Geometric Constructions with Discretized Random Variables

Hans-Peter Schröcker*   Johannes Wallner†

July 2005

We generalize the DEnv (Distribution envelope determination) method for bounding the result of arithmetic operations on random variables with unknown dependence to higher-dimensional settings. In order to minimize both the influence of the coordinate frame and information loss we suggest a nested thicket representation for random variables and a corresponding intersection algorithm.

1. Introduction

Algorithms computing with discretized random variables have been presented in a number of publications Williamson and Downs [1990], Berleant and Goodman-Strauss [1998], Berleant et al. [2003], Berleant and Zhang [2004b,a], Regan et al. [2004]. However, the algebra of imprecise numbers in its various senses often is a poor model for geometry with imprecisely defined objects. A vector of imprecise numbers, for instance, usually defines an interval-like geometric object (a bounding box, a set of bounding boxes with probability measures assigned etc.), which is not invariant with respect to geometric transformations. Furthermore accurate representation of a geometric shape may require a large number of intervals.

Using the idea of *tolerance zone* Requicha [1983] and replacing intervals by spheres, ellipsoids or general convex sets when dealing with geometric constructions Pottmann et al. [2000], Wallner et al. [2000] provides more flexibility and at the same times allows a relatively simple geometric treatment. The computational complexity of convex sets in general is, of course, higher than that of intervals. In the 2D case it is still reasonably low.

In this article we extend the DEnv method to higher-dimensional settings with convex sets as fundamental objects. DEnv stands for "distribution envelope determination" and is a method for bounding the results of arithmetic operations on discretized random variables. It has been introduced in Berleant and Goodman-Strauss [1998], an implementation is described in Berleant et al. [2003]; its relation to other methods for dealing with probabilistic errors in algebraic computations is the topic of Regan et al. [2004]. It is shown that these methods are equivalent to a great extent.

In univariate DEnv, a family of random variables is represented either by a set of intervals with associated probability masses (a *thicket*) or by a pair of upper and lower bounds on the random variables' cumulative distribution function (CDF). Both representations can be converted into each other. The result of an arithmetic operation on random variables is bounded

---

*Institut für Technische Mathematik, Geometrie und Bauinformatik, Universität Innsbruck.
†Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien.

by pointwise optimal distribution bounds that are found by linear programming techniques. In the original version, no assumption on the mutual dependence of the random variables is made, of course at the cost of wider distribution bounds. But it is possible to incorporate knowledge of dependence, for example independence or correlation Berleant and Zhang [2004a,b].

A generalization of DEnv to multivariate settings is more or less straightforward. We do this in Section 3, after a short introduction to the univariate DEnv algorithm in Section 2. The resulting multivariate DEnv method, however, is inherently information losing because multivariate thickets and their distribution bounds are no longer equivalent. In addition, the algorithm's outcome depends on the chosen coordinate system or, equivalently, on the position and orientation of the input data: Denote the output of multivariate DEnv applied to a certain set of input data by T and consider the output T' of multivariate DEnv applied to a rotated copy of the same input data. In general, T' is not a rotated copy of T, which is undesirable in geometric applications.

In order to minimize the loss of information and to weaken the influence of the coordinate system, we introduce the concept of nested thickets and an intersection algorithm for them in Section 4. Not all problems related to multivariate thickets can be solved by these methods, but we argue that they are good enough to make thickets a valuable tool in geometric constructions. The comprehensive example of Section 5 illustrates our ideas.

2. Univariate thickets

This section is meant as a short introduction to univariate thickets. For more details see Berleant and Goodman-Strauss [1998], Regan et al. [2004], Berleant and Zhang [2004a].

Definition 1. A *univariate thicket* is a finite set $T = \{(C_i, p_i)\}$ whose elements are pairs of nonempty open intervals $C_i = (a_i, b_i) \subset \mathbb{R}$ and associated probability values p_i such that $\sum p_i = 1$. Upper and lower distribution bounds of the thicket T are the cumulative distribution functions

$$U(t) := \sum_{a_i \leq t} p_i, \quad L(t) := \sum_{b_i \leq t} p_i. \tag{1}$$

A thicket can be thought of as a representation of a family $\mathcal{X}$ of random variables. We say that a random variable x, defined over the Borel sigma algebra $\mathcal{B}$ of $\mathbb{R}$, is *represented* by the thicket T if the cumulative distribution function F of x has the lower bound L and the upper bound U. We denote this relation by $x \lhd T$; the family $\mathcal{X}$ may be written as $\{x \mid x \lhd T\}$.

Since a univariate random variable is determined by its cumulative distribution function, it seems appropriate to use the CDF for defining the '$\lhd$'-relation. The reader however should be aware of the fact that this notion might not always conform with the intuition behind the concept of thickets. The following example was pointed out by the anonymous reviewer: Consider the random variable x concentrated on the points 1 and 4, with probability 0.5 each. Even if no probability is assigned to the interval $(2, 3)$, x is represented by the thicket $\{((1, 4), 0.5), ((2, 3), 0.5)\}$.

2.1. **Construction and deconstruction of distribution envelopes.** The graphs of the distribution envelopes U and L of a thicket T can be constructed geometrically (Figure 1). In order to find the upper envelope U we assume that the intervals C_i are numbered so that $a_i \leq a_{i+1}$. We visualize the pairs (C_i, p_i) by rectangles of height p_i, erected over the interval C_i. Beginning with C_1, we translate the rectangles C_i in y-direction so that its bottom edge is at the height of the top edge of C_{i-1}. The left hand vertices of these rectangles define the graph of U. The lower distribution envelope L is found by piling up the same rectangles after sorting them in ascending order by right hand interval borders b_i.

The inverse procedure can be used for deconstructing a pair of suitable distribution bounds into a thicket. Since we come back to this task in a more general context in Section 3, we will not
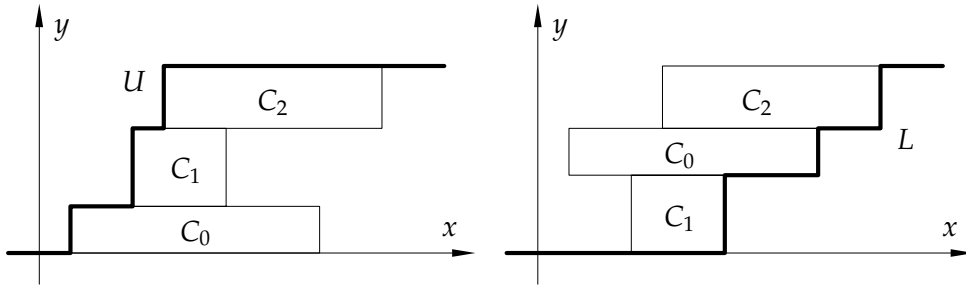
Figure 1: Construction of distribution bounds.

go into details here. The interested reader can find a formal description of the deconstruction algorithm for univariate thickets in [Regan et al., 2004, Section 4]. We only mention that the deconstruction of distribution bounds into a thicket is not necessarily unique (see Example 1), but always information preserving (i.e. is an invertible operation).

2.2. **Univariate DEnv.** Distribution envelope determination Berleant and Goodman-Strauss [1998] is a method for bounding the result of a binary arithmetic operation "$\star$" on random variables without making assumptions on their mutual dependence. As a prerequisite, the operation has to be extended to nonempty intervals. Then, the input random variables are discretized as thickets and DEnv computes distribution bounds on the resulting random variable, which is subsequently deconstructed into a thicket.

Algorithm 1 (Univariate DEnv).

Input: Two thickets $T' = \{(C'_i, p'_i)\}$, $T'' = \{(C''_j, p''_j)\}$ and a binary operation "$\star$", defined on $\mathbb{R}$ and extended to the case of nonempty open intervals as arguments.

