Probabilistic Robustness Radii with n:th Order Convex Distributions

Ilmari Kangasniemi and Juha-Pekka Nikkarila

***Abstract*—We analyze a probabilistic variation on the classic robustness radius. Instead of measuring the distance to the closest destabilizing state, we look at a set of probability distributions around our initial state. The probabilistic radius is the radius of the largest sphere, for which none of the distributions supported within it takes the expected values of system parameters outside a prescribed set. We show that when the set of acceptable parameter values is closed and convex, the radius for the family of distributions which are nondecreasing with respect to distance reduces to the same radius for uniform distributions. We generalize this result for distributions of higher order convexity or concavity with respect to distance, obtaining an equal radius using a simple family of polynomials.**

***Index Terms*—Robustness radii, distributional robustness, decision analysis, control theory, aircraft control.**

I. Introduction

The *robustness radius* is a classical tool for measuring local robustness of systems. When given a set of system states X equipped with a metric$d$, and an initial state$x$, the robustness radius is the smallest distance from x to an undesirable system state. The radius provides a worst-case estimate of how great an error is acceptable for the system to stay in a desirable state. Being a worst-case estimator, it is nonprobabilistic in nature.

Depending on the application, alternate terms describing robustness radii or their special cases may be used, such as robustness margins or stability radii. Robustness radii are used for example in job scheduling [1], [2], and various control theory applications such as aircraft control [3], [4].

While the sensitivity and nonprobabilistic nature of robustness radii is useful in some applications, it may also prove to be a limitation. If, for instance, we have some prior knowledge on how the initial state changes, the radius calculation may halt at an undesirable state that according to the prior information would be extremely unlikely. This problem is further exacerbated if the model for the system itself is also probabilistic.

To answer this problem, we reformulate the robustness radius in terms of *distributional robustness*. Distributional robustness is concerned with analyzing robustness when, instead of a probability distribution, a set of possible probability distributions describing the change of the system is given. This makes it possible to incorporate known probabilistic limitations while still obtaining a simple real-valued measurement of robustness.

Similar distributional variations on the robustness radius and its various special cases have been studied before, in [5,6]. In this article, we provide a general definition for the concept, and afterwards show how to reduce the radius computation for some specific families of distributions to computing it for significantly smaller subfamilies.

II. Formulation

Given a system of interest, we denote the set of possible states for it by X, and fix a metric d in X. We assume a model M, which assigns every state $x \in X$ a value $M(x)$ in $\mathbb{R}^k$. The values of M are vectors of probabilities and expected values describing our system. This description works even in the deterministic case, as Boolean states can be described as probabilities of either 0 or 1 depending on the state of the system, and numerical metrics are merely expected values with no deviation.

We assign the system an initial state of x, and then assume the system randomly changing into a final state x' close to x. In order to define the probabilistic robustness radius, we select for every x a family of probability distributions $P(x)$. Each element $p \in P(x)$ is a possible probability distribution for the final state given an initial state x. The only assumption on $P(x)$ is for every probability distribution in it to be supported in a sphere of finite radius.

For each one of these potential distributions for x', we can calculate the expected value of $M(x')$, and check whether it is desirable. Desirability is defined by fixing a set U in $\mathbb{R}^k$ of desirable states. Our radius is the supremum of all $r \in [0, \infty)$, for which all distributions of $P(x)$ which are supported in $B_d(x, r)$ yield a desirable expected value of $M(x')$. The resulting radius is dependent on the selected point x, the family of selected distributions $P(x)$, and the set of desirable expected values U.

Next, we present a more formal definition of our concept. Throughout this paper we make the following assumptions: (X, d) is a (pseudo)metric space, and we have a measure μ on X, for which sets $B_d(x, r)$ are measurable and the inequality $0 < \mu(B_d(x, r)) < \infty$ holds for every x in X and every r in the interval $(0, \infty)$. Note that the measurability of $B_d(x, r)$ is equivalent with the fact that every function $d_x : y \mapsto d(x, y)$ is measurable. In addition to this, we assume that our model M defines a measurable function from X to $\mathbb{R}^k$, and that M is bounded on every $B_d(x, r)$. Lastly, we denote by I_M the map

$$p \mapsto \int_X pM \, d\mu,$$

Manuscript received October 21, 2015; revised May 1, 2016.

I. Kangasniemi and J. P. Nikkarila are with the University of Helsinki, They are also with Finnish Defence Forces Research Agency, P.O.Box 10, FI-11311 Riihimäki, Finland (e-mail: ilmari.kangasniemi@helsinki.fi, juha-pekka.nikkarila@mil.fi).

DOI: 10.7763/IJMO.2016.V6.529

which maps probability distributions of X to values in $\mathbb{R}^k$.

Definition 1. For every element x of X, fix a family $P(x)$ of probability distributions on X, such that every p in $P(x)$ is supported in some $B_d(x,r)$. In the case of general probability distributions this means that

$$\int_{B_d(x,r)} p\, d\mu = 1$$

for some $r > 0$.

Given a value $r > 0$, we may define the set

$$P(r,x) = \{p \in P(x) |\ \mathrm{supp}\, P \subset B_d(x,r)\}.$$

We also define $P(0,x)$ to be empty.

The probabilistic robustness radius R_P is given by

$$R_P(x) = \sup\{r \in [0,\infty) \big| I_M(P(r,x)) \subset U\},$$

with $R_P(x) = \infty$ in case the set is unbounded.

A common way in literature to define probabilistic versions of robustness radii is to add a small acceptable probability of risk to the classical robustness radius. This is part of a line of research outlined by [7], and has been used to define robustness radii in for example [5]. The literary definition fixes a set of desirable system states X_D and a positive bound ε describing the accepted level of risk. The radius is then obtained as the supremum of radii r, for which the probability of x' being in X_D is at least $1-\varepsilon$ under all probability distributions of $P(x,r)$. Definition 1 yields this as a special case by selecting the interval $[1-\varepsilon, 1]$ as U, and the characteristic function of X_D, denoted by $\mathbf{1}_{X_D}$, as M.

