Why Taylor Models And Modified Taylor Models are Empirically Successful: A Symmetry-Based Explanation

Mioara Joldeş[1], Christoph Lauter[2],
Martine Ceberio [3], Olga Kosheleva[3], and Vladik Kreinovich[3]
[1]LAAS-CNRS, 7 Avenue du Colonel Roche, 31077 Toulouse,
Cedex 4 France, joldes@laas.fr
[2]LIP6, Université Pierre et Marie Curie Paris VI,
4 place Jussieu, 75252 Paris
Cedex 05, France, christoph.lauter@lip6.fr
[3]University of Texas at El Paso, 500 W. University,
El Paso, TX 79968, USA,
mceberio@utep.edu, olgak@utep.edu, vladik@utep.edu

Abstract

In this paper, we show that symmetry-based ideas can explain the empirical success of Taylor models and modified Taylor models in representing uncertainty.

1 Taylor Models and Modified Taylor Models: A Brief Reminder

General problem: reminder. In many practical situations:

- we know the equations that describe how desired quantities change with time (and space), and
- we want to use this knowledge to find the values of the desired quantities at different moments of time, at different locations, and for different values of the corresponding actions.

For example:

- we want to know how the location of a spaceship at time t depends on t and on the parameters of the launch: time, coordinates, orientation, etc.;
- another example is particle accelerators, in which particles move in a reasonable narrow tube with speeds close to the speed of light: we want to

be able to predict the location of a particle beam at different moments of time for different values of the corresponding parameters.

In all these cases, we want to know the dependence of each desired quantity y on quantities $x_1, \ldots, x_n$.

Often, we want guaranteed bounds. There are many numerical computation techniques for computing the desired dependence. However, numerical computations usually lead to an approximate description of the actual dependence: $y_{\rm approx}(x_1, \ldots, x_n) \approx y(x_1, \ldots, x_n)$.

In many real-life problems, it is important to provide not only an approximate value of y for given $x_1, \ldots, x_n$, but guaranteed bounds on this value, i.e., values $\underline{y}(x_1, \ldots, x_n)$ and $\overline{y}(x_1, \ldots, x_n)$ for which

$$\underline{y}(x_1, \ldots, x_n) \le y(x_1, \ldots, x_n) \le \overline{y}(x_1, \ldots, x_n).$$

- This is important in space exploration: we want to make sure that the spaceship does not collide with anything and reaches its target within desired accuracy.
- This is important for particle accelerators: we want to make sure that the particle beam does not self-destruct by hitting the tube walls and instead, reaches the desired target.

Taylor models. In many real-life problems, the deviations from nominal values are small, so small that we can safely ignore terms which are quadratic or of higher order in terms of these deviations. In such situations, we can expand the desired dependence in Taylor series and keep only linear terms in this expansion. Since linearization is one of the main tools in practical physics; see, e.g., [5].

To get a more accurate result, one can take into account quadratic terms. An even more accurate result emerges if we take into account cubic an higher order terms. In general, we take the sum of the first few terms in Taylor series and thus get a polynomial $P(x_1, \ldots, x_n)$, i.e., a linear combination of monomials

$$P(x_1, \ldots, x_n) = \sum c_{i_1 \ldots i_n} \cdot x_1^{i_1} \cdot \ldots \cdot x_n^{i_n}.$$

A natural way to transform this approximate model into a guaranteed model is to supplement the approximate polynomial $P(x_1, \ldots, x_n)$ with a guaranteed upper bound Δ on the absolute value of the approximation error $y(x_1, \ldots, x_n) - P(x_1, \ldots, x_n)$:

$$|y(x_1, \ldots, y_n) - P(x_1, \ldots, x_n)| \le \Delta.$$

Once we know this upper bound, we can conclude that for each combination of values $x_1, \ldots, x_n$), we have

$$y(x_1, \ldots, x_n) \in P(x_1, \ldots, x_n) + [-\Delta, \Delta] = \sum c_{i_1 \ldots i_n} \cdot x_1^{i_1} \cdot \ldots \cdot x_n^{i_n} + [-\Delta, \Delta]. \quad (1)$$

The right-hand side of this inclusion is known as a *Taylor model*. Taylor models has indeed been successfully used in many important applications; see, e.g., [1, 2, 3, 7, 8, 10].

