathbb{E}[1] = 1$ .

In summary,  $\overline{p}'(X) \ge \overline{p}_X^+(0)$ , where  $\overline{p}_X^+(x)$  is defined in Definition 3.9 of [6]. By Proposition 3.10 in [6],  $\overline{p}_X^+(0) = \min_{a \in \mathbb{R}^n_+} \mathbb{E}[a^\top X + 1]^+$ , which finishes the proof of  $\overline{p}'(X) \ge \overline{p}(X)$  for n = 1.

Now let *n* be an arbitrary positive integer. By contradiction, suppose that  $\overline{p}'(\mathbf{X}) < \overline{p}(\mathbf{X})$  and choose any  $k \in (\overline{p}'(\mathbf{X}), \overline{p}(\mathbf{X}))$ . Let

$$K = \{ \mathbf{y} \in \mathbb{R}^n | \mathbf{y} = \mathbb{E}[\mathbf{X}W] \text{ for some } W \in [0, 1] \text{ with } \mathbb{E}[W] \ge k \}.$$

Then *K* is a convex set, and  $\overline{p}'(\mathbf{X}) < k$  implies that  $\mathbb{R}^n_+ \cap K = \emptyset$ . By separation hyperplane theorem, this implies that there exists a vector  $\mathbf{c} \in \mathbb{R}^n$  and constant *b* such that (i)  $\mathbf{c}^\top \mathbf{y} \ge b$  for all  $\mathbf{y} \in \mathbb{R}^n_+$  but (ii)  $\mathbf{c}^\top \mathbf{y} < b$  for all  $\mathbf{y} \in K$ . Condition (i) implies that  $\mathbf{c} \ge 0$  and  $b \le 0$ . Then (ii) implies that for every *W* satisfying  $0 \le W \le 1$  and  $\mathbb{E}[W] \ge k$ , we have  $0 \ge b > \mathbf{c}^\top \mathbb{E}[\mathbf{X}W] = \mathbb{E}[YW]$ , where  $Y = \mathbf{c}^\top \mathbf{X}$ . But this implies that  $\overline{p}'(Y) < k$ . On the other hand,

$$\overline{p}(Y) = \inf_{a \in \mathbb{R}_+} \mathbb{E}[aY+1]^+ = \inf_{a \in \mathbb{R}_+} \mathbb{E}[a\mathbf{c}^\top \mathbf{X}+1]^+ \ge \inf_{\mathbf{a} \in \mathbb{R}_+^n} \mathbb{E}[\mathbf{a}^\top \mathbf{X}+1]^+ = \overline{p}(\mathbf{X}) > k.$$

Since *Y* is a one-dimensional random variable, this contradicts to  $\overline{p}'(X) \ge \overline{p}(X)$  for the case n = 1 proved above.

### B. Proof of proposition 2.3

With notation (48), Proposition 2.3 can be reformulated as

**Proposition B.1** On an atomless probability space  $\Omega$ ,  $\overline{p}(\mathbf{X})$  is the minimal Schurconvex upper bound for  $p(\mathbf{X}) = \mathbb{P}[\mathbf{X} \ge 0]$ .

**Proof** For any  $\mathbf{a} \in \mathbb{R}_{+}^{n}$ , function  $g(\mathbf{x}) = [\mathbf{a}^{\top}\mathbf{x} + 1]^{+}$  is a convex function of  $\mathbf{x}$ , hence  $\mathbf{X} \geq_{cx} \mathbf{Y}$  implies that  $\mathbb{E}[\mathbf{a}^{\top}\mathbf{X} + 1]^{+} \geq \mathbb{E}[\mathbf{a}^{\top}\mathbf{Y} + 1]^{+}$ . Taking infimum over all  $\mathbf{a} \in \mathbb{R}_{+}^{n}$ , we obtain  $\overline{p}(\mathbf{X}) \geq \overline{p}(\mathbf{Y})$ , and consequently, the function  $\overline{p}(\mathbf{X})$  is Schur-convex. Also, for any  $\mathbf{a} \in \mathbb{R}_{+}^{n}$ , we have  $[\mathbf{a}^{\top}\mathbf{X} + 1]^{+} \geq I_{\{\mathbf{X} \geq 0\}}$ , so that  $\overline{p}(\mathbf{X}) \geq \mathbb{E}[I_{\{\mathbf{X} \geq 0\}}] = \mathbb{P}[\mathbf{X} \geq 0] = p(\mathbf{X})$  is indeed an upper bound for  $p(\mathbf{X})$ . It is left to prove that this upper bound is minimal, i.e., for every  $\mathbf{X} \in L^{1,n}(\Omega)$ , there exist a  $\mathbf{Y} \in L^{1,n}(\Omega)$  such that  $\mathbf{X} \geq_{cx} \mathbf{Y}$  and  $\mathbb{P}[\mathbf{Y} \geq 0] \geq \overline{p}(\mathbf{X})$ .