We can also obtain the classical robustness radius as a special case. For this, denote the set of desirable states again by X_D. We set $U = \{1\}$, select $\mathbf{1}_{X_D}$ as our M, and define every $P(x)$ to consist of Dirac delta distributions $\delta_{x'}$, for all $x' \in X$. The resulting R_P -value is the classical robustness radius.

In order to keep R-values for different x comparable, one has to be careful in selecting the families $P(x)$. Due to this, families $P(x)$ are often limited to symmetric distributions, for which the probability density depends only on the distance from x. For the rest of the paper we will focus on these distributions. However, it is noteworthy to mention the existence of research on probabilistic robustness radii using nonsymmetric distributions, for example in [8].

The definition we will be using is the following:

Definition 2. Let $\mathcal{G}$ be a family of functions from $[0,\infty)$ into $[0,\infty)$. Assume that each g in $\mathcal{G}$ is bounded and supported in an interval $[0,h)$ for some $h \in (0,\infty)$. In addition to this, assume that $g \circ d_x$ is μ-integrable over X for every element x of X.

Fix an $x \in X$. For every $g \in \mathcal{G}$, we can calculate the integral

$$\int_X g \circ d_x\, d\mu\,.$$

If this is nonnegative, we may define a probability distribution function $p_{g,x}$ as follows:

$$p_{g,x} = \left(\int_X g \circ d_x\, d\mu\right)^{-1} g \circ d_x.$$

Using these, we may define the set $P_\mathcal{G}(x)$ as follows:

$$P_\mathcal{G}(x) = \left\{p_{g,x} \middle| g \in \mathcal{G} \text{ and } \int_X g \circ d_x\, d\mu > 0\right\}.$$

For a given $r > 0$, define the set $\mathcal{G}_r$ by

$$\mathcal{G}_r = \{g \in \mathcal{G} | \mathrm{supp}\, g \subset [0,r)\}.$$

Now, we have the following representation for $P_\mathcal{G}(r,x)$:

$$P_\mathcal{G}(r,x) = \{p_{g,x} \in P_\mathcal{G}(x) \big| g \in \mathcal{G}_r\}.$$

Note that $P_\mathcal{G}(0,x)$ is defined to be empty.

By definition 1, the set $P_\mathcal{G}(x)$ defines a robustness radius $R_{P_\mathcal{G}}(x)$. We denote this radius by $R_\mathcal{G}(x)$.

Note that if supp g is open in $[0,\infty)$ for every $g \in \mathcal{G}$, we can formulate $R_\mathcal{G}$ in an alternate way: We define the sets $\tilde{P}_\mathcal{G}(r,x)$ by

$$\tilde{P}_\mathcal{G}(r,x) = P_\mathcal{G}(r,x) \setminus \bigcup_{r'<r} P_\mathcal{G}(r',x)\,.$$

Since the supports are open, every distribution $p_{g,x} \in P_\mathcal{G}(x)$ is in some set $\tilde{P}_\mathcal{G}(r,x)$. Therefore, we may write out our radius as

$$R_\mathcal{G}(x) = \inf\left\{r \in [0,\infty) \middle| (X \setminus U) \cap I_M\left(\tilde{P}_\mathcal{G}(r,x)\right) \neq \emptyset\right\},$$

with $R_\mathcal{G}(x) = \infty$ if the set we're taking the infimum over is empty.

III. Subfamilies

We focus on the case where $U \subset \mathbb{R}^k$ is convex and closed. In this case, we obtain a way to simplify calculations of $R_\mathcal{G}$ for suitable families $\mathcal{G}$. We define a *nonnegative linear combination* of elements $a_1, \ldots, a_n$ to be a linear combination $\lambda_1 a_1 + \cdots + \lambda_n a_n$, where each coefficient λ_i is nonnegative.

Lemma 3. Let $\mathcal{G}$ and $\mathcal{G}^*$ be families of functions as in definition 2, with $\mathcal{G}^* \subset \mathcal{G}$. Fix a point x from X. Assume that U is convex and closed, and that for every combination of $r \in [0,\infty)$, $g \in \mathcal{G}_r$ and $\varepsilon > 0$, we have a nonnegative linear combination $p^*_{g,x}$ of elements of $P_{\mathcal{G}^*}(r,x)$ such that

$$\int_X \left| p^*_{g,x} - p_{g,x} \right| d\mu < \varepsilon.$$

In this case, we obtain the equality

$$R_\mathcal{G}(x) = R_{\mathcal{G}^*}(x).$$

Note that $p^*_{g,x}$ does not have to be a probability distribution, as it may integrate to something other than 1 over X. In fact, $p^*_{g,x}$ is a probability distribution precisely when it is a convex combination of elements of $P_{\mathcal{G}^*}(r,x)$.

Proof.

Since $\mathcal{G}^* \subset \mathcal{G}$, it is clear that $R_{\mathcal{G}}(x) \leq R_{\mathcal{G}^*}(x)$. Now, fix an arbitrary r that is less than $R_{\mathcal{G}^*}(x)$. We show that r is at most $R_{\mathcal{G}}(x)$. The result follows from this being true for all r defined as above. Hence, fix $g \in \mathcal{G}_r$. We want to show that $I_M(p_{g,x})$ is in U.

Fix an $\varepsilon \in (0, 1/2)$. By our assumptions, we get a non-negative linear combination $p^*_{g,x} = \lambda_1 p^*_1 + \cdots + \lambda_n p^*_n$ of elements of $P_{\mathcal{G}^*}(r,x)$ fulfilling

$$\int_X \left| p^*_{g,x} - p_{g,x} \right| d\mu < \varepsilon.$$

Since the integral of $p_{g,x}$ over X is 1, the integral of $p^*_{g,x}$ over X is finite and positive. Hence we can define a probability distribution $Cp^*_{g,x}$, where C is a normalizing constant, for which the value of the integral of $Cp^*_{g,x}$ over X is 1.