Modified Taylor models. In the traditional Taylor model, we use a single upper bound Δ to describe the approximation error for all possible combination of the values x_i. In many practical cases, however, the approximation error depends on the values x_i. For example, when we predict a trajectory, usually,

- the predictions are very accurate for small t,
- but become less and less accurate as the time t increases.

To get a better description of the model's accuracy, it is therefore desirable to take into account that the approximation error may depend on x_i.

One successful way to take this dependence into account was proposed in [4] under the name of *modified Taylor models.* In this description, each coefficient $c_{i_1,\ldots,i_n}$ is an interval:

$$y(x_1,\ldots,x_n)\in\sum[\underline{c}_{i_1\ldots i_n},\overline{c}_{i_1,\ldots,i_n}]\cdot x_1^{i_1}\cdot\ldots\cdot x_n^{i_n}. \tag{2}$$

Comment. The meaning of the formula (2) is that for each combination of the values $x_1,\ldots,x_n$ there exist values $c_{i_1\ldots i_n}\in[\underline{c}_{i_1\ldots i_n},\overline{c}_{i_1\ldots i_n}]$ for which

$$y(x_1,\ldots,x_n)=\sum c_{i_1\ldots i_n}\cdot x_i^{i_1}\cdot\ldots\cdot x_n^{i_n}.$$

Why? There are many ways to approximate a function: we can use Taylor series, we can use rational functions, we can use Fourier series, we can use neural networks. Why Taylor series approximation turned out to be among the most empirically efficient?

Once we have selected the Taylor model, there are many ways to take into account uncertainty: e.g., we could consider intervals multiplied not by monomials (as in formula (2)), but by more complex polynomials. So why is formula (2) empirically efficient?

What is known and what we do in this paper. A symmetry-based explanation of why Taylor models are efficient have been proposed in [9]. In this paper, we:

- first, use techniques from [6] to prove a stronger version of that result – thus providing a new justification for Taylor models, and
- then, extend this new result to a justification of the modified Taylor models.

2 Why Taylor Models: A New Justification

Towards formalization of the problem: we need to select a vector space. We want to select a family of functions $\mathcal{F}$, so that the results of our prediction have a form $y(x_1, \ldots, x_n) \in F(x_1, \ldots, x_n) + [-\Delta, \Delta]$ for some function $F \in \mathcal{F}$.

In the computer, all computations reduce to a sequence of arithmetic operations. Any function that is obtained by a sequence of arithmetic operations is *analytical*, i.e., it can be expanded into Taylor series. Thus, it is reasonable to restrict ourselves to analytical functions F.

We want to be able to represent functions from the class $\mathcal{F}$ inside a computer. If we use too many parameters, we will spend too much time processing these parameters — it might have been easier to decrease the excess width by dividing the original box into multiple subboxes. Therefore, it only makes sense to consider *finite-dimensional* families of functions.

It would be useful to select the family $\mathcal{F}$ in such a way that an application of any arithmetic operation $\odot$ does not lead to additional approximation error. In other words, ideally, we would like to select $\mathcal{F}$ in such a way that, if two intermediate results r and s belong exactly to $\mathcal{F}$, then $r \odot s$ should also belong to $\mathcal{F}$. However, if we require that, then, since

- we start with variables and
- the family is closed under addition and multiplication,

we will end up with arbitrary polynomials, which contradicts to $\mathcal{F}$ being finite-dimensional. So, we cannot require that the family $\mathcal{F}$ be closed under all arithmetic operations:

- since we cannot require that for *all* operations,
- we should at least require it for the *simplest* ones: $+$, $-$, and multiplication by a real number λ.

In other words, we require that if $F \in \mathcal{F}$ and $G \in \mathcal{F}$, then $F + G \in \mathcal{F}$ and $\lambda \cdot F \in \mathcal{F}$. So, the family $\mathcal{F}$ is a (finite-dimensional) *vector space* of functions.

We should select the optimal vector space. There are many possible vector spaces of functions. The question is: which of these vector spaces is the best ("optimal") for our purpose?

When we say "the best", we mean that on the set of all such spaces, there is a relation $\succeq$ describing which family is better or equal in quality. This relation must be transitive (if $\mathcal{F}$ is better than $\mathcal{G}$, and $\mathcal{G}$ is better than $\mathcal{H}$, then $\mathcal{F}$ is better than $\mathcal{H}$).