Output: A thicket representing all random variables $x' \star x''$ where $x' \lhd T'$ and $x'' \lhd T''$. No assumption on the dependence of x' and x'' is made.

Step 1: Compute the intervals $C_{ij} = (a_{ij}, b_{ij}) := C'_i \star C''_j$. The intervals C_{ij} will be the thicket intervals, but their probability masses are still unknown.

Step 2: Compute the upper distribution bound U by solving the optimization problem

$$U(t) = \max \sum_{(i,j):\, a_{ij} \leq t} p_{ij}, \tag{2}$$

for all $t \in \mathbb{R}$, where the unknowns p_{ij} are subject to the constraints

$$\sum_i p_{ij} = p'_i, \quad \sum_j p_{ij} = p''_j \quad \text{and} \quad 0 \leq p_{ij}. \tag{3}$$

Step 3: Compute the lower distribution bound L of T by solving the optimization problem

$$L(t) = \min \sum_{(i,j):\, b_{ij} \leq t} p_{ij}, \tag{4}$$

for all $t \in \mathbb{R}$; where the unknowns p_{ij} are subject to the constraints (3).

Step 4: Deconstruct U and L into the resulting thicket T.

Remark 1. The optimization problem in Steps 2 and 3 can be solved efficiently by linear programming, even for a large number of variables Karloff [1991]. Since U and L are *simple functions* (piecewise constant with finite range), it is sufficient to compute the values of $L(t)$ and $U(t)$ at a finite number of suitably chosen arguments $t \in \mathbb{R}$. Hence the algorithm's implementation is straightforward once libraries for linear programming and interval arithmetic are available.

For the computation of U one assumes a concentration of the probability mass p_{ij} at a_{ij}; for the computation of L one assumes a concentration at b_{ij}. This and the chosen optimization targets (minimum or maximum) ensure the correctness and pointwise optimality of the algorithm.

Dependence information formulated in terms of linear equalities or inequalities is easily incorporated in the univariate DEnv algorithm. An example of this is the Pearson correlation coefficient Berleant and Zhang [2004b]. Furthermore, it is obvious how to extend the algorithm to operations with more than two operands.

3. Multivariate thickets

In this section we introduce multivariate thickets. Even if this concept is in some sense a straightforward generalization of the univariate case, there are also essential differences.

Definition 2. A *multivariate thicket* is a finite set $T = \{(C_i, p_i)\}$ whose elements are pairs of nonempty open convex sets $C_i \subset \mathbb{R}^d$ and associated probability values p_i such that $\sum p_i = 1$.

Definition 3. The *lower and upper probability $\underline{p}, \overline{p}$* of a thicket $T = \{(C_i, p_i)\}$ are the functions $\underline{p}$ and $\overline{p}$ defined on the sigma algebra $\mathcal{B}_d$ of Borel sets of $\mathbb{R}^d$:

$$\underline{p}: X \in \mathcal{B}_d \mapsto \sum_{C_i \subset X} p_i, \qquad \overline{p}: X \in \mathcal{B}_d \mapsto \sum_{C_i \cap X \neq \varnothing} p_i, \tag{5}$$

Remark 2. The functions $\underline{p}$ and $\overline{p}$ satisfy the axioms of interval probability Walley [1996]:

1. $\forall X \in \mathcal{B}_d: \ 0 \leq \underline{p}(X) \leq \overline{p}(X) \leq 1$;

2. $\underline{p}(\mathbb{R}^d) = \overline{p}(\mathbb{R}^d) = 1, \ \underline{p}(\varnothing) = \overline{p}(\varnothing) = 0$;

3. $\overline{p}(X) + \underline{p}(\mathbb{R}^d \setminus X) = 1$;

4. $\forall X, Y \in \mathcal{B}_d: \ X \cap Y = \varnothing \implies$
   $\underline{p}(X) + \underline{p}(Y) \leq \underline{p}(X \cup Y) \leq \underline{p}(X) + \overline{p}(Y) \leq \overline{p}(X \cup Y) \leq \overline{p}(X) + \overline{p}(Y).$

Definition 4. A random variable x defined over $\mathcal{B}_d$ is *represented* by a thicket T, if $\underline{p}(X) \leq P(x \in X) \leq \overline{p}(X)$ holds for all $X \in \mathcal{B}_d$. We denote this relation by $x \lhd T$. Two thickets T_1, T_2 are called *equivalent* if they represent the same set of random variables. If every random variable represented by T_1 is also represented by T_2, we call T_1 a *subthicket* of T_2.

Remark 3. In the univariate case, the $\lhd$-relation between random variables and thickets is defined via the thicket's distribution envelopes. There exists a multivariate generalization of distribution envelopes (see Section 3) but it cannot be used for an analogous definition. The reason for this will become clear in Example 2.

Example 1. There is a trivial method for constructing a thicket equivalent to a given thicket $T = \{(C_i, p_i)\}$. We replace one thicket element (C_i, p_i) by the two new elements (C_i, ε), $(C_i, p_i - \varepsilon)$ where $0 < \varepsilon < p_i$. We call every thicket obtained from T by a series of these operations a *partition thicket* of T. Clearly two thickets are equivalent if they have a common partition thicket. In this case we call them *essentially identical*. The converse is not true, i.e., the same family of random variables can be represented by essentially different thickets. An example of this is given by two equivalent thickets

$$T_1 = \{((0,2), 0.3), ((1,3), 0.7)\}, \ T_2 = \{((0,3), 0.3), ((1,3), 0.4), ((1,2), 0.3)\}. \tag{6}$$
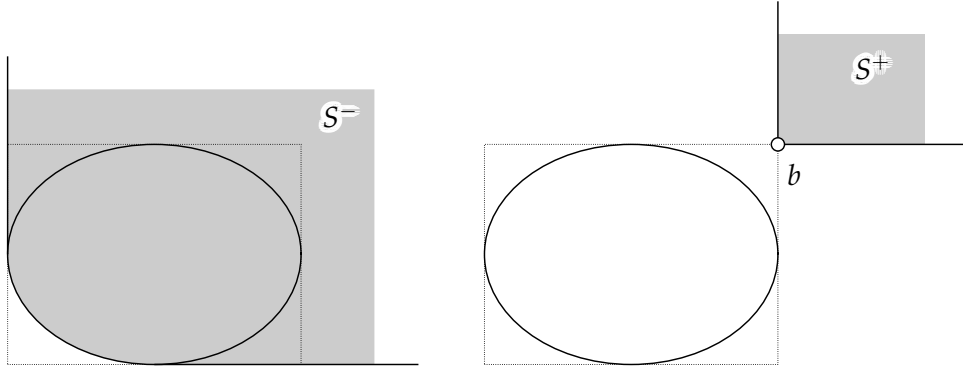
Figure 2: Definition of S^+ and S^-.

3.1. Distribution envelopes.

Definition 5. For an arbitrary subset S of $\mathbb{R}^d$ we define S^+ as the set of all points $(y_1, \ldots, y_d) \in \mathbb{R}^d$ such that $x_i \leq y_i$ holds for all points $(x_1, \ldots, x_d) \in S$ and all indices $i \in \{1, \ldots, d\}$. The set S^- is defined as $\bigcup_{x \in S} \{x\}^+$. When using coordinates with respect to a basis A of $\mathbb{R}^d$, we write $S^+(A)$ and $S^-(A)$.

Examples of S^+ and S^- are depicted in Figure 2. In 2D, $S^+ = b^+$ where b is the upper right corner of the bounding box.

Definition 6. The *upper and lower distribution envelope* of a thicket $T = \{(C_i, p_i)\}$ are the functions

$$U_T \colon \mathbb{R}^d \to [0, 1], \ t \mapsto \sum_{t \in C_i^-} p_i, \qquad L_T \colon \mathbb{R}^d \to [0, 1], \ t \mapsto \sum_{t \in C_i^+} p_i. \tag{7}$$

When using coordinates with respect to the basis A of $\mathbb{R}^d$, we write $L_T(A, t)$ and $U_T(A, t)$.