Since every p^*_i integrates to 1 over X, $C\lambda_1 + \cdots + C\lambda_n$ equals 1. Since C and each λ_i is also nonnegative, $Cp^*_{g,x}$ is a convex combination of elements of $P_{\mathcal{G}^*}(r,x)$. Hence, $I_M(Cp^*_{g,x})$ is in U. This is due to it being a convex combination of $I_M(p^*_i)$, all of which are in U, combined with the fact that U is convex.

Next we make the following estimate:

$$\begin{aligned}|1 - C^{-1}| &= \left| \int_X p_{g,x}\, d\mu - \int_X p^*_{g,x}\, d\mu \right| \\ &\leq \int_X \left| p^*_{g,x} - p_{g,x} \right| d\mu \\ &< \varepsilon.\end{aligned}$$

This results in an estimate of

$$\begin{aligned}&\left| \int_X p_{g,x}\, d\mu - \int_X Cp^*_{g,x}\, d\mu \right| \\ &\quad\leq \int_X \left| p_{g,x} - p^*_{g,x} \right| d\mu + \int_X \left| Cp^*_{g,x} - p^*_{g,x} \right| d\mu \\ &\quad< \varepsilon + |1 - C^{-1}| \int_X Cp^*_{g,x}\, d\mu \\ &\quad< 2\varepsilon.\end{aligned}$$

Consequently, we can find a sequence of probability distributions (q_i) with the following properties: Each q_i is zero outside $B_d(x,r)$, $I_M(q_i)$ is in U for every i, and the sequence (q_i) converges to $p_{g,x}$ in the L^1-norm. Now, when restricted to the set of probability distributions that vanish outside $B_d(x,r)$, I_M is a continuous map from the L^1-space of distributions to the Euclidean space $\mathbb{R}^k$. This is because I_M is a linear operator which fulfills

$$I_M(q) \leq \sum_{i=1}^{k} \left| \left(I_M(q)\right)_i \right| \leq k \int_X |qM|\, d\mu \leq kM_{x,r} \int_X |q|\, d\mu$$

for all distributions q which vanish outside $B_d(x,r)$. Here $M_{x,r}$ is the upper bound of $|M|$ on $B_d(x,r)$ guaranteed by our assumptions.

From the fact that (q_i) converges to $p_{g,x}$, the continuity of the restriction of I_M, and the closedness of U, we conclude that $I_M(p_{g,x})$ is in U. This concludes our proof.

In the next two chapters, we present a set of suitable families that this result can be applied on.

IV. Nonincreasing Functions

One natural property to use in defining G is nonincreasingness. For this, we define the families $\mathcal{D}^0_r$ to consist of all nonincreasing functions $g: [0,\infty) \to [0,\infty)$ supported in $[0,r)$. The set $\mathcal{D}^0$ is the union of these families. Note that nonincreasing functions are Borel, so $g \circ d_x$ is guaranteed to be μ-measurable for every $g \in \mathcal{D}^0$ and $x \in X$.

Our next goal is to provide an easier way to compute $R_{\mathcal{D}^0}$. Previous research on distribution families, such as [7], has shown a connection between the families of uniform distributions and nonincreasing symmetric distributions. Similarly to these previous results, we conclude that the subfamily of all uniform distributions on spheres yields exactly the same probabilistic robustness radius.

Theorem 4. Assume that U is convex and closed under the Euclidean metric of $\mathbb{R}^k$. Let $\mathcal{P}^0$ denote the set of characteristic functions $\{\mathbf{1}_{[0,r)} | r \in (0,\infty)\}$. Then for all $x \in X$ one obtains that

$$\begin{aligned}R_{\mathcal{D}^0}(x) &= R_{\mathcal{P}^0}(x) \\ &= \inf \left\{ r \in [0,\infty) \,\middle|\, \frac{\int_{B_d(x,r)} M\, d\mu}{\mu\left(B_d(x,r)\right)} \notin U \right\}.\end{aligned}$$

The latter form results from the fact that the supports of functions of $\mathcal{P}^0$ are open in $[0,\infty)$, combined with the set $\tilde{P}_{\mathcal{P}^0}(r,x)$ containing exactly one function for each positive r. Almost no assumptions are made on the model M, the only ones being its measurability and boundedness on all spheres of finite radius.

We prove this result by use of Lemma 3 on families $\mathcal{D}^0$ and $\mathcal{P}^0$. For this, the following result is required.

Lemma 5.

Let g be an element of $\mathcal{D}^0_r$. Fix $x \in X$ and $\varepsilon > 0$. Then there is a nonnegative linear combination $p_\varepsilon = \lambda_1 g_{1,x} + \cdots + \lambda_n g_{n,x}$ such that each g_i is in $\mathcal{P}^0_r$ and

$$\int_X \left| p_{g,x} - p_\varepsilon \right| d\mu < \varepsilon.$$

Proof. For the sake of notational convenience, we abbreviate $B_d(x,t)$ as $B_d(t)$. In addition to this, we denote by G the integral

$$\int_X g \circ d_x\, d\mu.$$

By the previously established notation, we can write out $p_{g,x}$ as $G^{-1} g \circ d_x$.

Denote $\mu(B(r))$ by A. By our assumptions A is finite. Now, define the function $h: [0,r] \mapsto [0,r]$ by $h(t) = \mu(B(t))$. The below-continuity of measures implies the left-continuity of h. In addition to this, h is nondecreasing, so it is continuous outside a countable amount of jump discontinuities, the positions of which we denote by $\{t_1, t_2, \dots\}$. Clearly these jump discontinuities are of the size $\{A_1, A_2, \dots\}$, where A_i equals $\mu(d_x^{-1}\{t_i\})$.

We denote by I_A the interval $[0, A]$ and by I_h the subset $h[0,r]$ of I_A. By previous observations, $I_A \setminus I_h$ is a countable union of disjoint intervals $I_{A_i} = (h(t_i), h(t_i) + A_i)$, where the upper endpoint may also be included. The lower endpoints of these intervals, on the other hand, are always in I_h, due to left-continuity of h.

