



HAL
open science

Generalized Homogeneous Artificial Neural Network and Applications

Andrey Polyakov

► **To cite this version:**

Andrey Polyakov. Generalized Homogeneous Artificial Neural Network and Applications. 2024. hal-04313941v2

HAL Id: hal-04313941

<https://inria.hal.science/hal-04313941v2>

Preprint submitted on 24 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Generalized Homogeneous Artificial Neural Network and Applications

Andrey Polyakov *

Abstract

The paper proposes an artificial neural network (ANN) being a global approximator for a special class of functions, which are known as *generalized homogeneous*. The homogeneity means a symmetry of a function with respect to a group of transformations having topological characterization of a dilation. In this paper, a class of the so-called linear dilations is considered. A homogeneous universal approximation theorem is proven. Procedures for a transformation of an existing ANN to a homogeneous one are developed. Theoretical results are illustrated by several academic examples highlighting potential applications of homogeneous ANNs in various domains such as systems theory, automatic control and computer science. In particular, a scaling invariant pattern recognition by homogeneous ANN is demonstrated. An ANN-based identification of generalized homogeneous dynamical system is considered. A feedback for robust homogeneous stabilization of multi-input linear time-invariant system is designed using the proposed ANN.

1 Introduction

1.1 State of the art

The universal approximation theorems [7], [13], [12] put limits on what artificial neural networks (ANNs) can theoretically learn. These theorems guarantee the existence of ANN, which approximates a continuous function on a compact set with an arbitrary high precision. A training of the ANN is based on compactly supported data as well, while, the trained ANN may be utilized next as a predictor of the function value for an input data non-belonging to the training set. Sometimes, the new input may be have a rather large distance even from a convex hull of the training set. In the latter case, the ANN is utilized as an extrapolator of the function and the approximation theorems are inapplicable. Moreover, the analysis of the extrapolation error is impossible if there is no information about the function behavior away from the training set. Therefore, a global extrapolation of a function based on a local data can be provided only under additional assumption about the function approximated by ANN. This paper deals with approximation of the so-called generalized homogeneous functions [36], [16], [4], [28] and introduces the corresponding homogeneous artificial neural network, which key feature is a global approximation based on local data.

The generalized homogeneity is a symmetry of an object (a function, a set, a vector field, etc) with respect to a group of the so-called generalized dilations. Recall that the standard (classical) homogeneity were introduced by Leonhard Euler in 18th century as the symmetry of a function $x \mapsto f(x)$ with respect to the standard (isotropic) dilation of its argument $x \mapsto \lambda x$, namely, $f(\lambda x) = \lambda^\nu f(x), \forall \lambda > 0, \forall x$, where $\nu \in \mathbb{R}$ is the homogeneity degree. To the best of author's knowledge, generalized (anisotropic) dilations are studied in systems theory since 1950s (see, e.g., [36]), where, instead of multiplication of the argument x on a scalar $\lambda > 0$, some group of transformations having topological characterization of a dilation is utilized (see, e.g. [17], [16], [14]). In particular, if $\mathbf{d}(s) : \mathbb{V} \rightarrow \mathbb{V}, s \in \mathbb{R}$ is a one parameter group of coordinate transformations on a normed vector space \mathbb{V} such that $\lim_{s \rightarrow \pm\infty} \|\mathbf{d}(s)x\| = e^{\pm\infty}, \forall x \neq \mathbf{0}$, then \mathbf{d} is a dilation in \mathbb{V} and a functional $h : \mathbb{V} \rightarrow \mathbb{R}$ satisfying $h(\mathbf{d}(s)x) = e^{\nu s}h(x)$ is said to be \mathbf{d} -homogeneous of degree ν . In this framework, the standard dilation corresponds to the dilation $\mathbf{d}(s)x = e^s x, s \in \mathbb{R}, x \in \mathbb{V}$.

Generalized homogeneous mathematical objects are studied in various domains of systems theory [9], [11], [16], [32], [8]. Many models of mathematical physics [28] are homogeneous in a generalized sense. Homogeneous models also appear as local approximations of nonlinear systems [1]. The control systems design is frequently based on generalized homogeneity (see, e.g., [10], [21], [26], [30]). This paper deals with functions $\mathbb{R}^n \rightarrow \mathbb{R}$, which are symmetric (\mathbf{d} -homogeneous) with respect to a linear dilation $\mathbf{d}(s) = e^{sG_{\mathbf{d}}}, s \in \mathbb{R}$, where $G_{\mathbf{d}} \in \mathbb{R}^{n \times n}$ is an anti-Hurwitz matrix¹. The homogeneity implies certain equivalence of local and global information, since any local knowledge about a function can be expanded globally using the dilation symmetry. The ANN-based approximation should take into account this symmetry, otherwise, an information about geometry of h may be lost and extrapolation capabilities of the ANN may be seriously degraded. For example, the dilation symmetry (scaling invariance) is important for the pattern recognition [5], which is expected to be invariant with respect to a zoom (standard dilation) of the image. The control theory provides many other examples of the generalized homogeneous systems (see, e.g., [10], [4], [6]), [28] and references therein).

*Univ. Lille, Inria, CNRS, UMR 9189 CRIStAL, Centrale Lille, F-59000 Lille, France, (andrey.polyakov@inria.fr)

¹All eigenvalues have positive real parts

The dilation symmetry (namely, standard homogeneity) is shown to be useful for analysis and learning of ANNs and, in particular, differential neural networks [3]. In [18], [24], [2], [15] it is shown that ANNs with standard homogeneous activation functions have various useful features such as faster training. However, the homogeneity of the activation function does not guarantee a dilation symmetry of the input-output mapping defined by ANN. This input-output symmetry is the key feature of the ANNs studied in this paper. Independently of the activation function, the introduced ANN is a generalized homogeneous function of a desired degree μ . To the best of author's knowledge, such homogeneous ANNs have not been studied before.