With the definitions $(u_1, \ldots, u_d) \leq (t_1, \ldots, t_d) \iff u_1 \leq t_1, \ldots, u_d \leq t_d$ and $t^\leq := \{u \in \mathbb{R}^d \mid u \leq t\}$, the MDF of x is defined as $F \colon \mathbb{R}^d \to \mathbb{R}$, $F(t) := P(x \leq t)$, i.e., the probability assigned to the set $t^\leq$. Definition 6 is the multivariate analogue to Definition 1. The definition of L_T and U_T ensures that they enclose the multivariate distribution function (MDF) of every random variable $x \lhd T$, regardless of the distribution of probability mass inside the convex sets C_i. For the definition of U_T, we assume a concentration of probability mass at points of the "lower left" border of C_i; for the definition of L_T we assume a concentration at points with maximal coordinate values.

Proposition 1. *The multivariate distribution function F of a random variable $x \lhd T$ has the property that $L_T(t) \leq F(t) \leq U_T(t)$ for all $t \in \mathbb{R}^d$.*

Proof. The proof is a matter of combining the notions defined above (Definitions 3, 4 and 6): $x \lhd T$ implies that for all t:

$$F(t) = P(x \leq t) \leq \overline{p}(t^\leq) = \sum_{i \colon (C_i \cap t^\leq) \neq \varnothing} p_i = \sum_{i \colon \exists u \in C_i \colon u \leq t} p_i = \sum_{i \colon t \in C_i^-} p_i = U_T(t). \tag{8}$$

The inequality $F(t) \geq L_T(t)$ is shown in an analogous way. □

The converse is not true, as demonstrated by the next example. Thus representing a thicket by distribution bounds loses information.

Example 2. Consider the one-element thickets $T = \{(C, 1)\}$, $T' = \{(C', 1)\}$, and $T'' = \{(C'', 1)\}$, where C, C' and C'' are bounded by arcs and straight lines as illustrated in Figure 3. Each thicket represents the random variables concentrated in the corresponding convex set, but the
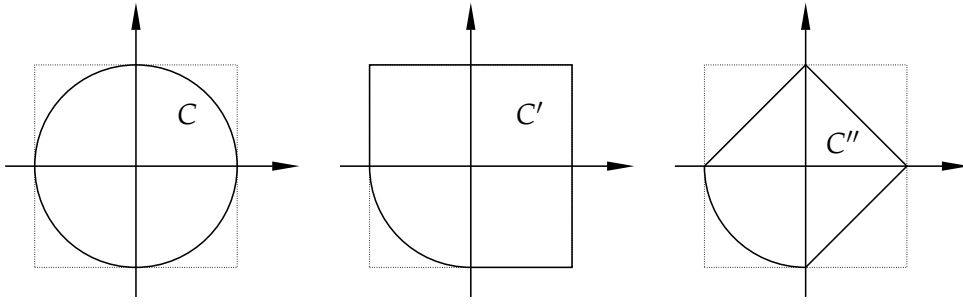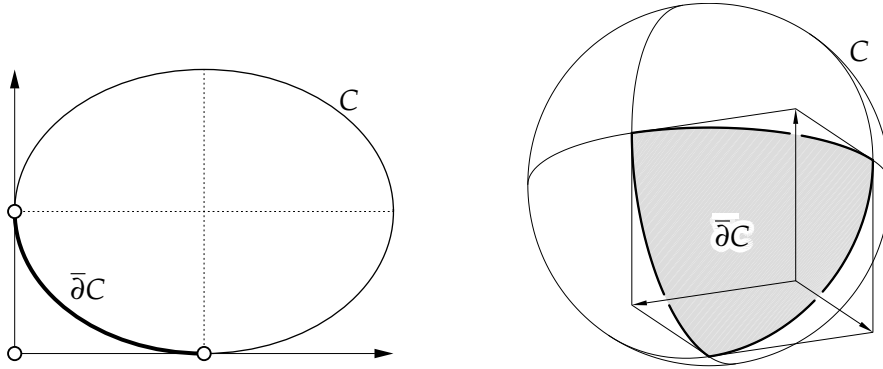
Figure 3: One-element thickets with identical distribution bounds.

Figure 4: Definition of $\overline{\partial}C$ in 2D (left) and 3D (right).

upper and lower distribution bounds L and U are the same for all three. T' represents precisely the random variables whose MDF lies between L and U, and C'' is the smallest convex set such that the corresponding one-element thicket has these lower and upper distribution bounds. If the loss of information is measured in terms of the area, then C' is 20% larger than C and 65% larger than C''.

3.2. **Deconstruction of distribution envelopes.** The deconstruction of a pair U, L of distribution envelopes into a thicket is a fundamental step of DEnv. The basic ideas are the same in both the univariate and the multivariate settings. We give a formal definition of the corresponding algorithm.

Definition 7. Assume that $F : \mathbb{R}^d \to [0,1]$ is a simple function with range$(F) = \{f_0, \ldots, f_n\}$ such that $0 = f_0 < f_1 < \cdots < f_n = 1$. We consider the level sets $[F = f_i]$ $(i = 0, \ldots, n)$, and for $z \in \mathbb{R}$ we define the level-type sets $[F \approx z] := \overline{[F \geq z]} \cap \overline{[F < z]}$ (here $\overline{X}$ denotes topological closure of X).

Definition 8. The bounding box of a subset $S \subset \mathbb{R}^d$ is the set BB$(S) := \pi_1(S) \times \cdots \times \pi_d(S)$ where $\pi_j(x_1, \ldots, x_d) := x_j$ is the projection onto the j-th coordinate. If C is a convex set with boundary ∂C, then $\overline{\partial}C$ is the set of all points $x \in \partial C$ which have a supporting hyperplane with exterior normal vector $n = (n_1, \ldots, n_d)$ such that $n_1, \ldots, n_d \leq 0$. Further we define $\eta\, C := C^- \cap$ BB(C). The convex hull of $M \subseteq \mathbb{R}^d$ is denoted by the symbol CH(M).

The definition of $\overline{\partial}C$ is illustrated in Figure 4. Examples of $\eta\, C$ are found in Figure 3, where we have $C' = \eta\, C = \eta\, C''$. Note that

$$C \text{ open and convex} \implies \text{BB}(\eta\, C) = \text{BB}(C),\ (\eta\, C)^+ = C^+,\ (\eta\, C)^- = C^-. \tag{9}$$

These properties are used in the following algorithm, which is similar to the deconstruction algorithm for univariate thickets Regan et al. [2004].

Algorithm 2 (Thicket deconstruction).

Input: Two simple functions L, U with the properties $L(t) \leq U(t)$ for all $t \in \mathbb{R}^d$, $0, 1 \in$ range$(L) \cap$ range(U), $u \leq t \implies L(u) \leq L(t), U(u) \leq U(t)$ (i.e., both L and U are multivariate distribution functions).

Output: Thicket T representing all random variables whose MDF F fulfills $L \leq F \leq U$.

Step 1: Let $Z :=$ range$(L) \cup$ range(U) and sort the members of Z in ascending order: $0 = z_0 < \cdots < z_n = 1$.

Step 2: For $i = 1, \ldots, n$ let

$$C_i = \eta \, \mathrm{CH} \left(\bar{\partial} \mathrm{CH} \left[U \approx z_i \right] \cup \bar{\partial} \mathrm{CH} \left[L \approx z_i \right] \right) \tag{10}$$

(see Figure 5) and define $p_i := z_i - z_{i-1}$.