This relation must also clearly be *reflexive*: $\mathcal{F} \succeq \mathcal{F}$ for every family $\mathcal{F}$.

This relation is not necessarily asymmetric, because we can have two families of the same quality. However, we would like to require that this relation be *final* in the sense that it should define a unique *best* family $\mathcal{F}_{\rm opt}$, for which $\forall \mathcal{G}\,(\mathcal{F}_{\rm opt} \succeq \mathcal{G})$. Indeed:

- if none of the families is the best,
- then this criterion is of no use.

So, there should be *at least one* optimal family. Similarly:

- if *several* different families are equally best,
- then we can use this ambiguity to optimize something else.

For example:

- if we have two families with the same approximating quality,
- then we can choose the one which is easier to compute.

As a result, the original criterion was not final: we obtain a new criterion: $\mathcal{F} \succeq_{\text{new}} \mathcal{G}$, if:

- either $\mathcal{F}$ gives a better approximation,
- or $\mathcal{F} \sim_{\text{old}} \mathcal{G}$ and $\mathcal{G}$ is easier to compute.

For this new optimality criterion, the class of optimal families is narrower.

We can repeat this procedure until we obtain a final criterion for which there is only one optimal family.

Scale invariance. The numerical value of each quantity x_i depends on the choice of the measuring unit. If instead of the original measuring unit, we choose a new one which is λ_i times smaller, then all numerical values are multiplied by λ_i: $x_i \to x'_i = \lambda_i \cdot x_i$. For example:

- if to measure height, we use centimeters instead of meters,
- then all numerical values of height are multiplied by 100: e.g., 2 m becomes 200 cm.

It is reasonable to require that the relative quality of two families should not change if we simply apply such re-scaling to one of the variables x_i.

Thus, we arrive at the following definition.

Definition 1. *Let $n > 0$ and $N > 0$ be integers.*

- *By a N-*dimensional family*, we mean a family $\mathcal{F}$ of all functions of the type*

$$\{C_1 \cdot F_1(x_1, \ldots, x_n) + \ldots + C_N \cdot F_N(x_1, \ldots, x_n)\},$$

  where $F_1, \ldots, F_N$ are given analytical functions, and $C_1, \ldots, C_N$ are arbitrary (real) constants.

- *By an* optimality criterion*, we mean a transitive reflexive relation $\succeq$ on the set of all N-dimensional families.*
- *We say that a criterion is* final *if there exists one and only one optimal family, i.e., family for which $\mathcal{F}_{\text{opt}}$ for which $\forall \mathcal{G}\, (\mathcal{F}_{\text{opt}} \succeq \mathcal{G})$.*

- *For every transformation $T = \lambda \cdot x$ ($\lambda > 0$), and for every i, we define*

$$(T_i(F))(x_1, \ldots, x_n) \stackrel{\text{def}}{=} F(x_1, \ldots, x_{i-1}, T(x_i), x_{i+1}, \ldots, x_n),$$

  and $T_i(\mathcal{F}) \stackrel{\text{def}}{=} \{T_i(F) \mid F \in \mathcal{F}\}$.

- *We say that a criterion $\succeq$ is* scale-*invariant if for every two families $\mathcal{F}$ and $\mathcal{G}$, for every i, and for every linear function $T(x) = \lambda \cdot x$, $\mathcal{F} \succeq \mathcal{G}$ implies $T_i(\mathcal{F}) \succeq T_i(\mathcal{G})$.*

Proposition 1. *Let $\succeq$ be a final scale-invariant optimality criterion on the set of all families. Then, every function F from the optimal family $\mathcal{F}_{\text{opt}}$ is a polynomial.*

Comments.

- This result justifies the Taylor models.
- This result is stronger than the result from [9], since there, we also required that the optimality criterion be invariant wif we change the starting point for measuring x_i.

Proof.

1°. Let us first prove that the optimal family $\mathcal{F}_{\text{opt}}$ is itself scale-invariant, i.e., that for every rescaling T and for every i, we have $T_i(\mathcal{F}_{\text{opt}}) = \mathcal{F}_{\text{opt}}$.