1.2 Basic idea of homogeneous approximation

Following the classical model of a controlled rotational motion of the rigid body in the 3-D space, the dynamics of the angular velocity of the rigid body is governed by

$$\dot{\omega}(t) = J^{-1}(-\omega \times J\omega + \tau), \quad t \geq 0, \quad (1)$$

where $\omega(t) = (\omega_1(t), \omega_2(t), \omega_3(t))^T \in \mathbb{R}^3$ is the angular velocity in the body frame, $J = J^T \in \mathbb{R}^{3 \times 3}$ is the inertia matrix, $\tau(t) \in \mathbb{R}^3$ is the external momentum (e.g., control input) and \times denotes the vector product, i.e., $\omega \times J\omega = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_2 & 0 \end{pmatrix} J\omega$. The function $h : \mathbb{R}^6 \rightarrow \mathbb{R}^3$ given by $h(\omega, \tau) = -J^{-1}(\omega \times J\omega + \tau)$ is homogeneous of degree 2 with respect to a generalized dilation:

$$h(e^s \omega, e^{2s} \tau) = e^{2s} h(\omega, \tau), \quad \forall s \in \mathbb{R}, \forall \omega, \tau \in \mathbb{R}^3.$$

This means that the right-hand side of the differential equation (1) is homogeneous. Considering the problem of data-driven approximation the dynamic equation, if h_ϵ is an ANN approximation of h on a compact $\Omega \subset \mathbb{R}^6$ such that $|h(\omega, \tau) - h_\epsilon(\omega, \tau)| \leq \epsilon$ for all $(\omega, \tau) \in \Omega$, then $h_\epsilon^s(\omega, \tau) := e^{2s} h_\epsilon(e^{-s} \omega, e^{-2s} \tau)$ is an approximation of h on the set $\mathbf{d}(s)\Omega$, where $\mathbf{d}(s) = \text{diag}(e^s I_3, e^{2s} I_3)$ is the generalized dilation. Indeed, due to \mathbf{d} -homogeneity of h , the approximation error for h_ϵ^s admits the estimate $|h(\omega, \tau) - h_\epsilon^s(\omega, \tau)| = e^{2s} |h(e^{-s} \omega, e^{-2s} \tau) - h_\epsilon(e^{-s} \omega, e^{-2s} \tau)| \leq e^{2s} \epsilon$ for all $(\omega, \tau) \in \mathbf{d}(s)\Omega$. Since $s \in \mathbb{R}$ is an arbitrary number, then the mapping h_ϵ^s can be interpreted as a homogeneous extrapolation of h_ϵ away from Ω . This paper utilizes such an extrapolation for a global ANN approximator design.

1.3 Contributions

The paper presents the following contributions to the approximation theory by ANNs and its applications:

- A \mathbf{d} -homogeneous ANN with one hidden level is designed. Such an ANN is shown to be a global approximator of any continuous \mathbf{d} -homogeneous function on \mathbb{R}^n . The corresponding homogeneous approximation theorem is proven (Theorem 3). To illustrate a possible application of homogeneous ANN in computer sciences, an example of the pattern recognition is considered.
- A methodology of transformation of existing well-trained ANN to a homogeneous one is developed, theoretically justified and supported by numerical examples from systems theory. An integral version of the homogeneous Euler function theorem is proven for this purpose (see Theorem 1). The upgrade assumes that a homogeneous mapping is already approximated by the conventional ANN locally, but the homogeneity parameters (such as the homogeneity degree or the dilation group) are unknown. Based on available data, these parameters are identified and, next, the existing ANN is homogenized (transformed to a homogeneous ANN) in order to make it a global approximator of the homogeneous function. This procedure can be interpreted as a *homogenization of ANN*.
- A generalized homogeneous stabilizer for linear time-invariant system is designed using proposed ANN approximation. The design uses homogeneous ANN for explicit approximation of the implicit homogeneous control [30], [34]. Stability of the closed-loop system with such an explicit homogeneous approximation of the implicit control law is studied (Theorem 5).

1.4 Structure of the paper

Preliminaries about generalized homogeneous functions are given in Section 2. Homogeneous ANNs are introduced in Section 3, where, in particular, the homogeneous universal approximation theorem is proven. As a possible application of homogeneous ANN, the problem of scaling invariant pattern recognition is considered. The Section 4 is devoted to transformation of the classical ANN to a homogeneous one. A methodology of the data-driven upgrade of a conventional ANN to a homogeneous one is developed. The proposed methodology is illustrated by ANN identification of the homogeneous dynamical system (1). Section 5 is devoted to application of homogeneous ANN to robust stabilization of linear time-invariant system. The MATLAB codes of examples considered in the paper are available: <https://gitlab.inria.fr/polyakov/homogeneous-artificial-neural-networks>.

1.5 Notation

$\mathbb{N} = \{1, 2, \dots\}$ is the set of natural number; \mathbb{R} is the field of reals; $\mathbb{R}_0^n = \mathbb{R}^n \setminus \{\mathbf{0}\}$, where $\mathbf{0}$ is the zero element of a vector space (e.g., $\mathbf{0} \in \mathbb{R}^n$ means that $\mathbf{0}$ is the zero vector); $\|\cdot\|$ is a norm in \mathbb{R}^n (to be specified later); $S = \{x \in \mathbb{R}^n : \|x\| = 1\}$; the matrix norm for $A \in \mathbb{R}^{n \times n}$ is defined as $\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$; $\lambda_{\min}(P)$ denote a minimal eigenvalue of a symmetric matrix $P = P^\top \in \mathbb{R}^{n \times n}$; $P \succ 0$ means that the symmetric matrix P is positive definite; $A^{-\top} = (A^{-1})^\top$ for invertible matrix $A \in \mathbb{R}^{n \times n}$; $C(\Omega_1, \Omega_2)$ denotes the set of continuous functions $\Omega_1 \subset \mathbb{R}^n \rightarrow \Omega_2 \subset \mathbb{R}^m$; $C_c(\mathbb{R}^n, \mathbb{R})$ denotes the set of compactly supported continuous functions; $L^p(\mathbb{R}^n, \mathbb{R})$ is the Lebesgue space.

2 Generalized Homogeneity

2.1 Linear Dilation

Recall that a family of operators $\mathbf{d}(s) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ parameterized by the scalar $s \in \mathbb{R}$ is a one-parameter group if $\mathbf{d}(0)x = x$, $\mathbf{d}(s)\mathbf{d}(t)x = \mathbf{d}(s+t)x$, $\forall x \in \mathbb{R}^n, \forall s, t \in \mathbb{R}$. A group \mathbf{d} is a) *continuous* if the mapping $s \mapsto \mathbf{d}(s)x$ is continuous, $\forall x \in \mathbb{R}^n$; b) *linear* if $\mathbf{d}(s)$ is a linear mapping (i.e., $\mathbf{d}(s) \in \mathbb{R}^{n \times n}$), $\forall s \in \mathbb{R}$; c) a *dilation* [16] in \mathbb{R}^n if $\liminf_{s \rightarrow +\infty} \|\mathbf{d}(s)x\| = +\infty$ and $\limsup_{s \rightarrow -\infty} \|\mathbf{d}(s)x\| = 0$, $\forall x \neq \mathbf{0}$. Any linear continuous group in \mathbb{R}^n admits the representation [25]

$$\mathbf{d}(s) = e^{sG_{\mathbf{d}}} = \sum_{j=1}^{\infty} \frac{s^j G_{\mathbf{d}}^j}{j!}, \quad s \in \mathbb{R}, \quad (2)$$

where $G_{\mathbf{d}} \in \mathbb{R}^{n \times n}$ is a generator of \mathbf{d} . A continuous linear group (2) is a dilation in \mathbb{R}^n if and only if $G_{\mathbf{d}}$ is an anti-Hurwitz matrix. In this paper we deal only with *linear continuous dilations*. A dilation \mathbf{d} in \mathbb{R}^n is i) *monotone* if the function $s \mapsto \|\mathbf{d}(s)x\|$ is strictly increasing, $\forall x \neq \mathbf{0}$; ii) *strictly monotone* if $\exists \beta > 0$ such that $\|\mathbf{d}(s)x\| \leq e^{\beta s} \|x\|$, $\forall s \leq 0, \forall x \in \mathbb{R}^n$.