Now, fix an integer $l > 0$ such that $l > AG^{-1}g(0)\,/\varepsilon$. We divide $[0,r]$ into intervals$[s_i, s_{i+1}]$, starting from $i = 0$, as follows: Observe the points $y_j = jA/l$, where $j = 0, \dots l$. These points all lie in I_A. Starting from y_0 and proceeding inductively, if our current y_j is in I_h, we select as the next s the largest element of I_A mapped to y_j by h. This element is well defined due to the left-continuity of h. Afterwards, we move on to y_{j+1}. If y_j is not in I_h, it has to be in some interval I_{A_i}. If no prior s maps to the lower endpoint of I_{A_i}, we select the greatest value mapping to it as our next s. In any case, we move on to the next y-point that is not in the interval I_{A_i}.

We denote the collection of points s_i for a given l by $\mathcal{S}_l$, and the value $|\mathcal{S}_l| - 1$ by S_l. From the above procedure it is clear that $S_l \leq l$. In addition to this, $s_0 = 0$ and $s_{S_l} = r$, by the definition of A and the fact that $B(0) = \emptyset$. Note that $g(r)$ equals zero.

Define the function $f_l : [0,\infty] \to \mathbb{R}$ as follows:

$$f_l(t) = \sum_{i=1}^{S_l} \big(g(s_{i-1}) - g(s_i)\big)\mathbf{1}_{[0,s_i)}(t)\,.$$

We set $p_\varepsilon = G^{-1} f_l \circ d_x$. Clearly p_ε is a nonnegative linear combination of elements of $\mathcal{P}_r^0$. Also, since $f_l(t)$ is at least $g(t)$ for every t in $[0, \infty]$, p_ε is at least $p_{g,x}$ in every point of $B(r)$.

Next, we estimate the integral of the difference:

$$\int_X |p_\varepsilon - p_{g,x}|\, d\mu = \sum_{l=0}^{S_l - 1} \int_{B(s_{i+1})\setminus B(s_i)} \big(G^{-1}g(s_i) - p_{g,x}\big)\, d\mu\,.$$

Fix a value of i and observe the corresponding integral on the right hand side. The measure of the set of integration is $h(s_{i+1}) - h(s_i)$. If this measure is at most A/l, the fact that g is nonincreasing results in an upper bound of $G^{-1}\big(g(s_i) - g(s_{i+1})\big)\,A/l$ for our integral.

If the measure is greater than A/l, the function g has to have a jump discontinuity t_j at $h(s_i)$. In this case, the value of $h(s_{i+1}) - h(s_i) + A_j$ is at most A/l. We divide $B(s_{i+1})\setminus B(s_i)$ into two disjoint parts: $\{x' | d(x', x) = s_i\}$ and $B(s_{i+1}) \setminus \bar{B}(s_i)$. As we have stated earlier, the former set has measure A_j. Therefore, using our earlier estimate and the σ-additivity of measures, the latter set has a measure of at most A/l. In addition to this, the value of $p_{g,x}$ in the former set is a constant value of $G^{-1}g(s_i)$. Hence, the integral of $G^{-1}g(s_i) - p_{g,x}$ over $\{x' | d(x', x) = s_i\}$ is zero, and again, we attain the upper bound of $G^{-1}\big(g(s_i) - g(s_{i+1})\big)\,A/l$ for our integral.

We have obtained an estimate of:

$$\begin{aligned}\int_X |p_\varepsilon - p_{g,x}|\, d\mu &\leq \frac{AG^{-1}}{l} \sum_{l=0}^{S_l - 1} \big(g(s_i) - g(s_{i+1})\big) \\ &= \frac{AG^{-1}g(0)}{l} \\ &< \varepsilon.\end{aligned}$$

Now, Theorem 4 follows directly via use of Lemma 3 on families $\mathcal{D}^0$ and $\mathcal{P}^0$.

V. Convex Functions of Order n

Due to the sets $\tilde{P}_{\mathcal{P}^0}(r, x)$ being singletons, the computation of $R_{D^0}(x)$ reduces to analyzing a single real function$r \mapsto I_M\left(p_{\mathbf{1}_{[0,r)},x}\right)$. This is due to $\mathcal{P}^0$ being of the form where, for every $r > 0$, it contains exactly one g_r with supp $g_r = [0, r)$.

A set of simple families that also have this property is given by

$$\{g_r : x \mapsto (r - x)^n \mathbf{1}_{[0,r)}(x) \,|\, r \in (0, \infty)\},$$

where n goes through all natural numbers. Note that for $n = 0$, this essentially gives us $\mathcal{P}^0$. Hence, we denote these families by P^n respectively. We are interested in finding large distribution families containing $\mathcal{P}^n$ that yield the same robustness radius. For this, we require a generalization of convex functions called *convex functions of order n*.

We use definitions similar to those used in [9]. Given a real function f and a set of distinct points $x_1, \dots, x_n$, one may calculate the *divided difference* $f[x_1, \dots, x_n]$. This difference is defined inductively with respect to n, using the following formulae:

$$f[x] = f(x)$$
$$f[x_1, \dots, x_n] = \frac{f[x_1, \dots, x_{n+1}] - f[x_2, \dots, x_n]}{x_n - x_1}.$$

The definition easily results in the following representation, which is mentioned for example in [10, Ch. 1.3]:

$$f[x_1, \dots, x_n] = \sum_{k=1}^{n} \frac{f(x_k)}{\prod_{l \neq k}(x_k - x_l)}.$$

Let A be a subset of R. We call a function $f: A \to \mathbb{R}$ *convex of the order n*, or more shortly, *n-convex*, if it fulfills the following condition: For every set $\{x_0, \dots, x_{n+1}\} \subset A$, where x_i are ordered in an ascending order, the divided difference $f[x_1, \dots, x_{n+1}]$ is at least 0. This definition results in the set of 1-convex functions consisting of all convex functions, and the set of 0-convex functions consisting of all nondecreasing functions. On open intervals for positive values of n, n-convexity is equivalent with the function admitting a convex $(n-1)$:th derivative $f^{(n-1)}$, with the 0:th derivative being the function itself – see for example [9].