Let  $\mathcal{W}$  be a feasible set in (49). Consider probability space  $\Omega' = \Omega \times [0, 1]$ , where [0, 1] is the unit interval with Lebesgue measure. Write elements of  $\Omega'$  as  $(\omega, t)$ , where  $\omega \in \Omega$  and  $t \in [0, 1]$ . Let  $A = \{(\omega, t) \in \Omega' : W(\omega) \leq t\}$  for arbitrary  $W \in W$ . Then  $\mathbb{P}'(A) = \mathbb{E}[W]$ , where  $\mathbb{P}'$  is the probability measure on  $\Omega'$ . For any  $\mathbf{X} \in L^{1,n}(\Omega)$ , let  $\mathbf{X}' \in L^{1,n}(\Omega')$  be the random vector defined by  $\mathbf{X}'(\omega, t) = \mathbf{X}(\omega)$ . Then random vectors **X** and **X'** have the same distribution. Let  $\mathbf{Y}' = \mathbb{E}[\mathbf{X}'|I_A]$ , where  $I_A$  is the indicator function of an event A. Then  $W \in W$  implies that  $\mathbb{E}[\mathbf{X}_i W] \ge 0, i = 1, ..., n$ , which in turn implies that  $\mathbb{E}[X'_i|I_A = 1] \ge 0$  for each component  $X'_i$  of  $\mathbf{X}'$ , or, equivalently, that  $\mathbf{Y}'(\omega, t) \ge 0$  whenever  $(\omega, t) \in A$ . Consequently,  $\mathbb{P}'[\mathbf{Y}' \ge 0] \ge \mathbb{P}'(A) = \mathbb{E}[W]$ . If there exists  $\mathbf{Y} \leq_{cx} \mathbf{X}$  such that  $\mathbb{P}[\mathbf{Y} \ge 0] = \mathbb{P}'[\mathbf{Y}' \ge 0]$ , then, since  $W \in \mathcal{W}$  is arbitrary, this would imply that  $\sup_{\mathbf{Y} \leq \mathbf{X}} \mathbb{P}[\mathbf{Y} \ge 0] \ge \sup_{W \in \mathcal{W}} \mathbb{E}[W] = \overline{p}(\mathbf{X})$ , where the last equality is (49). Indeed,  $\mathbf{Y}' = \mathbb{E}[\mathbf{X}'|I_A]$  implies that  $\mathbf{X}' \ge_{cr} \mathbf{Y}'$ . Since  $\Omega$  is atomless, there exists a pair of random vectors  $\mathbf{X}'' \in L^{1,n}(\Omega)$  and  $\mathbf{Y} \in L^{1,n}(\Omega)$  with the same joint distribution as X' and Y'. Then  $\mathbb{P}[\mathbf{Y} \ge 0] = \mathbb{P}'[\mathbf{Y}' \ge 0]$  and  $\mathbf{X}'' \ge_{cx} \mathbf{Y}$ . Since **X** and **X**<sup>"</sup> have the same distribution,  $\mathbf{X} \ge_{cx} \mathbf{Y}$ , as required. 

### C. Proof of proposition 2.5

Proposition 2.5 can be reformulated as follows.

### Proposition C.1 Let

$$\overline{F}_1(\mathbf{X}) = 1 - \underline{F}_{\mathbf{X}}(\mathbf{0}) = \inf_{\mathbf{a} \in \mathbb{R}^n_+} \mathbb{E}[\max\{a_1 X_1, \dots, a_n X_n\} + 1]^+,$$
(51)

$$\overline{F}_{2}(\mathbf{X}) = \sup_{\mathbf{w} \in L^{\infty,n}(\Omega)} \sum_{i=1}^{n} \mathbb{E}[W_{i}]$$

subject to  $\mathbb{E}[X_i W_i] \ge 0$ ,  $W_i \in [0, 1]$ , I = 1, ..., n,  $\sum_{i=1}^n W_i \in [0, 1]$ . (52)

Also, let

$$\overline{F}_{3}(\mathbf{X}) = \sup_{\mathbf{Y} \leq_{cx} \mathbf{X}} \mathbb{P}[\max\{Y_{1}, \dots, Y_{n}\} \ge 0].$$