The following result is the straightforward consequence of the existence of the quadratic Lyapunov function for stable linear system [22].

Corollary 1 *A linear continuous dilation in \mathbb{R}^n is strictly monotone with respect to the weighted Euclidean norm $\|x\| = \sqrt{x^\top P x}$ with $0 \prec P \in \mathbb{R}^{n \times n}$ if and only if $PG_{\mathbf{d}} + G_{\mathbf{d}}^\top P \succ 0, P \succ 0$.*

The standard dilation corresponds to $G_{\mathbf{d}_1} = I_n$ and $\mathbf{d}_1(s) = e^s I_n$. The generator of the weighted dilation [36] is $G_{\mathbf{d}_2} = \text{diag}(r_1, \dots, r_n)$.

2.2 Canonical homogeneous norm

By definition, a norm in \mathbb{R}^n is a function $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}_+$ satisfying the following three conditions: 1) *Positive Definiteness*: $\|u\| = 0$ if and only if $u = \mathbf{0}$; 2) *Dilation+Reflection Symmetry*: $\|\pm e^s u\| = e^s \|u\|$, $\forall s \in \mathbb{R}, \forall u \in \mathbb{R}^n$; 3) *Subadditivity (Triangle Inequality)*: $\|u+v\| \leq \|u\| + \|v\|$, $\forall u, v \in \mathbb{R}^n$.

In the literature, the second condition is frequently called by the homogeneity or the positive homogeneity, but, in fact, it combines two symmetries. One of them is, indeed, the standard homogeneity (= a symmetry with respect to the standard dilation). Omitting the third condition and replacing the standard dilation with a generalized one, the so-called homogeneous norm is introduced.

A continuous positive definite function $\|\cdot\|_{\mathbf{d}} : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfying $\|\pm \mathbf{d}(s)x\|_{\mathbf{d}} = e^s \|x\|_{\mathbf{d}}, \forall x \in \mathbb{R}^n$ is called a *\mathbf{d} -homogeneous norm* (or “dilated norm”) [16], [10], [1]. For the weighted dilation $\mathbf{d}(s) = \begin{pmatrix} e^{2s} & 0 \\ 0 & e^s \end{pmatrix}$, a \mathbf{d} -homogeneous norm can be defined, for example, as $\|x\|_{\mathbf{d}} = \sqrt{|x_1|} + |x_2|$, where $x = (x_1, x_2)^\top \in \mathbb{R}^2$.

The \mathbf{d} -homogeneous norm $\|\cdot\|_{\mathbf{d}}$ does not satisfy the triangle inequality in the general case, so, formally, this is not even a semi-norm. However, we would follow the tradition accepted in systems sciences and call the function $\|\cdot\|_{\mathbf{d}}$ satisfying the above property by a “norm”, but, basically, we use this name for the canonical homogeneous norm (see below) being a norm (in the classical sense) for the vector space $\mathbb{R}_{\mathbf{d}}^n$ homeomorphic to \mathbb{R}^n . The space $\mathbb{R}_{\mathbf{d}}^n$ has the same set of elements as \mathbb{R}^n , but the rules for addition of vectors and multiplication of a vector by a scalar are modified by means of a special homeomorphism on \mathbb{R}^n (see [28] for more details). Let a linear continuous dilation \mathbf{d} have a diagonalizable generator, i.e., $\exists J \in \mathbb{R}^{n \times n}, \det(J) \neq 0$ such that

$$G_{\mathbf{d}} = J^{-1} \Lambda J, \quad \Lambda = \text{diag}\{r_1, r_2, \dots, r_n\} \succ 0. \quad (3)$$

Then, for any $0 \prec Q = Q^\top \in \mathbb{R}^{n \times n}$, the function $\|\cdot\|_{\bar{r}} : \mathbb{R}^n \mapsto \mathbb{R}$ defined as follows

$$\|x\|_{\bar{r}} = \left(\Psi^\top(x) Q \Psi(x) \right)^{\frac{1}{2\bar{r}}}, \quad \Psi(x) = \begin{pmatrix} |e_1^\top J x|^{\frac{\bar{r}}{r_1}} \\ |e_2^\top J x|^{\frac{\bar{r}}{r_2}} \\ \dots \\ |e_n^\top J x|^{\frac{\bar{r}}{r_n}} \end{pmatrix}, \quad \bar{r} > 0, \quad (4)$$

is a \mathbf{d} -homogeneous norm in \mathbb{R}^n , where $e_i = (0, \dots, 1, \dots, 0)^\top \in \mathbb{R}^n$ is the unit vector and $[\cdot]^\alpha = |\cdot|^\alpha \text{sign}(\cdot)$. Indeed, on the one hand, taking into account $\mathbf{d}(s) = e^{sG_{\mathbf{d}}} = J^{-1}e^{s\Lambda}J$, we derive

$$\Psi(\mathbf{d}(s)x) = \begin{pmatrix} [e_1^\top J\mathbf{d}(s)x]^\frac{\bar{r}_1}{r_1} \\ [e_2^\top J\mathbf{d}(s)x]^\frac{\bar{r}_2}{r_2} \\ \dots \\ [e_n^\top J\mathbf{d}(s)x]^\frac{\bar{r}_n}{r_n} \end{pmatrix} = \begin{pmatrix} [e^{r_1 s} e_1^\top Jx]^\frac{\bar{r}_1}{r_1} \\ [e^{r_2 s} e_2^\top Jx]^\frac{\bar{r}_2}{r_2} \\ \dots \\ [e^{r_n s} e_n^\top Jx]^\frac{\bar{r}_n}{r_n} \end{pmatrix} = e^{\bar{r}s} \Psi(x),$$

and

$$\|\mathbf{d}(s)x\|_{\bar{r}} = e^s \|x\|_{\bar{r}}, \quad \forall s \in \mathbb{R}, \forall x \in \mathbb{R}^n.$$

On the other hand, since J is invertible then, by construction, Ψ is a homeomorphism on \mathbb{R}^n . The inverse of Ψ is given by

$$\Psi^{-1}(y) = J^{-1} \begin{pmatrix} [y_1]^\frac{r_1}{\bar{r}_1} \\ [y_2]^\frac{r_2}{\bar{r}_2} \\ \dots \\ [y_n]^\frac{r_n}{\bar{r}_n} \end{pmatrix}, \quad y = (y_1, \dots, y_n)^\top \in \mathbb{R}^n.$$

In this case, $\Psi(x) = \mathbf{0} \Leftrightarrow x = \mathbf{0}$, so $\|x\|_{\bar{r}}$ is a \mathbf{d} -homogeneous norm.