Step 3: Let $T = \{ (C_i, p_i) \mid i = 1, \ldots, n \}$.

Proof (of correctness of Algorithm 2). We claim that for a random variable x whose distribution function F fulfills $L(t) \leq F(t) \leq U(t)$ for all $t \in \mathbb{R}^d$, we have $x \lhd T$. First we observe that

$$C \subseteq C' \implies C^+ \supseteq (C')^+, \eta \, C \subseteq \eta \, C'; \quad C_i \subseteq C_{i,L} := \eta \, \mathrm{CH}(\bar{\partial} \, \mathrm{CH} \left[L \approx z_i \right]). \tag{11}$$

The "lower left corners" of $[L \approx z]$ and $[U \approx z]$ determine the shape of these sets:

$$(\mathrm{CH} \, \bar{\partial} \, \mathrm{CH} \left[U \approx z \right])^- = \mathrm{CH} \left[U \approx z \right], \quad t \in (\mathrm{CH} \, \bar{\partial} \, \mathrm{CH} \left[L \approx z \right])^+ \implies L(t) \geq z \tag{12}$$

(cf. Figures 2 and 4). We aim at the inequality $L_T(t) \leq L(t)$ and use (11) to compute

$$L_T(t) = \textstyle\sum_{i:\, t \in C_i^+} p_i \leq \sum_{i:\, t \in C_{i,L}^+} p_i = \sum_{i:\, t \in C_{i,L}^+} p_i. \tag{13}$$

By (9) and (12), $t \in C_{i,L}^+$ implies that $L(t) \geq z_i$. Thus

$$L_T(t) \leq \textstyle\sum_{i:\, L(t) \geq z_i} (z_i - z_{i-1}) = \max\{z_j \mid L(t) \geq z_j\} \leq L(t). \tag{14}$$

As to U_T, we define $C_{i,U}$ analogous to (11) and observe $L \leq U \implies C_{i,L}^- \subseteq C_{i,U}^-$:

$$\begin{aligned} U_T(t) &= \textstyle\sum_{i:\, t \in C_i^-} p_i = \sum_{i:\, t \in C_{i,U}^-} p_i = \sum_{i:\, t \in \mathrm{CH}[U \approx z_i]} p_i \\ &\geq \textstyle\sum_{i:\, t \in [U \approx z_i]} p_i = \max\{z_i \mid t \in [U \geq z_i]\} = U(t). \end{aligned} \tag{15}$$

With $L_T \leq L$ and $U \leq U_T$ as shown above and the assumption $L \leq F \leq U$ as required, we have shown $L_T \leq F \leq U_T$, i.e., $x \lhd T$. $\qquad\square$

In general a thicket T_0 is only a subthicket (c.f. Definition 4) of the thicket T constructed by Algorithm 2 from U_{T_0} and L_{T_0}, i.e., $x \lhd T$ does not imply $x \lhd T_0$. This is in contrast to the univariate case.

Definition 9. If finite sets $Z \subseteq Z' \subset \mathbb{R}$ have the property that $\min Z = \min Z'$ and $\max Z = \max Z'$, then Z' is called a *partition* of Z.

If the set Z in Step 1 of Algorithm 2 is replaced by a partition Z', the output thicket will be a partition thicket of the original one. Now we have provided all tools to formulate the multivariate version of the DEnv algorithm.
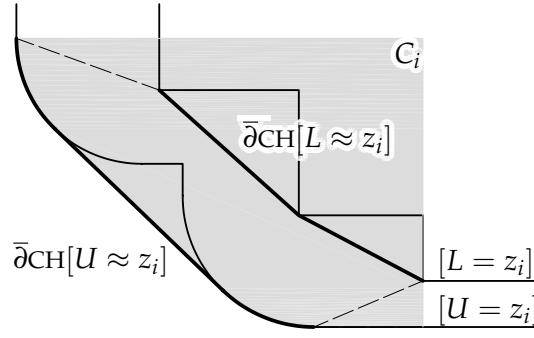
Algorithm 3 (Multivariate DEnv).

Figure 5: Constructing a thicket element C_i (shaded area) from level sets of U and L.

Input: Two thickets $T' = \{(C_i', p_i')\}$, $T'' = \{(C_i'', p_i'')\}$ and a binary operation "$\star$", defined on $\mathbb{R}^d$ and extended to convex sets in the sense that $t \in C$, $t' \in C' \implies t \star t' \in C \star C'$ and $C \star C'$ is convex.

Output: A thicket representing all random variables $x' \star x''$ where $x' \lhd T'$ and $x'' \lhd T''$. No assumption on the dependence of x' and x'' is made.

Step 1: Compute $C_{ij} = C_i' \star C_j''$.

Step 2: Compute the upper distribution bound U by solving the optimization problem

$$U(t) := \max \sum_{t \in C_{ij}^-} p_{ij}, \tag{16}$$

for each $t \in \mathbb{R}$, where the unknowns p_{ij} are subject to the constraints (3).

Step 3: Compute the lower distribution bound L of T by solving the optimization problem

$$L(t) := \min \sum_{t \in C_{ij}^+} p_{ij}. \tag{17}$$

for each $x \in \mathbb{R}$, where the unknowns p_{ij} are subject to the constraints (3).

Step 4: Deconstruct U and L into the resulting thicket T.

Proof (of correctness of Algorithm 3). The function U in (16) is an upper bound of the MDF of all random variables $x = x' \star x''$ because the level-type sets $[U \approx z_i]$ enclose the corresponding level sets of the MDF of x and the corresponding function values are maximal, subject to the necessary constraint (3). Analogously it follows that L is a lower bound of the MDF of x. $\square$

The basic steps of this algorithm are similar to the univariate DEnv algorithm (Algorithm 1). However, the result needs a different interpretation:

- The multivariate DEnv algorithm is information losing. This is due to the information losing intermediate representation of a thicket by its distribution bounds.

- The output of multivariate DEnv depends on the coordinate frame.

Both issues are drawbacks for the use of DEnv in geometric constructions. The loss of information in the DEnv algorithm can be reduced by a certain amount if we choose an appropriate basis of $\mathbb{R}^d$. In higher dimensions it apparently is not easy to minimize $\int (U - L)^2$ or a similar target functional over all possible orthonormal bases of $\mathbb{R}^d$ (in case we work in Euclidean geometry). Therefore we present an improved version of multivariate DEnv in the next section. The main idea is to combine DEnv results that were obtained with respect to different coordinate frames of $\mathbb{R}^d$.

4. Nested thickets and thicket intersection

In this section we introduce nested thickets and thicket intersection in order to reduce the shortcomings of the multivariate DEnv algorithm.

Definition 10. A thicket $T = \{(C_i, p_i) \mid i = 0, \ldots, n\}$ is called *nested* if there exists a permutation π of $\{0, \ldots, n\}$ such that $C_{\pi(i+1)} \subset C_{\pi(i)}$.

We tacitly assume that for a nested thicket $\{(C_i, p_i)\}$ the sets C_i are numbered such that $C_{i+1} \subseteq C_i$. The lower and upper probability of a nested thicket are of simple shape. If $X \in \mathcal{B}_d$, we have

$$\underline{p}(X) = \sum_{i=\underline{m}}^{n} p_i, \qquad \overline{p}(X) = \sum_{i=0}^{\overline{m}} p_i \tag{18}$$

where $\underline{m} := \min\{i \mid X \subset C_i\}$ and $\overline{m} := \max\{i \mid X \cap C_i \neq \varnothing\}$. In particular this implies

$$P(x \in C_i) \geq \sum_{j=i}^{n} p_j \quad \text{and} \quad P(x \in \mathbb{R}^d \setminus C_i) \leq \sum_{j=0}^{i-1} p_j. \tag{19}$$

Definition 11. The thicket T is the *intersection* of the thickets T_1 and T_2 if

$$x \lhd T \iff (x \lhd T_1) \text{ and } (x \lhd T_2). \tag{20}$$

The intersection of T_1 and T_2 is denoted by $T_1 \cap T_2$.