Then on an atomless probability space,

$$\overline{F}_1(\mathbf{X}) = \overline{F}_2(\mathbf{X}) = \overline{F}_3(\mathbf{X}).$$

**Proof** First, let us show that (51) is a dual problem to (52): introduce dual variables  $a_i \ge 0$  and  $0 \le M \in L^1(\Omega)$  for constraints  $\mathbb{E}[X_i W_i] \ge 0$  and  $\sum_{i=1}^n W_i \le 1$ , respectively. Minimization of the Lagrangian for (52) with respect to  $\mathbf{a} = (a_1, \dots, a_n)$  and M yields

$$\begin{split} &\inf_{\mathbf{a}\in\mathbb{R}^{n}_{+},0\leqslant M\in L^{1}(\Omega)}\mathbb{E}[M] + \sup_{\mathbf{w}\in L^{\infty,d},\ \mathbf{w}\in[0,1]}\sum_{i=1}^{n}\mathbb{E}[W_{i}(1+a_{i}X_{i}-M)] \\ &= \inf_{\mathbf{a}\in\mathbb{R}^{n}_{+},0\leqslant M\in L^{1}(\Omega)}\mathbb{E}\left[M + \sum_{i=1}^{n}[1+a_{i}X_{i}-M]^{+}\right] \\ &= \inf_{\mathbf{a}\in\mathbb{R}^{n}_{+}}\mathbb{E}\left[\max\{a_{1}X_{1}+1,\ldots,a_{n}X_{n}+1\}\right]^{+}, \end{split}$$

where the last equality follows from comparing M to  $m = \max\{a_1X_1 + 1, \dots, a_nX_n + 1\}$ . Indeed, if m < 0 then the subdifferential  $\partial(M + \sum_{i=1}^{n} [1 + a_iX_i - M]^+) = 1$  for  $M \ge 0$ , so M = 0 is optimal; otherwise, at M = m, the subdifferential is  $1 + k \cdot [-1, 0] + (n - k) \cdot 0 \ge 0$ , for some  $k \in \{1, \dots, n\}$ , so M = m is optimal.

Now, let  $\mathbf{Y} \leq_{cx} \mathbf{X}$  and  $\mathbf{a} \in \mathbb{R}^n_+$  be arbitrary. The function  $u(X) = [\max\{a_1X_1, \dots, a_nX_n\} + 1]^+$  is the maximum of linear functions and therefore is convex. Consequently,

$$E[u(\mathbf{X})] \ge E[u(\mathbf{Y})] \ge E[I_{\max\{Y_1,\dots,Y_n\}\ge 0\}}] = \mathbb{P}[\max\{Y_1,\dots,Y_n\}\ge 0]$$

This implies that

$$\overline{F}_1(\mathbf{X}) \geqslant \overline{F}_3(\mathbf{X}).$$

We next adapt the proof of Proposition 2.3 to prove that

$$\overline{F}_3(\mathbf{X}) \geqslant \overline{F}_2(\mathbf{X}).$$

Indeed, take any  $W_1, \ldots, W_n$  in the feasible set of (52). Consider probability space  $\Omega' = \Omega \times [0, 1]$ , where [0, 1] is the unit interval with Lebesgue measure. Write elements of  $\Omega'$  as  $(\omega, t)$ , where  $\omega \in \Omega$  and  $t \in [0, 1]$ . For every  $j = 1, \ldots, n$ , let  $A_j = \{(\omega, t) \in \Omega' : \sum_{i=1}^{j-1} W_i(\omega) < t \leq \sum_{i=1}^{j} W_i(\omega)\}$ . Also, let