Geometrically (see [31, §15]), a conventional norm $\|\cdot\|_*$ in \mathbb{R}^n can be constructed by means of a standard dilation applied to a convex central-symmetric compact neighborhood $B \subset \mathbb{R}^n$ of the origin (e.g., B can be an ellipsoid/ball in \mathbb{R}^n centered at zero). Indeed, let the boundary of the set B define the unit level set (i.e., we assign $\|x\|_* = 1, \forall x \in \partial B$), but the boundary of the scaled set $e^s B$ corresponds to the level e^s (i.e., we assign $\|x\|_* = e^s, \forall x \in \partial(e^s B)$), where $s \in \mathbb{R}$. Let $\|\mathbf{0}\|_* = 0$. The function $\|\cdot\|_*$ constructed in such a way is a norm in \mathbb{R}^n . Indeed, since the set B is a convex neighborhood of $\mathbf{0}$ then $\partial(e^s B) \cap \partial(e^t B) = \emptyset$ for $s \neq t$. This means that the function $x \mapsto \|x\|_*$ is single-valued. The positive definiteness comes from the fact that B is a neighborhood of the origin: $\mathbf{0} \in \text{int } B$. The homogeneity is obtained by construction. Reflection symmetry of $\|\cdot\|_*$ is due to the central symmetry of B . The triangle inequality comes from convexity of B and, consequently, convexity of $\|\cdot\|_*$. A \mathbf{d} -homogeneous norm can be designed similarly by means of a linear dilation \mathbf{d} being monotone with respect to a norm $\|\cdot\|$ in \mathbb{R}^n . Such a homogeneous norm was called *canonical* [27], [29].

Definition 1 Let a linear dilation \mathbf{d} in \mathbb{R}^n be continuous and monotone with respect to a norm $\|\cdot\|$. A function $\|\cdot\|_{\mathbf{d}} : \mathbb{R}^n \rightarrow [0, +\infty)$ defined as follows: $\|\mathbf{0}\|_{\mathbf{d}} = 0$ and

$$\|x\|_{\mathbf{d}} = e^{s_x}, \quad \text{where } s_x \in \mathbb{R} : \|\mathbf{d}(-s_x)x\| = 1, \quad x \neq \mathbf{0} \quad (5)$$

is said to be a *canonical \mathbf{d} -homogeneous norm* in \mathbb{R}^n

For standard dilation $\mathbf{d}(s) = e^s I_n$ we have $\|x\|_{\mathbf{d}} = \|x\|$. In other cases, the canonical homogeneous norm $\|x\|_{\mathbf{d}}$ with $x \neq \mathbf{0}$ is implicitly defined by the nonlinear algebraic equation (5), which always have a unique solution due to monotonicity of the dilation. In some particular cases, this implicit equation has explicit solution.

Lemma 1 [28] If a linear continuous dilation \mathbf{d} in \mathbb{R}^n is monotone with respect to a norm $\|\cdot\|$ then 1) the function $\|\cdot\|_{\mathbf{d}} : \mathbb{R}^n \rightarrow [0, +\infty)$ implicitly defined by (5) is single-valued and continuous on \mathbb{R}^n ; 2) there exist $\sigma_1, \sigma_2 \in \mathcal{K}_\infty$:

$$\sigma_1(\|x\|_{\mathbf{d}}) \leq \|x\| \leq \sigma_2(\|x\|_{\mathbf{d}}), \quad \forall x \in \mathbb{R}^n; \quad (6)$$

3) $\|\cdot\|$ is locally Lipschitz continuous on $\mathbb{R}^n \setminus \{\mathbf{0}\}$; 4) $\|\cdot\|_{\mathbf{d}}$ is continuously differentiable on $\mathbb{R}^n \setminus \{\mathbf{0}\}$ provided that $\|\cdot\|$ is continuously differentiable on $\mathbb{R}^n \setminus \{\mathbf{0}\}$ and \mathbf{d} is strictly monotone.

For the \mathbf{d} -homogeneous norm $\|x\|_{\mathbf{d}}$ induced by the weighted Euclidean norm $\|x\| = \sqrt{x^\top P x}$ (see Corollary 1) we have [28]

$$\frac{\partial \|x\|_{\mathbf{d}}}{\partial x} = \|x\|_{\mathbf{d}} \frac{x^\top \mathbf{d}^\top (-\ln \|x\|_{\mathbf{d}}) P \mathbf{d} (-\ln \|x\|_{\mathbf{d}})}{x^\top \mathbf{d}^\top (-\ln \|x\|_{\mathbf{d}}) P G_{\mathbf{d}} \mathbf{d} (-\ln \|x\|_{\mathbf{d}}) x}, \quad (7)$$

$$\sigma_1(\rho) = \begin{cases} \rho^\alpha & \text{if } \rho \leq 1, \\ \rho^\beta & \text{if } \rho > 1, \end{cases} \quad \sigma_2(\rho) = \begin{cases} \rho^\beta & \text{if } \rho \leq 1, \\ \rho^\alpha & \text{if } \rho > 1, \end{cases} \quad (8)$$

where $\alpha = \frac{\lambda_{\max}(P^{\frac{1}{2}} G_{\mathbf{d}} P^{-\frac{1}{2}} + P^{-\frac{1}{2}} G_{\mathbf{d}}^\top P^{\frac{1}{2}})}{2} \geq \beta = \frac{\lambda_{\min}(P^{\frac{1}{2}} G_{\mathbf{d}} P^{-\frac{1}{2}} + P^{-\frac{1}{2}} G_{\mathbf{d}}^\top P^{\frac{1}{2}})}{2} > 0$.

All \mathbf{d} -homogeneous norms are equivalent, i.e., there exist $0 < k_1 \leq k_2 < +\infty$ such that $k_1 \|x\|_{\mathbf{d},1} \leq \|x\|_{\mathbf{d},2} \leq k_2 \|x\|_{\mathbf{d},1}, \forall x \in \mathbb{R}^n$, where $\|\cdot\|_{\mathbf{d},1}$ and $\|\cdot\|_{\mathbf{d},2}$ are \mathbf{d} -homogeneous norms.