Note that the intersection of two thickets does not always exist but univariate thickets T_1, T_2 can always be intersected: If U_i and L_i are the distribution envelopes of T_i, we can deconstruct $T = T_1 \cap T_2$ from its distribution envelopes

$$U := \min\{U_1, U_2\}, \quad L := \max\{L_1, L_2\}. \tag{21}$$

In the same way it is possible to intersect multivariate thickets but the result will usually be only a superset of intersection. Note further that the distribution bounds U_i and L_i have to refer to the same basis of $\mathbb{R}^d$.

In contrast to general multivariate thickets, the intersection of nested thickets exists. The intersection algorithm is based on the representation of nested thickets via *nested stacks*:

Definition 12. For $i = 0, \ldots, n$ assume that $B_i \subset \mathbb{R}^{d+1}$ is the cylinder $C_i \times (0, h_i)$ with a nonempty convex domain C_i as basis. If $C_{i+1} \subseteq C_i$ and $h_{i+1} \geq h_i$, the union $\bigcup_{i=0}^{n} C_i$ is called a *nested stack*.

For a nested thicket $T = \{(C_i, p_i) \mid i = 0, \ldots, n\}$ we may construct the nested stack defined by cylinders $B_i = C_i \times (0, h_i)$ with $h_i = \sum_{j=0}^{i} p_j$. Conversely a nested stack defined by cylinders $C_i \times (0, h_i)$ can be deconstructed into a nested thicket $\{(C_i, p_i) \mid i = 0, \ldots, n\}$ by recursively defining $p_0 = h_0$ and $p_i = h_i - h_{i-1}$ for $i > 0$. The deconstruction of the nested stack into a thicket does not necessarily reproduce the original thicket, but at least an essentially identical thicket (cf. Example 1). Therefore we can represent nested thickets by nested stacks and vice versa.

Lemma 1. *The intersection of the two thickets* $T_1 = \{(C_1, 1)\}$ *and* $T_2 = \{(C_2, 1)\}$ *is the thicket* $T_3 = \{(C_3, 1)\}$ *with* $C_3 = C_1 \cap C_2$.

Proof. The case $C_1 = C_2$ being trivial, we assume $C_1 \neq C_2$. Since T_3 is a subthicket of T_1 and T_2, every random variable represented by T_3 is also represented by T_1 and T_2. We have to show the converse. Assume that $x \lhd T_1$, $x \lhd T_2$, but $x \not\lhd T_3$. We denote the lower and upper probabilities of T_i by $\underline{p}_i$ and $\overline{p}_i$. There exists $X \in \mathcal{B}_d$ such that $\underline{p}_i(X) \leq P(x \in X) \leq \overline{p}_i(X)$ for $i = 1, 2$ and either $\underline{p}_3(X) > P(x \in X)$ (case 1) or $\overline{p}_3(X) < P(x \in X)$ (case 2), where $P(x \in X)$ is the probability of x being contained in X.

In case 1, $x \lhd T_2$ and $\underline{p}_2(C_1 \setminus C_2) = \overline{p}_2(C_1 \setminus C_2) = 0$ imply $P(x \in C_1 \setminus C_2) = 0$, and $x \lhd T_1$, $\underline{p}_1(C_1) = 1$ imply $P(x \in C_1) = 1$. Thus

$$P(x \in C_3) = P(x \in C_1 \setminus C_2) + P(x \in C_3) = P(x \in C_1) = 1 \tag{22}$$

which leads to the contradiction $1 \geq \underline{p}_3(X) > P(x \in X) \geq P(x \in C_3) = 1$.

Case 2 is possible only if $X \cap C_3 = \varnothing$ but $X \cap C_1 \neq \varnothing$ and $X \cap C_2 \neq \varnothing$. This implies the contradiction

$$0 = \overline{p}_3(X) < P(x \in X) = P(x \in X \cap C_1) + P(x \in X \setminus C_1) \leq P(x \in C_1 \setminus C_2) = 0 \tag{23}$$

and finishes the proof. $\qquad\square$

Lemma 2. *The intersection of the thickets $T_i = \{(\mathbb{R}^d, p), (C_i, 1 - p)\}$ $(i = 1, 2)$ with $0 < p < 1$ is the thicket $T_3 = \{(\mathbb{R}^d, p), (C_3, 1 - p)\}$, where $C_3 = C_1 \cap C_2$.*

Proof. The random variable $x' = \max\{0, (x - p)/(1 - p)\}$ is represented by the thicket $T_i' = \{(C_i, 1)\}$ if and only if $x \lhd T_i$. Therefore, the result follows from Lemma 1. $\qquad\square$

Theorem 1. *Assume that the nested thicket $T_i = \{(C_{ij}, p_{ij}) \mid j = 0, \ldots, n_i\}$ $(i = 1, 2)$ is represented by the nested stack B_i. Then the intersection of T_1 and T_2 is represented by the nested stack $B_1 \cap B_2$.*

Proof. Without loss of generality we assume $n_1 = n_2 =: n$ and $p_{1j} = p_{2j} =: p_j$. This condition can always be fulfilled if we replace T_1 and T_2 by suitable partition thickets. The set $B_3 := B_1 \cap B_2$ is a nested stack and corresponds to a thicket $T_3 = \{(C_{3j}, p_j) \mid j = 0, \ldots, n\}$, where $C_{3j} = C_{1j} \cap C_{2j}$. T_3 is a subthicket of T_1 and T_2, i.e., every random variable represented by T_3 is also represented by T_1 and T_2. We have to show the converse.

We assume that this is not the case: There exists a random variable with $x \lhd T_1$ and $x \lhd T_2$ but $x \not\lhd T_3$. If $\underline{p}_i$ and $\overline{p}_i$ denote the lower and upper probability of T_i, there exists a set $X \in \mathcal{B}_d$ such that $\underline{p}_3(X) > P(x \in X)$ or $\overline{p}_3(X) < P(x \in X)$. We consider the case $\underline{p}_3(X) > P(x \in X)$ first. Nestedness of the thickets T_i implies

$$\underline{p}_i(X) = \sum_{C_{ij} \subset X} p_j = \sum_{j=\underline{m}_i}^{n} p_j \tag{24}$$

where

$$\underline{m}_i := \min\{j \mid C_{ij} \subset X\}, \quad i = 1, 2, 3. \tag{25}$$

Hence $\underline{m}_3 < \min\{\underline{m}_1, \underline{m}_2\}$, i.e., $C_{3\underline{m}_3} \subset X$ but neither $C_{1\underline{m}_3} \subset X$ nor $C_{2\underline{m}_3} \subset X$. We let $T_i' = \{(C_{i0}', p_0'), (C_{i1}', p_1')\}$ where

$$C_{i0}' = \mathbb{R}^d, \quad C_{i1}' = C_{i\underline{m}_3}, \quad p_0' = \sum_{i=0}^{\underline{m}_3 - 1} p_i, \quad p_1' = 1 - p_0'. \tag{26}$$

and denote the corresponding lower and upper probabilities by $\underline{p}_i'$ and $\overline{p}_i'$. Because T_i is a subthicket of T_i', x is also represented by T_1' and T_2'. Because of $\underline{p}_3'(X) = \underline{p}(X) > P(x \in X)$, x is not represented by T_3'. This is a contradiction to Lemma 2. Now we consider the case $\overline{p}(X) < P(x \in X)$. The proof is similar to the previous case. We have

$$\overline{p}_i(X) = \sum_{C_{ij} \cap X \neq \varnothing} p_j = \sum_{j=0}^{\overline{m}_i}, \tag{27}$$

where

$$\overline{m}_i := \max\{j \mid C_{ij} \cap X \neq \varnothing\}, \quad i = 1, 2, 3. \tag{28}$$