Indeed, let T and i be given. Since $\mathcal{F}_{\text{opt}}$ is optimal, for every other family $\mathcal{G}$, we have $\mathcal{F}_{\text{opt}} \succeq T_i^{-1}(\mathcal{G})$ (where T_i^{-1} means the inverse transformation). Since the optimality criterion $\succeq$ is invariant, we conclude that $T_i(\mathcal{F}_{\text{opt}}) \succeq T_i(T_i^{-1}(\mathcal{G})) = \mathcal{G}$. Since this is true for every family $\mathcal{G}$, the family $T_i(\mathcal{F}_{\text{opt}})$ is also optimal. But since our criterion is final, there is only one optimal family and therefore, $T_i(\mathcal{F}_{\text{opt}}) = \mathcal{F}_{\text{opt}}$.

2°. Since the family $\mathcal{F}_{\text{opt}}$ is scale-invariant, in particular, it means that for every function $F(x_1, \ldots, x_n)$ from this family, and for every $\lambda > 0$, the function

$$F_\lambda(x_1, \ldots, x_n) \stackrel{\text{def}}{=} F(\lambda \cdot x_1, \ldots, \lambda \cdot x_n)$$

also belongs to the optimal family.

3°. Let us now take any function $F(x_1, \ldots, x_n)$ from the optimal family $\mathcal{F}_{\text{opt}}$ and prove that this function is a polynomial.

The family consists of analytical functions, thus the selected function $F(x_1, \ldots, x_n)$ is also analytical.

By definition, an analytical function $f(x_1, \ldots, x_n)$ is an infinite series consisting of monomials $m(x_1, \ldots, x_n)$ of the type

$$a_{i_1 \ldots i_n} \cdot x_1^{i_1} \cdot \ldots \cdot x_n^{i_n}.$$

For each such term, by its *total order*, we will understand the sum $i_1 + \ldots + i_n$. The meaning of this total order is simple: if we multiply each input of this monomial by λ, then the value of the monomial is multiplied by λ^k:

$$m(\lambda \cdot x_1, \ldots \lambda \cdot x_n) = a_{i_1 \ldots i_n} \cdot (\lambda \cdot x_1)^{i_1} \cdot \ldots \cdot (\lambda \cdot x_n)^{i_n} =$$

$$\lambda^{i_1 + \ldots + i_n} \cdot a_{i_1 \ldots i_n} \cdot x_1^{i_1} \cdot \ldots \cdot x_n^{i_n} = \lambda^k \cdot m(x_1, \ldots, x_n).$$

For each order k, there are finitely many possible combinations of integers $i_1, \ldots, i_n$ for which $i_1 + \ldots + i_n = k$, so there are finitely many possible monomials of this order. Let $P_k(x_1, \ldots, x_n)$ denote the sum of all the monomials of order k from the series describing the function $F(x_1, \ldots, x_n)$. Then, we have

$$F(x_1, \ldots, x_n) = P_0 + P_1(x_1, \ldots, x_n) + P_2(x_1, x_2, \ldots, x_n) + \ldots$$

Some of these terms may be zeros – if the original expansion has no monomials of the corresponding order. Let k_0 be the first index for which the term $P_{k_0}(x_1, \ldots, x_n)$ is not identically 0. Then,

$$fF(x_1, \ldots, x_n) = P_{k_0}(x_1, \ldots, x_n) + P_{k_0+1}(x_1, x_2, \ldots, x_n) + \ldots$$

Since the family $\mathcal{F}_{\rm opt}$ is scale-invariant, it also contains the function

$$F_\lambda(x_1, \ldots, x_n) = F(\lambda \cdot x_1, \ldots, \lambda \cdot x_n).$$

At this re-scaling, each term P_k is multiplied by λ^k; thus, we get

$$F_\lambda(x_1, \ldots, x_n) = \lambda^{k_0} \cdot P_{k_0}(x_1, \ldots, x_n) + \lambda^{k_0+1} \cdot P_{k_0+1}(x_1, x_2, \ldots, x_n) + \ldots$$

Since $\mathcal{F}_{\rm opt}$ is a linear space, it also contains a function

$$\lambda^{-k_0} \cdot F_\lambda(x_1, \ldots, x_n) = P_{k_0}(x_1, \ldots, x_n) + \lambda \cdot P_{k_0+1}(x_1, x_2, \ldots, x_n) + \ldots$$

Since $\mathcal{F}_{\rm opt}$ is finite-dimensional, it is closed under turning to a limit. In the limit $\lambda \to 0$, we conclude that the term $P_{k_0}(x_1, \ldots, x_n)$ also belongs to the family $\mathcal{F}_{\rm opt}$.