In Section 5 we present an ANN-based approximation of $\|\cdot\|_{\mathbf{d}}$, which uses the so-called \mathbf{d} -homogeneous projector on the unit sphere

$$\pi_{\mathbf{d}}(x) = \mathbf{d}(-\ln \|x\|_{\mathbf{d}})x, \quad x \neq \mathbf{0}.$$

Obviously, $\|\pi_{\mathbf{d}}(x)\|_{\mathbf{d}} = 1$ and $\pi_{\mathbf{d}}(\pi_{\mathbf{d}}(x)) = \pi_{\mathbf{d}}(x), \forall x \neq \mathbf{0}$, so π is, indeed, a projector on the unit sphere $\{x \in \mathbb{R}^n : \|x\|_{\mathbf{d}} = 1\}$. In the case of the canonical homogeneous norm this sphere coincides with $S = \{x \in \mathbb{R}^n : \|x\| = 1\}$, where $\|\cdot\|$ is the conventional norm in \mathbb{R}^n that induces $\|\cdot\|_{\mathbf{d}}$ (see Definition 1).

2.2.1 Homogeneous functions

Below we study functions being symmetric with respect to a linear dilation \mathbf{d} . The dilation symmetry introduced by the following definition is known as the generalized homogeneity [36], [16], [32], [4], [28].

Definition 2 [16] Let \mathbf{d} be a dilation in \mathbb{R}^n . A function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is \mathbf{d} -homogeneous of degree $\nu \in \mathbb{R}$ if

$$h(\mathbf{d}(s)x) = e^{\nu s}h(x), \quad \forall s \in \mathbb{R}, \quad \forall x \in \mathbb{R}^n. \quad (9)$$

Local behavior of any homogeneous function h can be expanded globally taking into account the homogeneity degree of the function. The sign of the homogeneity degree also defines some properties of the function [4].

Proposition 1 Let $h : \mathbb{R}^n \rightarrow \mathbb{R}$ be a \mathbf{d} -homogeneous function of degree $\nu \in \mathbb{R}$, $h \neq \mathbf{0}$ and $\sup_{x \in S} |h(x)| < +\infty$.

- If $\nu > 0$ then h is bounded on any \mathbf{d} -homogeneous ball $B_{\mathbf{d}}(r) = \{x \in \mathbb{R}^n : \|x\|_{\mathbf{d}} \leq r\}$ with $r > 0$ and

$$\lim_{\|x\| \rightarrow 0} h(x) = 0 \quad \text{and} \quad \limsup_{\|x\| \rightarrow +\infty} |h(x)| = +\infty.$$

- If $\nu < 0$ then h is bounded on any set $\mathbb{R}^n \setminus B_{\mathbf{d}}(r)$,

$$\lim_{\|x\| \rightarrow +\infty} h(x) = 0 \quad \text{and} \quad \limsup_{\|x\| \rightarrow 0} |h(x)| = +\infty.$$

- If $\deg_{\mathbf{d}}(h) = 0$ then h is uniformly bounded on \mathbb{R}^n ; moreover, $h \equiv \text{const}$ if h is continuous at $\mathbf{0}$.
- A non-constant homogeneous function h is continuous at zero if and only if $\deg_{\mathbf{d}}(h) > 0$.

Since the identity (9) with $s = -\ln \|x\|_{\mathbf{d}}$ yields

$$h(u) = \|x\|_{\mathbf{d}}^{\nu} h(\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x)$$

then the properties of h claimed above follow from the inclusion $\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x \in S$ and from the properties of the canonical homogeneous norm (see Lemma 1).

Euler homogeneous function theorem is one of fundamental results characterizing homogeneous mappings. Traditionally it relates a smooth homogeneous function and its derivative. The following extension of the Euler homogeneous function theorem also links a locally Lipschitz homogeneous function with its integral and deals with the generalized homogeneity. Let us denote

$$R(x, x^*) = \left\{ (y_1, \dots, y_n)^{\top} \in \mathbb{R}^n : \underline{x}_i \leq y_i \leq \bar{x}_i, i = 1, \dots, n \right\},$$

where $x = (x_1, \dots, x_n)^{\top} \in \mathbb{R}^n$, $x^* = (x_1^*, \dots, x_n^*)^{\top} \in \mathbb{R}^n$ and $\underline{x}_i = \min\{x_i, x_i^*\}$, $\bar{x}_i = \max\{x_i, x_i^*\}$.

Theorem 1

Let a function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ be locally Lipschitz continuous on $\mathbb{R}^n \setminus \{\mathbf{0}\}$ and $h(\mathbf{0}) = 0$. Then the following three claims are equivalent

- 1) h is \mathbf{d} -homogeneous of degree $\nu \in \mathbb{R}$;
- 2) the identity

$$\frac{\partial h(x)}{\partial x} G_{\mathbf{d}} x \stackrel{\text{a.e.}}{=} \nu h(x), \quad (10)$$

holds almost everywhere on \mathbb{R}^n , where $x \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and $G_{\mathbf{d}} \in \mathbb{R}^{n \times n}$ is a generator of the linear dilation \mathbf{d} ;

- 3) the identity

$$\int_{x^*}^x (\text{trace}(G_{\mathbf{d}}) + \nu) h(y) - \sum_{i: x_i \neq x_i^*} \Delta_i(h, y, x, x^*, G_{\mathbf{d}}) dy = 0 \quad (11)$$

holds for $x = (x_1, \dots, x_i, \dots, x_n)^{\top}$, $x^* = (x_1^*, \dots, x_i^*, \dots, x_n^*)^{\top}$ such that $R(x, x^*) \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, where $\int_{x^*}^x = \int_{x_1^*}^{x_1} \dots \int_{x_n^*}^{x_n}$, $dy = dy_n \dots dy_1$, $i = 1, \dots, n$, $e_i = (0, \dots, 1, \dots, 0)^{\top}$ and

$$\Delta_i(h, y, x, x^*, G_{\mathbf{d}}) = \frac{h(y_{x_i}) e_i^{\top} G_{\mathbf{d}} y_{x_i} - h(y_{x_i^*}) e_i^{\top} G_{\mathbf{d}} y_{x_i^*}}{x_i - x_i^*}$$

with $y_{x_i} = (y_1, \dots, x_i, \dots, y_n)^{\top}$, $y_{x_i^*} = (y_1, \dots, x_i^*, \dots, y_n)^{\top}$.

Moreover, if h is differentiable on \mathbb{R}^n (resp., $\mathbb{R}^n \setminus \{\mathbf{0}\}$), the identity (10) is defined everywhere on \mathbb{R}^n (resp., $\mathbb{R}^n \setminus \{\mathbf{0}\}$).

Proof. Let $\|\cdot\|_{\mathbf{d}}$ be the canonical homogeneous norm induced by a weighted Euclidean norm.