Hence $\overline{m}_3 > \max\{\overline{m}_1, \overline{m}_2\}$, i.e., $C_{3\overline{m}_3} \cap X \neq \varnothing$ but $C_{1\overline{m}_3} \cap X = \varnothing$ and $C_{2\overline{m}_3} \cap X = \varnothing$. We let $T_i'' = \{(C_{i0}'', p_0''), (C_{i1}'', p_1'')\}$, where

$$C_{i0}'' = \mathbb{R}^d, \quad C_{i1}'' = C_{i\overline{m}_3}, \quad p_0'' = \sum_{i=0}^{\overline{m}_3 - 1} p_i, \quad p_1'' = 1 - p_0''. \tag{29}$$

and denote the corresponding lower and upper probabilities by $\underline{p}_i''$ and $\overline{p}_i''$. Because T_i is a subthicket of T_i'', x is also represented by T_1'' and T_2''. Because of $\overline{p}_3''(X) = \overline{p}(X) > P(x \in X)$, x is not represented by T_3''. Again, this is a contradiction to Lemma 2. $\qquad\square$
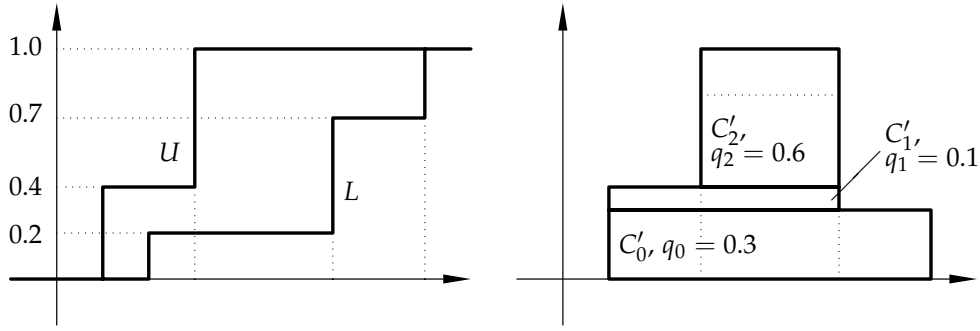
Figure 6: Nested thicket representation.

Example 3 (Nested thicket intersection). The intersection of the nested thickets

$$T_1 = \{((1,6),0.5),\ ((3,4),0.5)\}, \quad T_2 = \{((0,7),0.3),\ ((2,5),0.7)\} \tag{30}$$

is the nested thicket $T_1 \cap T_2 = \{((1,6),0.3),\ ((2,5),0.2),\ ((3,4),0.5)\}$.

In order to intersect arbitrary thickets, we have to transform them to nested thickets. We cannot expect that to be possible without losing information:

Algorithm 4 (Nested thicket conversion)**.**

Input: Thicket $T = \{(C_i, p_i)\}$.

Output: Nested thicket T' of which T is a subthicket.

Step 1: Denote by U_T and L_T the distribution envelopes of T. Let Z be the union of range(U_T) and range$(1 - L_T)$ and sort its elements in ascending order: $0 = z_0 < \cdots < z_n = 1$. If desired, Z can be replaced by a partition of Z.

Step 2: For $k = 0, \ldots, n-1$ let u_k be the largest value in range(U_T) that is not larger than z_k and let l_k be the smallest value in range$(1 - L_T)$ that is not larger than z_k.

Step 3: Let

$$C'_k = \eta \operatorname{CH}\!\left(\bar{\partial}[U_T \approx u_k] \ \cap \ \bar{\partial}[L_T \approx 1 - l_k]\right). \tag{31}$$

If $k = n-1$ or if C'_k is empty, let $k_0 := k$ and continue with the next step.

Step 4: For $k = 0, \ldots, k_0 - 2$ let $p'_k = z_{k+1} - z_k$ and let $p'_{k_0-1} = 1 - \sum_{l=0}^{k_0-2} p'_l$.

Step 5: The output of the algorithm is the thicket $T' = \{(C'_i, p'_i) \mid i = 0, \ldots, k_0 - 1\}$.

Proof (of correctness of Algorithm 4). The thicket T' is well defined because C'_0 is never empty. It is no loss of generality to assume range$(U_T) = $ range$(1 - L_T)$. If this condition is not fulfilled, we can replace T by a suitable partition thicket. By construction, the thicket T' is nested and there exists an injection $\iota_1 \colon \{0, \ldots, k_0 - 1\} \to \{0, \ldots, n\}$ such that $C'^+_i = C^+_{\iota_1(i)}$ and, for $i \neq k_0 - 1$, $p'_i = p_{\iota_1(i)}$. Furthermore, p'_{k_0-1} is the sum over all values p_i where $i \notin$ range(ι_1). This implies $L_{T'} \leq L_T$.

Analogously, there exists an injection $\iota_2 \colon \{0, \ldots, k_0 - 1\} \to \{0, \ldots, n\}$ such that $C'^-_i = C^-_{\iota_2(i)}$ and, for $i \neq k_0 - 1$, $p'_i = p_{\iota_2(i)}$. Again p'_{k_0-1} is the sum over all values p_i where $i \notin$ range(ι_2). This implies $U_T \leq U_{T'}$ and T is a subthicket of T'. $\qquad\square$

Example 4. We convert the univariate thicket T represented by the distribution envelopes U, L of Figure 6, left, into a nested thicket. First we unite the ranges of U and $1 - L$ and obtain the set $Z = \{0.0, 0.3, 0.4, 0.8, 1.0\}$. Now we construct the sets C'_0, C'_1 and C'_2 according to Step 2 of the nested thicket conversion algorithm. Since C'_3 is empty, we leave Step 3 with $k_0 = 2$ and assign the complete remaining probability mass 0.2 to C'_2.

Nested thicket conversion with Algorithm 4 works for all thickets. The procedure however is information losing and depends on the coordinate frame. In fact, there are two information leaks. One is due to the intermediate representation of a thicket by distribution envelopes. The second potential loss of information occurs if we leave Step 3 with $k_0 < n - 1$. In this case, we have to assign the remaining probability mass to the set C'_{k_0} which means extending both distribution bounds.

The thicket of Example 4 cannot be converted into an equivalent nested thicket while the thicket T_1 of Equation (6) can. The equivalent nested thicket is the thicket T_2 of (6).

The main reason for using nested thickets in multivariate DEnv is the possibility to combine the results DEnv with respect to different coordinate frames by nested thicket intersection.

Algorithm 5 (Multivariate DEnv with intersection).

Input: Two thickets $T' = \{(C'_i, p'_i)\}$, $T'' = \{(C''_i, p''_i)\}$ and a binary operation "$\star$", defined on $\mathbb{R}^d$ and extended to nonempty open convex sets.

Output: A thicket representing all random variables $x' \star x''$ where $x' \lhd T'$ and $x'' \lhd T''$. No assumption on the dependence of x' and x'' is made.

Step 1: Choose a certain number of bases A_i of $\mathbb{R}^d$. For each of them perform the multivariate DEnv algorithm (Algorithm 3). Denote the resulting thickets by T_i.

Step 2: Use nested thicket conversion (Algorithm 4) to transform the thicket T_i to a nested thicket T'_i. Use the basis A_i for this transformation.

Step 3: Return the thicket $T = \bigcap T'_i$.

Proof (of correctness of Algorithm 5). The algorithm is correct because the thickets T_i represent every random variable $x' \star x''$ if $x' \lhd T'$ and $x'' \lhd T''$, and the nested thicket transformation produces thickets T'_i with the same property. According to Theorem 1 this is also true for the output thicket T. $\qquad\square$

Remark 4. In an actual implementation of the algorithm it is not necessary to deconstruct the distribution bounds U_i, L_i into thickets T_i in Step 1, since we need precisely these distribution bounds (and not the thickets themselves) for the nested thicket conversion in Step 2.