 $A = \{(\omega, t) \in \Omega' : \sum_{i=1}^{n} W_i(\omega) < t\}$ . Then sets  $A_1, \dots, A_n, A$  form a partition of  $\Omega'$ , which we denote  $\mathcal{F}'$ . Note that  $1 - \mathbb{P}'(A) = \mathbb{E}[e^{\top}W]$ , where  $\mathbb{P}'$  is the probability measure on  $\Omega'$  and  $e^{\top}W = \sum_{i=1}^{n} W_i$ . For any  $\mathbf{X} \in L^{1,n}(\Omega)$ , let  $\mathbf{X}' \in L^{1,n}(\Omega')$  be the random vector defined by  $\mathbf{X}'(\omega, t) = \mathbf{X}(\omega)$ . Then random vectors  $\mathbf{X}$  and  $\mathbf{X}'$  have the same distribution. Let  $\mathbf{Y}' = \mathbb{E}[\mathbf{X}' | \mathcal{F}']$ . Since  $W_1, \ldots, W_n$  are in the feasible set of (52), we have  $\mathbb{E}[\mathbf{X}_i W_i] \ge 0, i = 1, ..., n$ , which in turn implies that  $\mathbb{E}[\mathbf{X}'_i | I_{A_i} = 1] \ge 0$ for each component  $X'_i$  of  $\mathbf{X}'$ , or, equivalently, that  $\max_i Y'_i(\omega, t) \ge 0$  whenever  $(\omega, t) \notin A$ . Consequently,  $\mathbb{P}'[\max\{Y'_1, \dots, Y'_n\} \ge 0] \ge 1 - \mathbb{P}'[A] = \mathbb{E}[\mathbf{e}^\top \mathbf{w}]$ . If there exists  $\mathbf{Y} \leq_{cx} \mathbf{X}$  such that  $\mathbb{P}[\max\{Y_1, \dots, Y_n\} \ge 0] = \mathbb{P}'[\max\{Y'_1, \dots, Y'_n\} \ge 0]$ , then, since  $W_1, \ldots, W_n$  in the feasible set of (52) are arbitrary, this would imply that  $\overline{F}_{3}(\mathbf{X}) = \sup_{\mathbf{Y} \leq_{c_{1}} \mathbf{X}} \mathbb{P}[\max\{Y_{1}, \dots, Y_{n}\} \ge 0] \ge \sup_{W \in \mathcal{W}} \mathbb{E}[\mathbf{e}^{\mathsf{T}}\mathbf{w}] = \overline{F}_{2}(\mathbf{X}).$ Indeed.  $\mathbf{Y}' = \mathbb{E}[\mathbf{X}'|\mathcal{F}']$  implies that  $\mathbf{X}' \ge_{cx} \mathbf{Y}'$ . Since  $\Omega$  is atomless, there exists a pair of random vectors  $\mathbf{X}'' \in L^{1,n}(\Omega)$  and  $\mathbf{Y} \in L^{1,n}(\Omega)$  with the same joint distribution as  $\mathbf{X}'$ and **Y**'. Then  $\mathbb{P}[\max\{Y_1, \dots, Y_n\} \ge 0] = \mathbb{P}'[\max\{Y'_1, \dots, Y'_n\} \ge 0]$  and **X**''  $\ge_{cx}$  **Y**. Since **X** and **X**'' have the same distribution,  $\mathbf{X} \ge_{cx} \mathbf{Y}$ , as required.

Finally,  $\overline{F}_2(\mathbf{X}) \ge \overline{F}_1(\mathbf{X})$  follows from the fact that (51) is the problem dual to (52).

#### D. The algorithm for solving problem (28)

Algorithm 1 (bCDF and rCDF in objective).

Step 1. Set  $\mathbf{a} = (1, ...1)$ ,  $\epsilon = \text{small number}$ , I = F or H. Step 2. Solve  $\min_{\mathbf{w} \in \mathbf{W}} \overline{p}_{g1(\mathbf{a}, \mathbf{w})}(0)$ : let  $\widetilde{\mathbf{w}}$  be its optimal solution. Step 3. Solve  $\min_{\mathbf{a} \in \mathbf{A}} \overline{p}_{g1(\mathbf{a}, \widetilde{\mathbf{w}})}(0)$ : let  $\widetilde{\mathbf{a}}$  be its optimal solution. Set  $\widetilde{p} = \overline{p}_{g1(\widetilde{\mathbf{a}}, \widetilde{\mathbf{w}})}(0)$ . Step 4. Use  $\widetilde{\mathbf{w}}$  as an initial approximation to solve  $\min_{\mathbf{w} \in \mathbf{W}} \overline{p}_{g1(\widetilde{\mathbf{a}}, \mathbf{w})}(0)$ : let  $\mathbf{w}^*$  be its optimal solution. Step 5. Use  $\widetilde{\mathbf{a}}$  as an initial approximation to solve  $\min_{\mathbf{a} \in \mathbf{A}} \overline{p}_{g1(\widetilde{\mathbf{a}}, \mathbf{w})}(0)$ : let  $\mathbf{a}^*$  be its optimal solution. Set  $p^* = \overline{p}_{g1(\mathbf{a}^*, \mathbf{w}^*)}(0)$ . Step 6. If  $\widetilde{p} - p^* < \epsilon$ , then Stop, otherwise set  $\widetilde{\mathbf{w}} = \mathbf{w}^*$ ,  $\widetilde{\mathbf{a}} = \mathbf{a}^*$ , and  $\widetilde{p} = p^*$ , and go to Step 4.

### E. The algorithm for solving problem (37)

Algorithm 2 (bCDF and rCDF in constraint).