Since $\mathcal{F}_{\rm opt}$ is a linear space, this means that the difference

$$F(x_1, \ldots, x_n) - P_{k_0}(x_1, \ldots, x_n) =$$

$$P_{k_0+1}(x_1, x_2, \ldots, x_n) + P_{k_0+2}(x_1, x_2, \ldots, x_n) + \ldots$$

also belongs to $\mathcal{F}_{\rm opt}$. If we denote, by k_1, the first index $k_1 > k_0$ for which the term $P_{k_1}(x_1, \ldots, x_n)$ is not identically 0, then we can similarly conclude that this term $P_{k_1}(x_1, \ldots, x_n)$ also belongs to the family $\mathcal{F}_{\rm opt}$, etc.

We can therefore conclude that for every index k for which term $P_k(x_1, \ldots, x_n)$ is not identically 0, this term $P_k(x_1, \ldots, x_n)$ also belongs to the family $\mathcal{F}_{\rm opt}$.

Monomials of different total order are linearly independent. Thus, if there were infinitely many non-zero terms P_k in the expansion of the function

$F(x_1, \ldots, x_n)$, we would have infinitely many linearly independent function in the family $\mathcal{F}_{\text{opt}}$ – which contradicts to our assumption that the family $\mathcal{F}_{\text{opt}}$ is a finite-dimensional linear space.

So, in the expansion of the function $F(x_1, \ldots, x_n)$, there are only finitely many non-zero terms. Hence, the function $F(x_1, \ldots, x_n)$ is a sum of finitely many monomials – i.e., a polynomial.

The proposition is proven.

Comment. As we can see form the proof, to show that every function from the optimal family is a polynomial,w e do not even need to use scale-invariance with respect to each of the variables: it is sufficient to require that the optimality criterion is invariant with respect to a simultaneous re-scaling of all the variable:

$$x_1 \to \lambda \cdot x_1, \ldots, x_n \to \lambda \cdot x_n.$$

3 Why Modified Taylor Models: A Justification

Discussion. In the original Taylor model, coefficients at the unknown functions were real numbers. The main idea behind modified Taylor models is that we can have interval-valued coefficients. Thus, we arrive at the following definition.

Definition 2. *Let $n > 0$, $N > 0$, and $M > 0$ be integers.*

- *By a (N, M)*-family*, we mean a family $\mathcal{F}$ of all interval-valued functions of the type*

$$\{C_1 \cdot F_1(x_1, \ldots, x_n) + \ldots + C_N \cdot F_N(x_1, \ldots, x_n) +$$
$$\mathbf{C}_1 \cdot G_1(x_1, \ldots, x_n) + \ldots + \mathbf{C}_M \cdot G_M(x_1, \ldots, x_n)\},$$

  where $F_1, \ldots, F_N$ and $G_1, \ldots, G_M$ are given analytical functions, $C_1, \ldots, C_N$ are arbitrary (real) constants, and $\mathbf{C}_k = [\underline{C}_k, \overline{C}_k]$ are arbitrary intervals.

- *By an* optimality criterion*, we mean a transitive reflexive relation $\succeq$ on the set of all (N, M)-families.*

- *We say that a criterion is* final *if there exists one and only one optimal family, i.e., family for which $\mathcal{F}_{\text{opt}}$ for which $\forall \mathcal{G}\, (\mathcal{F}_{\text{opt}} \succeq \mathcal{G})$.*

- *For every transformation $T = \lambda \cdot x$ $(\lambda > 0)$, and for every i, we define*

$$(T_i(F))(x_1, \ldots, x_n) \stackrel{\text{def}}{=} F(x_1, \ldots, x_{i-1}, T(x_i), x_{i+1}, \ldots, x_n),$$

  and $T_i(\mathcal{F}) \stackrel{\text{def}}{=} \{T_i(F) \,|\, F \in \mathcal{F}\}$.