1) \Rightarrow 2). Let $h \in \mathcal{H}_{\mathbf{d}}(\mathbb{R}^n)$ and $\nu = \deg_{\mathbf{d}}(h) \in \mathbb{R}$. Due to homogeneity, we have

$$\begin{aligned} h(x) &= h(\mathbf{d}(\ln \|x\|_{\mathbf{d}}) \mathbf{d}(-\ln \|x\|_{\mathbf{d}}) x) \\ &= \|x\|_{\mathbf{d}}^{\nu} h(\mathbf{d}(-\ln \|x\|_{\mathbf{d}}) x), \end{aligned} \quad \forall x \neq \mathbf{0}.$$

Since, by Rademacher Theorem, the locally Lipschitz continuous function h is differentiable almost everywhere on $\mathbb{R}^n \setminus \{\mathbf{0}\}$ then for almost all $x \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ using product and chain rules for derivatives we obtain

$$\begin{aligned} \frac{\partial h(x)}{\partial x} &\stackrel{a.e.}{=} \nu \|x\|_{\mathbf{d}}^{\nu-1} h(\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x) \frac{\partial \|x\|_{\mathbf{d}}}{\partial x} \\ &+ \|x\|_{\mathbf{d}}^{\nu} \frac{\partial h(z)}{\partial z} \frac{\partial}{\partial x} (\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x), \end{aligned}$$

where $z = \mathbf{d}(-\ln \|x\|_{\mathbf{d}})x$. Taking into account the identities $\frac{d}{ds} \mathbf{d}(s) = G_{\mathbf{d}} \mathbf{d}(s) = \mathbf{d}(s) G_{\mathbf{d}}, \forall s \in \mathbb{R}$ we conclude

$$\frac{\partial (\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x)}{\partial x} = \mathbf{d}(-\ln \|x\|_{\mathbf{d}}) - \frac{G_{\mathbf{d}} \mathbf{d}(-\ln \|x\|_{\mathbf{d}})x}{\|x\|_{\mathbf{d}}} \frac{\partial \|x\|_{\mathbf{d}}}{\partial x}.$$

Finally, since $\frac{\partial \|x\|_{\mathbf{d}}}{\partial x} G_{\mathbf{d}} x = \|x\|_{\mathbf{d}}, \forall x \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ due to the formula (7), then $\frac{\partial (\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x)}{\partial x} G_{\mathbf{d}} x = \mathbf{0}$ and the identity (10) holds.

2) \Rightarrow 1). For an arbitrary $x \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ let us consider the function $s \mapsto h(\mathbf{d}(s)x)$. Since $\mathbf{d}(s)x \neq \mathbf{0}$ for all $s \in \mathbb{R}$ then the considered function is locally Lipschitz continuous on \mathbb{R} and for almost all $x \neq \mathbf{0}$ we have

$$\left. \frac{dh(\mathbf{d}(s)x)}{ds} \stackrel{a.e.}{=} \frac{\partial h(z)}{\partial z} \right|_{z=\mathbf{d}(s)x} \left. \frac{d\mathbf{d}(s)x}{ds} \stackrel{a.e.}{=} \frac{\partial h(z)}{\partial z} G_{\mathbf{d}} z \right|_{z=\mathbf{d}(s)x}$$

due to $\frac{d}{ds} \mathbf{d}(s) = G_{\mathbf{d}} \mathbf{d}(s), \forall s \in \mathbb{R}$. Using the identity (10) we derive

$$\frac{dh(\mathbf{d}(s)x)}{ds} \stackrel{a.e.}{=} \nu h(\mathbf{d}(s)x).$$

Hence, taking into account $h(\mathbf{d}(0)x) = h(x)$ we derive

$$h(\mathbf{d}(s)x) = e^{\nu s} h(x), \forall s \in \mathbb{R}, \quad \forall x \neq \mathbf{0}.$$

Since by assumption $h(\mathbf{0}) = 0$ then $h \in \mathcal{H}_{\mathbf{d}}(\mathbb{R}^n)$.

2) \Leftrightarrow 3) Since we have $\frac{\partial h}{\partial x} G_{\mathbf{d}} x \stackrel{a.e.}{=} \sum_{i=1}^n \frac{\partial h}{\partial x_i} e_i^{\top} G_{\mathbf{d}} x = h(x)$ then, taking into account local Lipschitz continuity of h on $\mathbb{R}^n \setminus \{\mathbf{0}\}$ we conclude that the identity (10) is equivalent

$$\sum_{i=1}^n \int_{x_1^*}^{x_1} \dots \int_{x_n^*}^{x_n} \frac{\partial h}{\partial y_i} e_i^{\top} G_{\mathbf{d}} y \, dy_1 \dots dy_n = \nu \int_{x^*}^x h(y) \, dy$$

for all $x, x^* \in \mathbb{R}^n$ such that $R(x, x^*) \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. Changing the integration order

$$\int_{x_1^*}^{x_1} \dots \int_{x_n^*}^{x_n} \dots dy = \int_{x_1^*}^{x_1} \dots \int_{x_{i-1}^*}^{x_{i-1}} \int_{x_{i+1}^*}^{x_{i+1}} \dots \int_{x_n^*}^{x_n} \int_{x_i^*}^{x_i} \dots dy_i dy_n dy_{n-1} \dots$$

and using the integration by parts

$$\int_{x_i^*}^{x_i} \frac{\partial h}{\partial y_i} e_i^{\top} G_{\mathbf{d}} y \, dy_i = h(y) e_i^{\top} G_{\mathbf{d}} y \Big|_{x_i^*}^{x_i} - \int_{x_i^*}^{x_i} h(y) \frac{\partial e_i^{\top} G_{\mathbf{d}} y}{\partial y_i} \, dy_i$$

we derive

$$\begin{aligned} &\int_{x_1^*}^{x_1} \dots \int_{x_n^*}^{x_n} \frac{\partial h}{\partial y_i} e_i^{\top} G_{\mathbf{d}} y \, dy = \\ &\int_{x_1^*}^{x_1} \dots \int_{x_{i-1}^*}^{x_{i-1}} \int_{x_{i+1}^*}^{x_{i+1}} \dots \int_{x_n^*}^{x_n} h(y) e_i^{\top} G_{\mathbf{d}} y \Big|_{x_i^*}^{x_i} \, dy_n \dots dy_{i+1} dy_{i-1} \dots dy_1 - \\ &\int_{x_1^*}^{x_1} \dots \int_{x_{i-1}^*}^{x_{i-1}} \int_{x_{i+1}^*}^{x_{i+1}} \dots \int_{x_n^*}^{x_n} \int_{x_i^*}^{x_i} h(y) e_i^{\top} G_{\mathbf{d}} e_i \, dy_i dy_n \dots dy_{i+1} dy_{i-1} \dots dy_1. \end{aligned}$$

Hence, using the identity $\frac{1}{x_i - x_i^*} \int_{x_i^*}^{x_i} 1 \, dy_i = 1$ we derive the formula (11). \blacksquare

The identity (11) is utilized below for identification of the generator $G_{\mathbf{d}}$ of a dilation group \mathbf{d} by means of ANNs.