Remark 5. The applications one has in mind should determine which bases to choose in Step 1 of the above algorithm. Probably it is often sufficient to choose 2^d bases A_i so that $\bigcup 0^+(A_i) = \mathbb{R}^d$. In 2D, these could be for example

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{32}$$

We use the bases (32) in the example of Section 5.

The multivariate DEnv with intersection reduces the loss of information due to intermediate thicket representation by distribution envelopes. It induces, however, a loss of information in conversion to nested thickets. What is worse depends on the particular examples. Certain families of random variables can be represented very well by nested thickets while others cannot. Some arguments in favor of the refined second version of multivariate DEnv are:
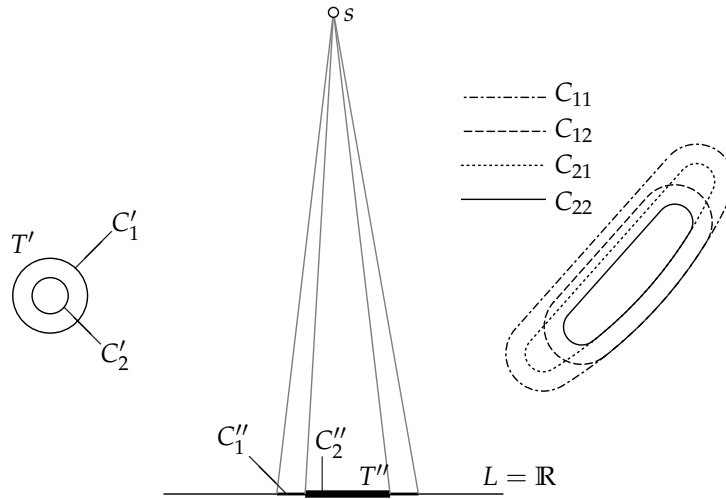
Figure 7: Reflection of a point in a line.

- The loss of information due to thicket representation by distribution envelopes might be considerable (Example 2) while the nested thicket conversion is often information preserving or nearly information preserving.

- In geometric constructions with imprecise input data (as well as in interval arithmetic) the excessive growth of tolerance regions is a problem. The multivariate DEnv algorithm with intersection allows better control of the size of the convex sets C_i.

- The refined multivariate DEnv algorithm at least approaches geometric invariance.

5. An Example

This section provides a comprehensive example illustrating the algorithms of this article.

5.1. **Definition of a binary operation.** In the Euclidean plane we are given a straight line L and a point $s \notin L$. The straight line L is identified with $\mathbb{R}$; every point $r \in L$ corresponds to a real number and vice versa. We use s and L to define a binary operation $\star \colon \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$. The point $a \star r$ is defined as the reflection image of the point a in the straight line through s and $r \in L$.

In Figure 7 this situation is visualized. We anticipatorily assume that both a and r are random variables, represented by thickets T' and T'', respectively.

5.2. **Extension to intervals.** We have to extend "$\star$" to convex subsets of $\mathbb{R}^2$ and $\mathbb{R}$: All points of a convex set A are to be reflected in all straight lines S that contain the point s and intersect L in an interval R. The set M that is swept by this one-parameter family of reflected copies of A is described in Hu and Wallner [2004]. It is bounded by parts of the boundary of the reflection of A in the two lines through the boundary points of R and two circular arc segments with center s. Since our algorithm requires convex sets as result of the operation "$\star$", we replace M by its convex hull $\mathrm{CH}(M)$, i.e., the extension of "$\star$" to convex sets is defined by $A \star R := \mathrm{CH}(M)$.

5.3. **Representation by thickets.** Now we meet the requirements for an extension of "$\star$" to multivariate thickets. We consider a concrete example. The position of the point a as well as the intersection of the straight line S with L are random variables represented by thickets T' and T'', respectively. We choose $T' = \{(C_1', 0.6), (C_2', 0.4)\}$, where C_1' and C_2' are the concentric

disks shown by Figure 7. Further, we let $T'' = \{(C_1'', 0.8), (C_2'', 0.2)\}$, where C_1'' and C_2'' are the two intervals indicated in Figure 7.

5.4. Upper and lower distribution envelopes.

The first step in multivariate DEnv with intersection consists of computing the convex sets $C_{ij} = C_i' \star C_j''$. In our example, each C_{ij} is bounded by three circular arcs and one straight line segment (see Figure 7). We distinguish the different $C_{ij}'s$ by line style.

The next step is the computation of distribution envelopes with respect to different bases of $\mathbb{R}^2$. We choose the initial basis A_0 of Equation (32) and compute the corresponding distribution envelopes $U_0 = U_0(A_0)$ and $L_0 = L_0(A_0)$:

Both upper and lower distribution envelope are simple functions of finite range. The value $U_0(t)$ is defined as the solution of the optimization problem

$$U_0(t) = \max \sum_{t \in C_{ij}^-} p_{ij}. \tag{33}$$

The variables to solve for are the unknown probability values p_{ij} of C_{ij}; the necessary constraints are

$$p_{11} + p_{12} = 0.6, \quad p_{21} + p_{22} = 0.4, \quad p_{11} + p_{21} = 0.8, \quad p_{12} + p_{22} = 0.2, \quad p_{ij} \geq 0. \tag{34}$$

Figure 8 shows that $\mathbb{R}^2$ can be divided into five regions such that the target functional (33) is the same for all points of one region. Hence the range of U_0 is found by solving five optimization problems, two of which are trivial: the empty sum yields zero and the sum over all values p_{ij} yields one. The three remaining target functionals and the corresponding solutions, found by linear programming, are

$$0.6 = \max\{p_{11}\}, \quad 0.8 = \max\{p_{11} + p_{12}\}, \quad 1.0 = \max\{p_{11} + p_{12} + p_{21}\}. \tag{35}$$

Hence the range of U_0 is the set $\{0.0, 0.6, 0.8, 1.0\}$.

The range of L_0 is computed in a similar way. We have to solve five minimization problems. The three nontrivial solutions are

$$0.0 = \min\{p_{22}\}, \quad 0.2 = \min\{p_{22} + p_{12}\}, \quad 0.4 = \min\{p_{22} + p_{12} + p_{21}\} \tag{36}$$

and the range of L_0 is $\{0.0, 0.2, 0.4, 1.0\}$.

Both U_0 and L_0 are discontinuous on points of three curves shown in bold by Figure 8. These lines of discontinuity are composed of parts of the convex sets C_{ij} and half lines parallel to the coordinate directions of A_0.

We repeat the above steps for $i = 1, 2, 3$ in order to compute distribution bounds U_i, L_i with respect to the remaining bases A_i of (32). The target functionals to be minimized are always those of (36). The minimal values are the function values of the lower distribution bounds. The target functionals to be maximized yield the function values of the upper distribution bounds. These functionals are the same for U_0 and U_3. For U_1 and U_2 they are

$$\begin{aligned} 0.6 &= \max\{p_{11}\}, & 0.8 &= \max\{p_{11} + p_{12}\}, \\ 0.8 &= \max\{p_{11} + p_{21}\}, & 1.0 &= \max\{p_{11} + p_{12} + p_{21}\}. \end{aligned} \tag{37}$$

We don't have to deconstruct the distribution bounds L_i, U_i into a thicket T_i because the distribution bounds and not the thickets themselves are needed in the next step.