We define the sets $\mathcal{I}^n$ as follows:

$$\mathcal{I}^n = \left\{ f : [0,\infty) \to [0,\infty) \middle| \begin{matrix} f \text{ is } n\text{-convex and} \\ \operatorname{supp} f \subset [0,r) \text{ for some } r \end{matrix} \right\}.$$

Besides convexity, we can also define *concavity of order* n. In this case, the definition is given by $f[x_1,\dots,x_{n+1}]$ being at most 0, for all x_i ordered in an ascending order. The derivative condition for n-concavity on open intervals is similarly $f^{(n-1)}$ being concave. If f is n-convex, $-f$ is n-concave, and vice versa.

Based on this, we define the sets

$$\mathcal{D}^n = \left\{ f : [0,\infty) \to [0,\infty) \middle| \begin{matrix} f \text{ is } n\text{-concave and} \\ \operatorname{supp} f \subset [0,r) \text{ for some } r \end{matrix} \right\}.$$

Since 0-concavity is equivalent with nonincreasingness, this coincides with our earlier definition of $\mathcal{D}^0$.

For all of our previously defined sets $\mathcal{P}^n$, $\mathcal{I}^n$ and $\mathcal{D}^n$, the sets $\mathcal{P}^n_r$, $\mathcal{I}^n_r$ and $\mathcal{D}^n_r$ are defined as previously by restricting to elements supported in the set $[0,r)$. Using the definition via divided differences, we prove the following lemma:

Lemma 6. Fix a nonnegative integer n, and a radius $r \in [0,\infty]$. Let g be an element of $\mathcal{I}^n_r$. If n is odd, g is nonincreasing. If n is even, g is nondecreasing. As a corollary, for even values of n, the set $\mathcal{I}^n$ contains only the zero function.

A similar result holds for functions g of $\mathcal{D}^n_r$: For odd values of n, g is nondecreasing, and for even values of n, g is nonincreasing. Therefore, whenever n is odd, $\mathcal{D}^n$ contains only the zero function.

The case $n = 0$ is clear, so we may assume $n > 0$. We prove the case where g is in $\mathcal{I}^n_r$ and n is odd. Let x, y be elements of $[0,\infty]$, with the assumption that $x < y$, and let A be a real number greater than both y and r.

We derive the inequality

$$\begin{aligned} &0 \le g[x,y,A,A+1,\dots A+(n-1)] \\ &= \frac{g(x)}{(x-y)(x-A)\cdots\big(x-(A+n-1)\big)} \\ &+ \frac{g(y)}{(y-x)(y-A)\cdots\big(y-(A+n-1)\big)} \\ &= \left(\frac{(-1)^n}{(y-x)A^n}\right)\left(\frac{g(y)}{\left(1-\frac{y}{A}\right)\cdots\left(1-\frac{y+1-n}{A}\right)}\right. \\ &\left. + \frac{g(x)}{\left(1-\frac{x}{A}\right)\cdots\left(1-\frac{x+1-n}{A}\right)}\right). \end{aligned}$$

Since n is odd, the value within the right pair of parantheses is nonpositive for all previously defined A. Since the limit of this value at $A \to \infty$ is $g(y) - g(x)$, we obtain the nonpositivity of $g(y) - g(x)$. Since this holds for all $0 \le x < y < \infty$, g is nonincreasing. The other three cases are proven analoguously.

Next, define the following subfamilies:

$$\tilde{\mathcal{I}}^n = \{f \in \mathcal{I}^n | f \text{ is continuous}\},$$
$$\tilde{\mathcal{D}}^n = \{f \in \mathcal{D}^n | f \text{ is continuous}\}.$$

Again, subfamilies $\tilde{\mathcal{I}}^n_r$ and $\tilde{\mathcal{D}}^n_r$ are defined as previously. Due to convex functions being continuous on open intervals, every element of $\mathcal{I}^n$ is continuous at every point, except possibly $x = 0$. Hence, for every element f of $\mathcal{I}^n$, we may associate an element $\tilde{f}$ of $\tilde{\mathcal{I}}^n$ by setting $\tilde{f}(x) = \lim_{t\to x^+} f(t)$. We see that $\tilde{f}$ is in $\tilde{\mathcal{I}}^n$, since $f[x,x_1,\dots,x_{n+1}]$ is continuous in $[0,x_1)$. Similar results hold for $\mathcal{D}^n$.

By using the derivative condition and taking the limit at the point $x = 0$, we reach the conclusion of $\mathcal{P}^n \subset \tilde{\mathcal{I}}^n$ for odd n, and $\mathcal{P}^n \subset \tilde{\mathcal{D}}^n$ for even n. According to the next result, $\mathcal{P}^n$ actually gives the same R-value.

Lemma 7. Let n be a positive integer and $r \in (0,\infty)$. If n is odd, select an element g of $\tilde{\mathcal{I}}^n_r$. Otherwise, select g from $\tilde{\mathcal{D}}^n_r$. Fix $\varepsilon > 0$. Then there is a nonnegative linear combination of elements of $\mathcal{P}^n_r$, denoted by $f = f_1\lambda_1 + \cdots + f_l\lambda_l$, for which $|f(x) - g(x)| < \varepsilon$ on all of $[0,\infty)$. In other words, we can approximate g uniformly with nonnegative linear combinations of elements of $\mathcal{P}^n_r$.

Proof. We use induction on n, starting from case $n = 1$. We assume a convex, continuous g supported in $[0,r)$. Using Lemma 6, we obtain the nonincreasingness of g.

The first half of the proof shows the base case of $n = 1$. We select an integer l such that $g(0)/l < \varepsilon$. Since g is continuous and nonincreasing, $g([0,r]) = [0, g(0)]$. Next, we select points $x_0,\dots,x_l$ for which $g(x_i) = ig(0)/l$, taking care to select $x_l = 0$ and $x_0 = r$. Due to the nonincreasingness of g, the points x_i have the order $x_0 > x_1 > \cdots > x_l$.