- *We say that a criterion $\succeq$ is* scale*-invariant if for every two families $\mathcal{F}$ and $\mathcal{G}$, for every i, and for every linear function $T(x) = \lambda \cdot x$, $\mathcal{F} \succeq \mathcal{G}$ implies $T_i(\mathcal{F}) \succeq T_i(\mathcal{G})$.*

Proposition 2. *Let $\succeq$ be a final scale-invariant optimality criterion on the set of all families. Then, every function F from the optimal family $\mathcal{F}_{\rm opt}$ is a sum of finitely many monomials with interval coefficients.*

Comment. This result justifies the modified Taylor models.

Proof.

1°. First let us notice that each interval $[\underline{C}_k, \overline{C}_k]$ can be represented as $\underline{C}_k + [0, W_i k]$, where $W_k \stackrel{\rm def}{=} \overline{C}_k - \underline{C}_k$ is the width of the k-th interval. Substituting this expression into the general formula for a family, we conclude that each interval-valued function from the family is a linear combination of:

- real-valued functions F_j and G_k and
- functions G_k multiplied by an interval $[0, W_k]$.

Thus, without losing generality, we can safely assume that in each interval $\mathbf{C}_k$, the lower endpoint is 0, i.e., that each such interval has the form $[0, W_k]$ for some $W_k \geq 0$.

2°. Similarly to the proof of Proposition 1, we can prove that the optimal family is scale-invariant, i.e., remains unchanged if we re-scale each variable $x_i \to \lambda_i \cdot x_i$.

In other words, for each interval-valued function from the optimal family,

- if we re-scale all the variables,
- we get an interval-valued function from the same family – but probably corresponding to different coefficients C'_k and W'_k.

So, the re-scaled interval-valued function

$$F_\lambda(x_1, \ldots, x_n)+$$

$$[0, W_1] \cdot \cdot G_1(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n) + \ldots + [0, W_M] \cdot G_M(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n),$$

where

$$F_\lambda(x_1, \ldots, x_n) \stackrel{\rm def}{=} C_1 \cdot F_1(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n) + \ldots + C_N \cdot F_N(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n),$$

coincides with the interval-valued function

$$F'(x_1, \ldots, x_n) + [0, W'_1] \cdot G_1(x_1, \ldots, x_n) + \ldots + [0, W'_M] \cdot G_M(x_1, \ldots, x_n),$$

where

$$F'(x_1, \ldots, x_n) \stackrel{\rm def}{=} C'_1 \cdot F_1(x_1, \ldots, x_n) + \ldots + C'_N \cdot F_N(x_1, \ldots, x_n).$$

Each interval-valued function is a convex set in the class of all functions, obtained by taking all possible values $w_k \in [0, W_k]$.

Since the convex sets coincide, this means that their sets of extreme points should also coincide. These extreme points correspond to extreme values 0 and W_k of the parameters $w_k \in [0, W_k]$. Thus, for the re-scaled family, they are:

$$F_\lambda(x_1, \ldots, x_n), \quad F_\lambda(x_1, \ldots, x_n) + W_1 \cdot G_1(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n), \ldots,$$

$$F_\lambda(x_1, \ldots, x_n) + W_M \cdot G_M(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n),$$

$$F_\lambda(x_1, \ldots, x_n) + W_1 \cdot G_1(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n) + W_2 \cdot G_2(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n), \ldots$$

For the new family. the extreme points are:

$$F'(x_1, \ldots, x_n), \quad F'(x_1, \ldots, x_n) + W'_1 \cdot G_1(x_1, \ldots, x_n), \ldots,$$

$$F'(x_1, \ldots, x_n) + W'_M \cdot G_M(x_1, \ldots, x_n),$$

$$F'(x_1, \ldots, x_n) + W'_1 \cdot G_1(x_1, \ldots, x_n) + W'_2 \cdot G_2(x_1, \ldots, x_n), \ldots$$

For $\lambda_1 = \ldots = \lambda_n = 1$, the first function in the first list coincides with the first function in the second list, etc. Since the dependence on λ_i is continuous, we cannot switch to different equalities, so always:

- the first extreme function from the first list must coincide with the first extreme function from the second list,
- the second extreme function from the first list must coincide with the second extreme function from the second list, etc.