Step 1. Set  $\mathbf{a} = (1, ...1)$ ,  $\epsilon =$  small number.

Step 2. Solve (37) with respect to w: let  $\tilde{w}$  be its optimal solution. Set  $\tilde{v} = V(\tilde{w})$ .

Step 3. Solve (38) and (39) with  $\mathbf{w} = \widetilde{\mathbf{w}}$ :let  $\widetilde{\mathbf{a}}_F$  and  $\widetilde{\mathbf{a}}_H$  be optimal solutions of (38) and (39), respectively.

Step 4. Use  $\tilde{\mathbf{w}}$  as an initial approximation to solve (37) with respect to  $\mathbf{w}$  with  $\mathbf{a}_F = \tilde{\mathbf{a}}_F$  and  $\mathbf{a}_H = \tilde{\mathbf{a}}_H$ : let  $\mathbf{w}^*$  be its optimal solution. Set  $v^* = V(\mathbf{w}^*)$ .

Step 5. Use  $\tilde{\mathbf{a}}$  as an initial approximation to solve (38) and (39) with  $\mathbf{w} = \mathbf{w}^*$ : let  $\mathbf{a}_F^*$  and  $\mathbf{a}_H^*$  be optimal solutions of (38) and (39), respectively.

Step 6. If  $\tilde{v} - v^* < \epsilon$ , then Stop, otherwise set  $\tilde{\mathbf{w}} = \mathbf{w}^*$ ,  $\tilde{\mathbf{a}}_F = \mathbf{a}_F^*$ , and  $\tilde{\mathbf{a}}_H = \mathbf{a}_H^*$  and  $\tilde{v} = v^*$ , and go to Step 4.

**Acknowledgements** We are grateful to Mr. Edward Cummings from the Stony Brook University for the numerical implementation of the case study. We are also grateful to the referees for their comments and suggestions, which helped to improve the quality of the paper.

**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/.

### References

- 1. Bertsekas, D.: Nonlinear programming. J. Oper. Res. Soc. 48(3), 334 (1997)
- Grechuk, B., Molyboha, A., Zabarankin, M.: Chebyshev's inequalities with law invariant deviation measures. Probab. Eng. Inf. Sci. 24, 145–170 (2010)
- Grippof, L., Sciandrone, M.: Globally convergent block-coordinate techniques for unconstrained optimization. Optim. Methods Softw. 10(4), 587–637 (1999)
- Grippof, L., Sciandrone, M.: On the convergence of the block nonlinear Gauss-Seidel method under convex constraints. Oper. Res. Lett. 26(3), 127–136 (2000)
- Lyu, H.: Convergence and complexity of block coordinate descent with diminishing radius for nonconvex optimization, arXiv:2012.03503 (2020)
- Mafusalov, A., Uryasev, S.: Buffered probability of exceedance: mathematical properties and optimization. SIAM J. Optim. 28(2), 1077–1103 (2018)
- Norton, M., Uryasev, S.: Maximization of AUC and buffered AUC in binary classification. Math. Program. 174(1–2), 575–612 (2019)
- Pertaia, G., Prokhorov, A., Uryasev, S.: A new approach to credit ratings. J. Bank. Finance 140, 106097 (2022)
- 9. Pinelis, I.: On the convexity of certain set of random vectors. Mathoverflow, https://mathoverflow. net/questions/415870/
- 10. Powell, M.: On search directions for minimization algorithms. Math. Program. 4(1), 193–201 (1973)
- 11. Rockafellar, R.T.: Convex Analysis, p. 451. Princeton University Press, Princeton, NJ (1970)
- 12. Rockafellar, R.T., Royset, J.O.: Random variables, monotone relations, and convex analysis. Math. Program. **148**, 297–331 (2014)
- 13. Rockafellar, R.T., Royset, J.O.: On buffered failure probability in design and optimization of structures. Reliab. Eng. Syst. Saf. **95**, 499–510 (2011)
- 14. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk 2, 21-42 (2000)
- 15. Rockafellar, R.T., Uryasev, S.: Conditional value-at-Risk for general loss distributions. J. Bank. Finance **26**, 1443–1471 (2002)
- Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Generalized deviations in risk analysis. Finance Stoch. 10(1), 51–74 (2006)
- 17. Veremyev, A., Tsyurmasto, P., Uryasev, S.: Optimal structuring of CDO contracts: optimization approach. J. Credit Risk 8(4), 133–155 (2012)
- Zabarankin, M., Uryasev, S.: Statistical Decision Problems: Selected Concepts and Portfolio Safeguard Case Studies. Springer, Berlin (2014)

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