Now, we may select our f_i and λ_i as follows for $i = 1,\dots l$:

$$f_i(x) = (x_{i-1} - x)\mathbf{1}_{[0,x_{i-1})},$$
$$\lambda_i = \frac{g(x_i) - g(x_{i-1})}{x_{i-1} - x_i} - \sum_{j<i} \lambda_j .$$

These yield a linear combination f with the property $f(x_i) = g(x_i)$ for every $i = 0,\dots,l$. This is best seen by induction: As the base case, we have $f(x_0) = 0 = g(x_0)$. For the induction step, we assume that $f(x_i)$ equals $g(x_i)$. Now, on the interval $[x_{i+1}, x_i]$ our f is a linear function of the form $x \mapsto C - Kx$, where x_i maps to $g(x_i)$ and K is the sum $\lambda_1 + \cdots + \lambda_{i+1}$, which by the definition of λ_{i+1} equals the quotient $(g(x_{i+1}) - g(x_i))/(x_i - x_{i+1})$. Hence, $f(x_{i+1})$ has to equal $g(x_{i+1})$.

Next, we show that f is a nonnegative linear combination of functions f_i. The nonnegativeness of λ_1 is clear. For $i > 1$, we obtain that

$$\begin{aligned} \lambda_i &= \frac{g(x_i) - g(x_{i-1})}{x_{i-1} - x_i} - \lambda_{i-1} - \sum_{j<i-1} \lambda_j \\ &= \frac{g(x_i) - g(x_{i-1})}{x_{i-1} - x_i} - \left(\frac{g(x_{i-1}) - g(x_{i-2})}{x_{i-2} - x_{i-2}} - \sum_{j<i-1} \lambda_j \right) \\ &\quad - \sum_{j<i-1} \lambda_j \\ &= (x_{i-2} - x_i) g[x_i, x_{i-1}, x_{i-2}] \ge 0. \end{aligned}$$

Therefore, every λ_i is nonnegative.

We have shown that f is a nonnegative linear combination of f_i which fulfills $f(x_i) = g(x_i)$ for $i = 0, \dots, l$. Since f_i are all nonincreasing, f is also nonincreasing. Hence, for any $x \in [0, r)$, there is an interval $[x_{i+1}, x_i]$ containing x, and therefore

$$|f(x) - g(x)| \leq g(x_{i+1}) - g(x_i) = g(0)/l < \varepsilon.$$

Since $g(x)$ and $f(x)$ are both zero for all $x \geq r$, f is the desired uniform approximation of g.

Next, we show the induction step. We prove the case of $g \in \tilde{\mathcal{I}}_r^n$, as the proof of case $g \in \tilde{\mathcal{D}}_r^n$ is essentially identical. We assume that n is odd and g is in $\tilde{\mathcal{I}}_r^n$. Under these assumptions, g is a continuous function which differs from a constant one only in a compact set. Therefore, g is uniformly continuous, and there is a $\delta > 0$ for which $|g(x) - g(y)| < \varepsilon/2$ whenever $|x - y| \leq \delta$.

We define a function $h: [0, \infty) \to [0, \infty)$ by setting $h(x) = g(x + \delta)$. Clearly h is also a function of $\tilde{\mathcal{I}}_r^n$. Since h can be extended to an n-convex function on $(-\delta, \infty)$, $h^{(n-1)}$ is convex, and therefore h' is $(n-1)$-convex. In addition to this, due to lemma 6, h is nonincreasing. Consequently, h' is nonpositive, which results in $-h'$ being an element of $\tilde{\mathcal{D}}_r^{n-1}$.

By the induction assumption, we find a nonnegative linear combination of functions of $\mathcal{P}_r^{n-1}$ which approximates $-h'$ uniformly:

$$\left| -h'(x) - \sum_{i=1}^{m} \lambda_i (r_i - x)^{n-1} \mathbf{1}_{[0,r_i)}(x) \right| < \frac{\varepsilon}{2r} \quad \forall x \in [0, \infty).$$

By defining f via

$$f(x) = \sum_{i=1}^{m} \frac{\lambda_i}{n} (r_i - x)^n \mathbf{1}_{[0,r_i)}(x),$$

we obtain for all x in $[0, r)$ the bound

$$\begin{aligned} &|h(x) - f(x)| \\ &\quad \leq \int_x^r \left| h'(x) + \sum_{i=1}^{m} \lambda_i (r_i - x)^{n-1} \mathbf{1}_{[0,r_i)}(x) \right| dx \\ &\quad \leq \int_x^r \frac{\varepsilon}{2r} dx \leq \frac{\varepsilon}{2} \end{aligned}$$

Since in $[r, \infty)$ both h and f are zero, we have obtained the uniform bound $|h - f| \leq \varepsilon/2$. In addition to this, our selection of δ and h yielded the uniform bound of $|h - g| < \varepsilon/2$. Therefore, by using the triangle inequality, we conclude that f fulfills our requirements.

When combined with lemma 3, our previous lemma is enough to conclude that the R-values for $\mathcal{P}^n$ are the same as those of $\tilde{\mathcal{I}}^n$ for odd n, and $\tilde{\mathcal{D}}^n$ for even n. However, with a small technical fix, we can show the same result for $\mathcal{I}^n$ and $\mathcal{D}^n$.

Lemma 8. Let n be a positive integer and $r \in [0, \infty)$. If n is even, select an element g from $\mathcal{D}_r^n$. Otherwise select g from $\mathcal{I}_r^n$. Fix $x \in X$ and $\varepsilon > 0$. Then there is a nonnegative linear combination $p_\varepsilon = \lambda_1 p_{g_1,x} + \dots + \lambda_m p_{g_m,x}$ for which each g_i is in $\mathcal{P}_r^n$, and

$$\int_X |p_{g,x} - p_\varepsilon| \, d\mu = \varepsilon.$$

Proof. As previously, we outline the case of odd n, as the other case is essentially identical. Again, we denote $B_d(x, t)$ by $B_d(t)$, and $\mu(B_d(t))$ by $h(t)$. We define A_0 to be $\mu(\{y \in X | d(x, y) = 0\})$. In addition, we denote by G the integral

$$\int_X g \circ d_x \, d\mu.$$

As previously, h is left-continuous and nondecreasing, and $\lim_{t \to 0^+} h(t) = A_0$. Hence, we can select a $\delta > 0$ which fulfills $h(\delta) < A_0 + G\varepsilon/(3g(0))$.