Equality of the first terms means that, for every tuple $C_1, \ldots, C_N$ and for every tuple $\lambda_1, \ldots, \lambda_n$, there exist values $C'_1, \ldots, C'_N$ for which

$$F_\lambda(x_1, \ldots, x_n) = C_1 \cdot F_1(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n) + \ldots + C_N \cdot F_N(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_n) =$$

$$F'(x_1, \ldots, x_n) = C'_1 \cdot F_1(x_1, \ldots, x_n) + \ldots + C'_N \cdot F_N(x_1, \ldots, x_n).$$

This means that the class of all functions

$$C_1 \cdot F_1(x_1, \ldots, x_n) + \ldots + C_N \cdot F_N(x_1, \ldots, x_n)$$

corresponding to different values C_j is scale-invariant. Thus, based on Proposition 1, all the functions $F_j(x_1, \ldots, x_n)$ are polynomials – i.e., a sum of finitely many monomials.

For each k from 1 to M, since both the first and the $(k+1)$-st terms in the two lists are equal to each other, we conclude the differences between these term should also be equal. Thus, we conclude that

$$W_k \cdot G_k(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_k) = W'_k \cdot G_k(x_1, \ldots, x_n).$$

For $W_1 = 1$, this means that for every tuple $\lambda_1, \ldots, \lambda_n$, there exists a value W'_k for which

$$G_k(\lambda_1 \cdot x_1, \ldots, \lambda_n \cdot x_k) = W'_k \cdot G_k(x_1, \ldots, x_n).$$

The function $G_k(x_1, \ldots, x_n)$ is an analytical function and is, thus, the sum of monomials $c_{i_1 \ldots i_n} \cdot x_1^{i_1} \cdot \ldots \cdot x_n^{i_n}$. Under re-scaling $x_i \to \lambda_i \cdot x_i$, each monomial is multiplied by different coefficients $\lambda_1^{i_1} \cdot \ldots \cdot \lambda_n^{i_n}$. So, the only case when the whole sum of monomials is multiplied by the same number W'_k is when the function $G_k(x_1, \ldots, x_n)$ consists of a single monomial.

Thus, each interval-valued function from the optimal family is indeed a sum of finitely many monomials $G_k(x_1, \ldots, x_n)$ with interval coefficients. The proposition is proven.

Acknowledgments

This work was supported in part by the US National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

References

[1] M. Berz and K. Makino, "Rigorouus global search using Taylor models", *Proceedings of the 2009 Conference on Symbolic Numeric Computations SNC'09*, Kyoto, Japan August 3–5, 2009. pp. 11–20.

[2] M Berz and K. Makino, *COSY INFINITY*, http://cosyinfinity.org/

[3] M. Berz, K, Makino, and Y.-K. Kim, "Long-term stability of the Tevatron by verified global optimization", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors, and Associated Equipment*, 2006, Vol. 558, No. 1, pp. 1–10.

[4] S. Chevillard, J. Harrison, M. Joldeş, and Ch. Lauter, "Efficient and accurate computation of upper bounds of approximaton error", *Theoretical Computer Science*, 2011, Vol. 412, pp. 1523–1543.

[5] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics, Addison Wesley*, Boston, Massachusetts, 2005.

[6] O. Kosheleva, M. Ceberio, and V. Kreinovich, "Why tensors?", In: M. Ceberio and V. Kreinovich (eds.), *Constraint Programming and Decision Making*, Springer Verlag, Berlin, Heidelberg, 2014, pp. 75–78.

[7] K. Makino, *Rigorous Analysis of Nonlinear Motion in Particle Accelerators*, Ph.D. Dissertation, Michigan State University, East Lansing, Michigan, 1998.

[8] K. Makino and M. Berz, "Taylor models and other validated functional inclusion methods", *International Journal of Pure and Applied Mathematics*, 2003, Vol. 4, No. 4, pp. 379–456.

[9] N. S. Nedialkov, V. Kreinovich, and S. A. Starks, "Interval Arithmetic, affine arithmetic, Taylor series methods: why, what next?", *Numerical Algorithms,* 2004, Vol. 37, pp. 325–336.

[10] A. Neumaier, "Taylor forms – use and limits", *Reliable Computing*, 2003, Vol. 9, No. 1, pp. 43–79.