3 Homogeneous ANNs

The statements of universal approximation theorems [7], [13], [12] vary dependently of the restriction to an activation function $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ of ANN. To the best of author's knowledge, the weakest restriction to σ is given by the following theorem.

Theorem 2 (Universal approximation theorem, [20]) *Let $\sigma \in C(\mathbb{R}, \mathbb{R})$ be a non-polynomial function and $\Omega \subset \mathbb{R}^n$ be a compact set. If $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuous function then for any $\epsilon > 0$ there exists $N \in \mathbb{N}, A \in \mathbb{R}^{N \times n}, b \in \mathbb{R}^N$ and $C \in \mathbb{R}^{1 \times N}$ such that*

$$|g(x) - g_{\epsilon}(x)| \leq \epsilon, \quad \forall x \in \Omega, \quad (12)$$

$$g_{\epsilon}(x) = C\sigma(Ax + b), \quad (13)$$

where the function σ in the above formula is applied in the component-wise manner.

The function g_{ϵ} given by (13) is an ANN with one hidden layer (also known as Shallow ANN). The above theorem shows that the considered ANN is the universal approximator of any continuous mapping on a compact. For simplicity, all theoretical results are proven below for scalar-valued functions $\mathbb{R}^n \rightarrow \mathbb{R}$. The extension to the vector-valued functions $\mathbb{R}^n \rightarrow \mathbb{R}^m$ is straightforward.

3.1 Homogeneous universal approximation theorem

Let us consider, first, an unknown \mathbf{d} -homogeneous mappings having a known homogeneity degree. We introduce the \mathbf{d} -homogeneous artificial neural network being a global universal approximator of a continuous \mathbf{d} -homogeneous function.

Definition 3 An ANN g_ϵ is said to be \mathbf{d} -homogeneous if g_ϵ is a \mathbf{d} -homogeneous function of a degree $\mu \in \mathbb{R}$.

In the case of homogeneous function, the universal approximation theorem can be revisited as follows.

Theorem 3 Let \mathbf{d} be a linear continuous dilation in \mathbb{R}^n . Let $\|\cdot\|_{\mathbf{d}}$ be an arbitrary (explicit or implicit) \mathbf{d} -homogeneous norm in \mathbb{R}^n . Let $\sigma \in C(\mathbb{R}, \mathbb{R})$ be a non-polynomial function.

If $h \in C(\mathbb{R}^n \setminus \{\mathbf{0}\}, \mathbb{R})$ is a \mathbf{d} -homogeneous function of degree $\nu \in \mathbb{R}$ then for any $\epsilon > 0$ there exist $N \in \mathbb{N}$, $A \in \mathbb{R}^{N \times n}$, $b \in \mathbb{R}^N$ and $C \in \mathbb{R}^{1 \times N}$ such that

$$|h(x) - h_\epsilon(x)| \leq \epsilon \|x\|_{\mathbf{d}}^\nu, \quad \forall x \in \mathbb{R}^n \setminus \{\mathbf{0}\}, \quad (14)$$

$$h_\epsilon(x) = \|x\|_{\mathbf{d}}^\nu C \sigma(\mathbf{A} \mathbf{d}(-\ln \|x\|_{\mathbf{d}})x + b), \quad (15)$$

where the function σ in the above formula is applied in the component-wise manner.

Proof. By definition, the \mathbf{d} -homogeneous norm $\|\cdot\|_{\mathbf{d}}$ is a continuous positive definite function satisfying $\|\mathbf{d}(s)x\|_{\mathbf{d}} = e^s \|x\|_{\mathbf{d}}$. Then $\|\cdot\|_{\mathbf{d}}$ is radially unbounded and the set

$$S_{\mathbf{d}} = \{x \in \mathbb{R}^n : \|x\|_{\mathbf{d}} = 1\}$$

is a compact such that $\mathbf{0} \notin S_{\mathbf{d}}$. Since $\|\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x\|_{\mathbf{d}} = e^{-\ln \|x\|_{\mathbf{d}}} \|x\|_{\mathbf{d}} = 1$ for any $x \neq \mathbf{0}$ then $\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x \in S_{\mathbf{d}}$. Let $\|\cdot\|$ be a norm in \mathbb{R}^n and

$$r_1 := \inf_{x \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \|\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x\| > 0,$$

$$r_2 := \sup_{x \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \|\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x\| \in [r_1, +\infty).$$

Let us consider the set

$$\Omega = \{y \in \mathbb{R}^n : r_1 \leq \|y\| \leq r_2\}.$$

By construction, the set Ω is a compact and $\mathbf{0} \notin \Omega$. Since the function h is continuous on Ω , then, by Theorem 2, for any $\epsilon > 0$ there exists $N \in \mathbb{N}$, $A \in \mathbb{R}^{N \times n}$, $b \in \mathbb{R}^N$ and $C \in \mathbb{R}^{1 \times N}$ such that

$$\sup_{y \in \Omega} |h(y) - C \sigma(\mathbf{A}y + b)| \leq \epsilon.$$

Since, due to \mathbf{d} -homogeneity of h , we have

$$h(x) = \|x\|_{\mathbf{d}}^\nu \underbrace{h(\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x)}_{=y \in \Omega}, \quad \forall x \neq \mathbf{0},$$

then the estimate (14) holds for h_ϵ given by (15). ■

The key feature of the homogeneous ANN is a global approximation based on local measurements. Indeed, by the formula (15) the homogeneous ANN uses the conventional ANN $C \sigma(\mathbf{A} \pi_{\mathbf{d}} + b)$ with the input $\pi_{\mathbf{d}} = \mathbf{d}(-\ln \|x\|_{\mathbf{d}})x$ being compactly supported on the unit sphere $\{\pi \in \mathbb{R}^n : \|\pi\|_{\mathbf{d}} = 1\}$ in \mathbb{R}^n . The classical universal approximation theorem (see Theorem 2) guarantees the error estimate (12) only on this compact. Due to homogeneity, any local estimate of the homogeneous function can be extended globally (see (14)). The approximation error in this case depends of the norm of the vector x and the homogeneity degree ν of the function h . If $\nu > 0$ then, due to Proposition 1, h is continuous at $\mathbf{0}$ and the above error estimate holds at zero as well. For $\nu < 0$ the function h is discontinuous and unbounded at $\mathbf{0}$. In this case, the approximation error may grow to infinity as $x \rightarrow \mathbf{0}$. For $\nu = 0$ the error is estimated uniformly on \mathbb{R}^n .