Using the previous Lemma, we find a nonnegative linear combination $g^* = \lambda_1^* g_1 + \dots + \lambda_m^* g_m$ of functions of $\mathcal{P}_r^n$ fulfilling the uniform bound $|g^* - \tilde{g}| < G\varepsilon/(2h(r))$. Recall that here, $\tilde{g}$ is the previously defined function of $\tilde{\mathcal{D}}_r^n$ given by the function g of $\mathcal{D}_r^n$. We define

$$\begin{aligned} p_\varepsilon^* = G^{-1} g^* \circ d_x &= G^{-1} \sum_{i=1}^{m} \lambda_i^* (g_i \circ d_x) \\ &= \sum_{i=1}^{m} \left(\lambda_i^* G^{-1} \int_X g_i \circ d_x \, d\mu \right) p_{g_i,x}. \end{aligned}$$

Since $\tilde{g} = g$ at all points t save for possibly $t = 0$, we obtain for all y with $d(x, y) > 0$ the bound

$$\begin{aligned} |p_\varepsilon^* - p_{g,x}(y)| &= |G^{-1} g^*(d(x,y)) - G^{-1} g(d(x,y))| \\ &< \frac{\varepsilon}{2h(r)}. \end{aligned}$$

Since g is nonincreasing thanks to Lemma 6, $g(0)$ is at least $\tilde{g}(0)$. Hence, in the case of $g^*(0) \geq g(0)$, we obtain the chain of inequalities

$$\begin{aligned} g(0) - \frac{G\varepsilon}{2h(r)} < g(0) \leq g^*(0) &< \tilde{g}(0) + \frac{G\varepsilon}{2h(r)} \\ &\leq g(0) + \frac{G\varepsilon}{2h(r)}, \end{aligned}$$

which results in the bound $|g(0) - g^*(0)| < G\varepsilon/(2h(r))$. In this case, p_ε^* is a nonnegative linear combination of the desired form, as

$$\int_X |p_\varepsilon^*(y) - p_{g,x}(y)| \, d\mu < h(r) \frac{\varepsilon}{2h(r)} < \varepsilon.$$

Next, we observe the remaining case of $g^*(0) < g(0)$. We select $g_0(t) = (\delta - t)^n \mathbf{1}_{[0,\delta)}$ and

$$\begin{aligned} p_\varepsilon &= p_\varepsilon^* + \left(G^{-1}(g(0) - g^*(0))\delta^{-n}\right) g_0 \circ d_x \\ &= p_\varepsilon^* + \left(G^{-1}(g(0) - g^*(0))\delta^{-n} \int_X g_0 \circ d_x \, d\mu\right) p_{g_0,x}. \end{aligned}$$

The function p_ε has the property

$$\begin{aligned} p_\varepsilon(x) &= G^{-1}g^*(0) + \left(G^{-1}\left(g(0) - g^*(0)\right)\delta^{-n}\right)(\delta - 0)^n \\ &= G^{-1}g(0) \\ &= p_{g,x}(x), \end{aligned}$$

and therefore we obtain the bound (note: [11])

$$\begin{aligned} &\int_X |p_\varepsilon - p_{g,x}|\, d\mu \\ &\leq \int_{d_x^{-1}\{0\}} |p_\varepsilon - p_{g,x}|\, d\mu \\ &\quad + \int_{d_x^{-1}(0,\infty)} \left(|p_\varepsilon - p_\varepsilon^*| + |p_\varepsilon^* - p_{g,x}|\right) d\mu \\ &\leq \int_{d_x^{-1}\{0\}} 0\, d\mu + \int_{d_x^{-1}(0,\delta)} \frac{\left(g(0) - g^*(0)\right)\delta^n}{G\delta^n} d\mu \\ &\quad + \int_{d_x^{-1}(0,r)} \frac{\varepsilon}{2h(r)} d\mu \\ &\leq 0 + (h(\delta) - A_0)\frac{g(0) - g^*(0)}{G} + h(r)\frac{\varepsilon}{2h(r)} \\ &\leq \frac{G\varepsilon}{3g(0)}\frac{g(0)}{G} + \frac{\varepsilon}{2} \\ &< \varepsilon \end{aligned}$$

We have found a suitable nonnegative linear combination in every possible case, which concludes the proof.

By combining Lemmas 8, 5, and 3, we have finally proven the main result:

Theorem 9. Let U be convex and closed under the Euclidean metric of R^k. Fix a natural number n. Then for all $x \in X$, if n is even,

$$\begin{aligned} R_{\mathcal{D}^n}(x) &= R_{\mathcal{P}^n}(x) \\ &= \inf\left\{ r \in (0,\infty) \,\middle|\, \int_{B_d(x,r)} \frac{\left(r - d_x(x')\right)^n M(x')\, d\mu(x')}{\int_{B_d(x,r)} \left(r - d_x(x'')\right)^n d\mu(x'')} \notin U \right\}, \end{aligned}$$

and if n is odd,

$$\begin{aligned} R_{\mathcal{J}^n}(x) &= R_{\mathcal{P}^n}(x) \\ &= \inf\left\{ r \in (0,\infty) \,\middle|\, \int_{B_d(x,r)} \frac{\left(r - d_x(x')\right)^n M(x')\, d\mu(x')}{\int_{B_d(x,r)} \left(r - d_x(x'')\right)^n d\mu(x'')} \notin U \right\}. \end{aligned}$$

VI. Difference in Results for Different n

The derivative condition for n-convexity and n-concavity yields that $\mathcal{P}^n \subset \mathcal{D}^{n+1}$ for odd n, and $\mathcal{P}^n \subset \mathcal{J}^{n+1}$ for even n. Due to this, $R_{\mathcal{P}^n}(x) \leq R_{\mathcal{P}^{n+1}}(x)$ for all x and n. Next, we present a simple example which further illustrates the effects of convexity order on results.