5.5. Nested thicket conversion and intersection.

We convert the distribution bounds U_i and L_i into nested thickets and intersect them. In the present example, nested thicket conversion turns out to be information preserving for all four thickets. The resulting thickets consist of two elements with probability mass 0.6, 0.2 and 0.2, respectively. The result of the intersection according to Theorem 1 is depicted in Figure 9 as set of nested thickets and as nested stack B_T.
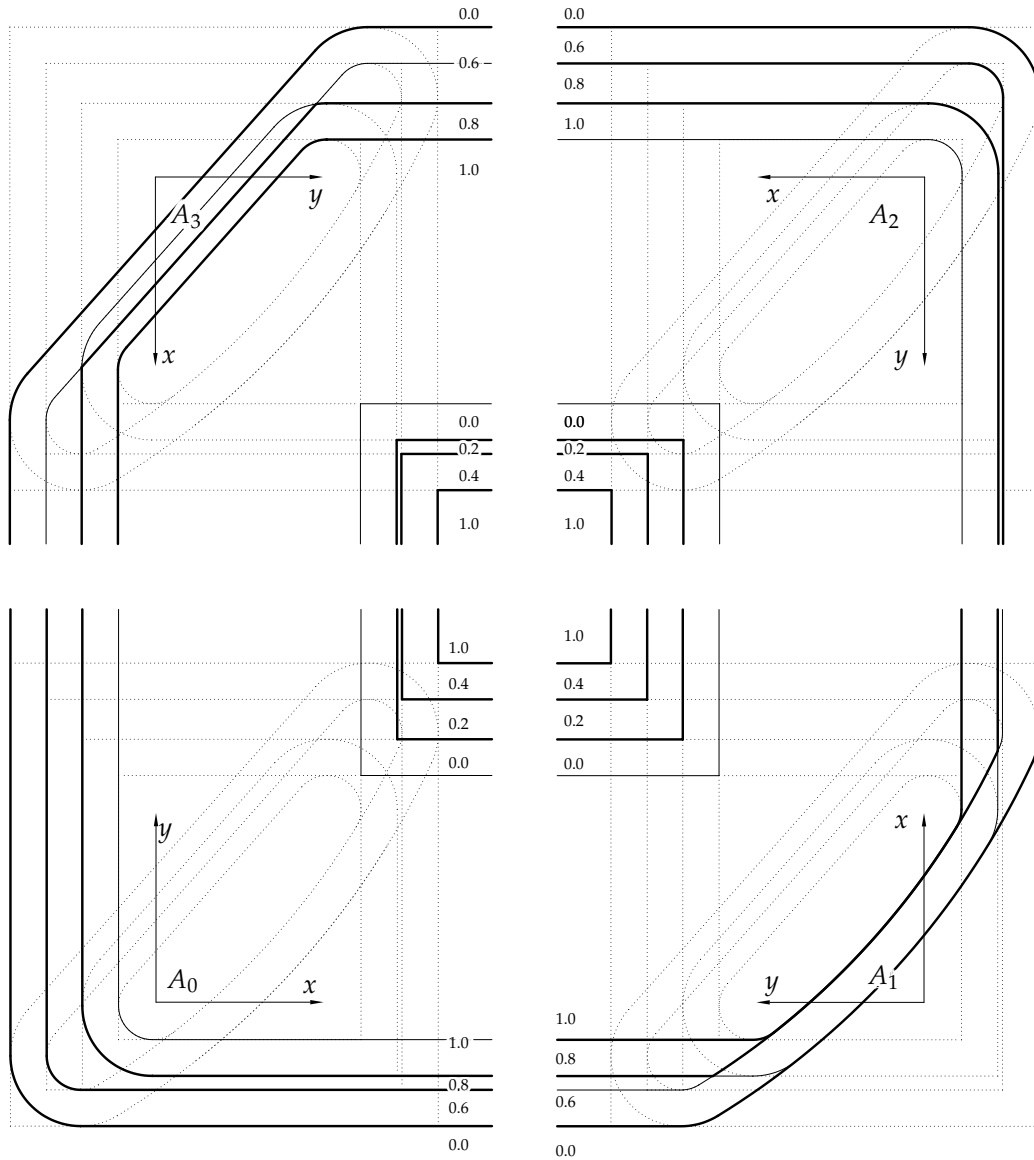
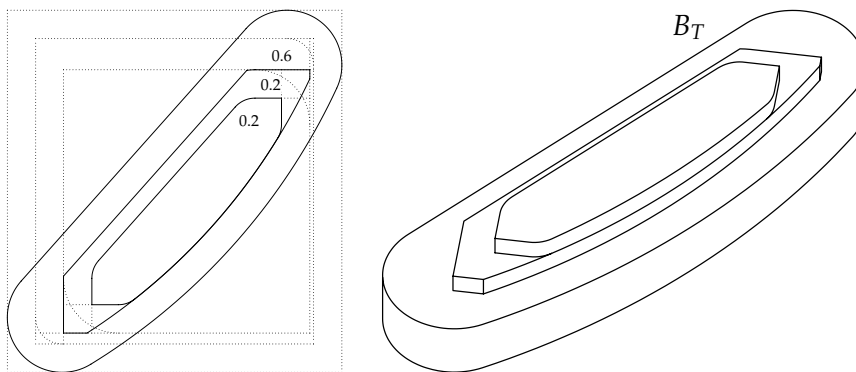Figure 8: Distribution envelopes with respect to different coordinate frames.

Figure 9: Resulting thicket

6. Conclusion

We extended the univariate DEnv algorithm for bounding the result of arithmetic operations on thickets to the multivariate case. Most concepts turned out to be straightforward generalizations. The multivariate DEnv algorithm is, however, information losing. A further obstacle to the use of DEnv in geometric constructions is its dependence on the coordinate frame.

In order to overcome these drawbacks, we introduced the concept of nested thickets and an intersection algorithm for them, which allows to minimize the effects mentioned above.

Incorporating dependence information as in typical geometric constructions is a topic of future research.

Acknowledgments

References

D. Berleant and C. Goodman-Strauss. Bounding the results of arithmetic operations on random variables of unknown dependency using intervals. *Reliab. Comput./Nadezhn. Vychisl.*, 4(2): 147–165, 1998.

D. Berleant, L. Xie, and J. Zhang. Stattool: A tool for distribution envelope determination (DEnv), an interval-based algorithm for arithmetic operations on random variables. *Reliab. Comput./Nadezhn. Vychisl.*, 9(2):91–108, 2003.

D. Berleant and J. Zhang. Representation and problem solving with distribution envelope determination. *Reliability Engineering and System Safety*, 85(1–3):153–168, 2004a.

D. Berleant and J. Zhang. Using pearson correlation to improve envelopes around the distributions of functions. *Reliab. Comput./Nadezhn. Vychisl.*, 10(2):139–161, 2004b.

S.-M. Hu and J. Wallner. Error propagation through geometric transformations. *J. Geom. Graphics*, 8(2):171–183, 2004.

H. Karloff. *Linear Programming*. Birkhäuser, Boston, 1991.

H. Pottmann, B. Odehnal, M. Peternell, J. Wallner, and R. Ait Haddou. On optimal tolerancing in Computer-Aided Design. In R. Martin and W. Wang, editors, *Geometric Modeling and Processing 2000*, pages 347–363. IEEE Computer Society, Los Alamitos, Calif., 2000.

H. A. Regan, S. Ferson, and D. Berleant. Equivalence of methods for uncertainty propagation of real-valued random variables. *Internat. J. Approx. Reason.*, 36:1–30, 2004.

A. A. G. Requicha. Towards a theory of geometric tolerancing. *Int. J. of Robotics Research*, 2: 45–60, 1983.

P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.

J. Wallner, R. Krasauskas, and H. Pottmann. Error propagation in geometric constructions. *Computer-Aided Design*, 32:631–641, 2000.

R. C. Williamson and T. Downs. Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. *Internat. J. Approx. Reason.*, 4(2):89–158, 1990.