As our example case, we use $X = [0,\infty)$, $k = 1$, and the following M:

$$M(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & 1 < x \leq 2 \\ 1, & 2 < x < \infty \end{cases}.$$

We define our set U to be $[2/3\,, 1]$. Clearly the value of $R_{\mathcal{P}^0}(0)$ is $3/2$. On the other hand, $R_{\mathcal{P}^1}(0)$ ends up being infinitely large. To see this, let r be an element of $[2,\infty)$. In this case,

$$\begin{aligned} \int_{B_d(0,r)} \frac{M(r-x)}{\int_{B_d(0,r)} (r-x)\, d\mu} d\mu &= \frac{r^2 - (r-1)^2 + (r-2)^2}{r^2} \\ &= 1 - 2r^{-1} + 3r^{-2}. \end{aligned}$$

When r^{-1} gains values from $(0, 1/2]$, $1 - 2r^{-1} + 3r^{-2}$ has a minimum value of $2/3$. Since the values of our integrals are nonincreasing when $0 \leq r \leq 2$, this is the global minimum.

As seen in the previous example, a small order of convexity n results in our radius calculation halting at comparatively smaller gaps of undesirable results. As the order of convexity grows larger, our probability distributions become increasingly focused around our intended choice. This lets the probabilistic robustness radius to pass relatively greater gaps in desirable results, resulting in a less conservative measure of robustness.

Unlike changing the size of U, this sensitivity adjustment via changing assumed convexity order has less effect on models with little variation. In the previous example, enlargening U to $[1/2, 1]$ would result in $R_{\mathcal{P}^0}(0)$ passing the drop in M. However, it would also vastly improve the robustness radii for models M which are under $2/3$ in large areas, the extreme example being a constant M of $1/2$.

VII. Conclusion

In this article we have analyzed a version of the classical robustness radius based on probability distributions. The probabilistic robustness radius was defined in the spirit of distributional robustness, by observing the worst case out of families of possible probability distributions. We limited the analysis to families consisting of symmetric distributions that have bounded supports. In this case, if the set of acceptable results is closed and convex, we deduced a condition for a subfamily of distributions yielding the same robustness radius.

These results were then applied to families of distribution functions that are n-convex or n-concave with respect to distance from the initial point. We reduced the calculation for each n to a simple family of polynomial distributions. As a special case, we reduced the radius calculation for nonincreasing symmetric distributions to the respective calculation for uniform distributions. We also noted how one can tune the sensitivity of the probabilistic robustness radius by assuming a different convexity order.

References

[1] Y. Sotskov, N. Y. Sotskova, and F. Werner, "Stability of an optimal schedule in a job shop," *Omega*, vol. 25, no. 4, pp. 397-414, 1997.

[1] L. Palopoli, C. Pinello, A. Bicchi, and A. Sangiovanni-Vincentelli, Maximizing the stability radius of a set of systems under real-time scheduling constraints," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1790-1795, 2005.

[2] S. S.-M. Lam, "Real robustness radii and performance limitations of LTI control systems," Ph.D. thesis, Univ., Toronto, 2011.

[3] G. Ferreres and J.-M. Biannic, "Reliable computation of the robustness margin for flexible aircraft," *Control Engineering Practice*, vol. 9, no. 12, pp. 1267-1278, 2001.

[4] C. M. Lagoa, P. S. Shcherbakov, and B. R. Barmish, "Probabilistic enhancement of classical robustness margins: The unirectangularity concept," *Systems and Control Letters*, vol. 35, no. 1, pp. 31-43, 1998.

[5] G. Calafiore, F. Dabbene, and R. Tempo, “The probabilistic real stability radius,” in *Proc. the 14:th World Congress of IFAC*, Beijing, China, 1999.

[6] B. Barmish and C. M. Lagoa.”The uniform distribution: A rigorous justification for its use in robustness analysis,” *Mathematics of Control, Signals and Systems,* vol. 10, no. 3, pp. 203-222, 1997.

[7] C. M. Lagoa, “Probabilistic enhancement of classical robustness margins: A class of nonsymmetric distributions,” *IEEE Transactions on Automatic Control,* vol. 48, no. 11, pp. 1990-1994, 2003.

[8] A. W. Roberts, J. Roberts, D. E. Varberg, *Convex functions*, vol. 57, Academic Press, 1974.

[9] L. M. Milne-Thomson, *The Calculus of Finite Differences*, American Mathematical Soc., 2000.

[10] M. Kuczma, *An Introduction to the Theory of Functional Equations and Inequalities: Cauchy’s Equation and Jensen’s Inequality*, Springer, 2008.

K. Ilmari Kangasniemi is a master’s student at the University of Helsinki. He became a bachelor of sciences in the Beginning of fall 2015, majoring mathematics, and is soon to finish his Master’s degree.

He served from fall 2014 to summer 2015 as a conscript researcher for the Finish Defence Forces Research Agency (FDRA). This article was written during the aforementioned period of service.

Juha-Pekka Nikkarila has a PhD in physics (2008) and serves as a researcher at Finnish Defence Research Agency (FDRA). He has acquired Master’s degrees from physics (2006) and electrics (2016).

He has served at FDRA from 2012 with research interests in operation analysis, electronic warfare and Cyber, and currently serves at FDRA as a First Lieutenant (Eng.). Earlier he has worked as a researcher in Marioff Corporation / United Technologies (2009-2012), Inspecta (2007-2009), a PhD student in University of Jyvaskyla (2006-2007), and as a physics lecturer in Metropolia University of Applied Sciences (2009-2014). He served as a conscript soldier at LapItR in 2001-2002.