Usually algorithms of ANN learning optimize the parameters A, b, C based on available measurements

$$y_j = h(x_j), \quad j = 1, \dots, M,$$

where M is a number of experiments. Mathematically, the goal of the ANN learning is to minimize a norm of the residual

$$e_j = y_j - \|x_j\|_{\mathbf{d}}^\nu C \sigma(\mathbf{A} \mathbf{d}(-\ln \|x_j\|_{\mathbf{d}})x_j + b). \quad (16)$$

It is worth stressing that the homogeneous ANN has structurally the same dependence on the parameters A, b, C . So, all conventional learning algorithms are applicable for its training. Their study goes out of the scope of this paper since, due to the similarity in structures of ANN and homogeneous ANN, all main advantages and disadvantages of the corresponding methods will be the same in our case. The aim of this paper is to demonstrate that the use of dilation symmetry in ANN design may improve the approximation precision of ANN provided that the mapping, which needs to be approximated, is homogeneous.

同质人工神经网络

Figure 1: The characters for recognition

Example 1 The scaling invariance is desirable feature for a pattern recognition by ANN. There are several approaches to exhibit the required invariance (see, e.g., [5, page 262]). One of them is to build the invariance properties into the structure of a neural network. The present example uses the generalized homogeneity for this purpose.

For simplicity, we assume that we deal with gray-scale images. Following the conventional approach, we associated any image with a function $(z_1, z_2) \mapsto \phi(z)$ of two arguments, which value $\phi(z) \in [0, 255]$ define a tone of the gray color at the coordinates $z = (z_1, z_2)^\top$. The value $\phi(z) = 0$ corresponds to the white color, while the value $\phi(z) = 255$ means that the point with the coordinate z has the black color. A pattern depicted in the image is assumed to be centered. The origin $z = (0, 0)$ of the coordinate frame corresponds to the center of the image. In computer, the coordinates of the points z as well as the possible values of $\phi(z)$ are represented by integers, i.e., $z \in \mathbb{Z}^2, \phi(z) \in \{0, 1, 2, \dots, 255\}$. We extend the domain of ϕ to \mathbb{R}^2 and the range of ϕ to the interval $[0, 255]$, respectively. This extension of the function ϕ can always be made in a continuous manner. For the points z out of the image range we assign $\phi(z) = 0$. In this case, the size of image does not impact to the definition of $\phi \in C_c(\mathbb{R}^2, \mathbb{R})$.

Traditionally, the values of $\phi(z)$ are utilized as inputs to ANN. Since the function ϕ is not homogeneous even with respect to the standard dilation $z \mapsto e^s z$, then the efficiency of a homogeneous ANN is questionable in this case. To homogenize the problem, we introduce the following transformations

$$x_i(\phi) = \int_{\mathbb{R}^2} b_i(z) \phi(z) dz, \quad i = 1, 2, \dots, n \quad (17)$$

where $b_i : \mathbb{R}^2 \mapsto \mathbb{R}$ are standard homogeneous functions of positive degree $r_i > -2$, i.e., $b_i(e^s z) = e^{r_i s} b_i(z), \forall s \in \mathbb{R}, \forall z \in \mathbb{R}^2$ and $n \in \mathbb{N}$. For example, we may take $b_i(z) = \frac{z_1^{p_i} z_2^{q_i}}{r_i!}$, where $m_i \in \{1, 2\}$ and $p_i, q_i \in \mathbb{N} \cup \{0\}$, $r_i = p_i + q_i$. In this case, the functions b_i are elements of polynomial basis in $L^2(\Omega, \mathbb{R})$, where $\Omega \subset \mathbb{R}^2$ contains a support of ϕ . So, any function $\phi \in C_c(\mathbb{R}^2, \mathbb{R}) \cap L^2(\Omega, \mathbb{R})$ can be approximated as linear combination of the polynomial basis functions b_i with the coefficients dependent on the vector $x = (x_1^\top, \dots, x_n^\top)^\top \in \mathbb{R}^{2n}$ related with coordinates of ϕ in the polynomial basis. Since the function ϕ is compactly supported with the support dependent on the image size, the computation of x_i is not a difficult problem and it can be easily parallelized.

Notice that the zoom of the image corresponds to the coordinate transformation $z \mapsto e^s z$ with $s \in \mathbb{R}$, which implies the transformation $\phi \mapsto \phi_s$, where $\phi_s(z) := \phi(e^{-s} z)$. Hence, we derive

$$\begin{aligned} x_i(\phi_s) &= \int_{\mathbb{R}^2} b_i(z) \phi(e^{-s} z) dz \\ &= e^{2s} \int_{\mathbb{R}^2} e^{i s} b_i(e^{-s} z) \phi(e^{-s} z) d(e^{-s} z) \\ &= e^{(2+r_i)s} x_i(\phi) \end{aligned}$$

In other words, the dilation $z \mapsto e^s z$ (the zoom of the image) implies the dilation of the coordinates $x_i \mapsto e^{(2+r_i)s} x_i, i = 0, \dots, n$. The vector $x = (x_1, \dots, x_n)^\top$ can be utilized as an input to an ANN, which is going to be designed \mathbf{d} -homogeneous of degree 0 with respect to the dilation

$$\mathbf{d}(s) = \text{diag}(e^{(2+r_1)s}, \dots, e^{(2+r_n)s}), \quad s \in \mathbb{R}.$$

In this case, the scaling/zoom of the image will not impact the output of the \mathbf{d} -homogeneous ANN. Therefore, once being trained, such ANN can be utilized for the pattern recognition independently of the scaling of the function $\phi \in C_c(\mathbb{R}^2, \mathbb{R})$. However, scaling of the image maps a function ϕ having a discrete domain \mathbb{Z}^2 . In this case a reduction of the image size may lead to a degradation of the image quality and impossibility of its recognition. So, the \mathbf{d} -homogeneous ANN will have the natural limitation in recognition of very reduced images.

For numerical illustration, we consider the 8 images of the characters shown on Figure 1. Each image has the size 100×100 pixels. The components of the input vector x are defined by the formula (17) with $n = 8$ and $b_1 = 1, b_2 = z_1, b_3 = z_2, b_4 = 0.5z_1z_2, b_5 = 0.5x^2, b_6 = 0.5y^2, b_7 = x^2y/6, b_8 = xy^2/6$. To demonstrate the computational robustness of the method, the coordinates x_i are computed using the following approximation of the integral (17):

$$x_i(\phi) \approx \sum_{k=1}^{100} \sum_{j=1}^{100} b_i(k-50, j-50) \phi(k, j).$$

The ANN is assumed to have $k = 8$ outputs. For the i -th character at the input, only the i -th output must have the value 1. All other outputs have to be zero in this case. The training set consist of nominal images

of the size 100×100 pixels and the scaled images the size 75×75 pixels, 50×50 pixels and 35×35 pixels, respectively. The total number of images in the training set is 32. The matrices $A \in \mathbb{R}^{N \times n}$ and $B \in \mathbb{R}^{N \times 1}$ are selected randomly with $N = 8$. The matrix $C \in \mathbb{R}^{k \times N}$ is obtained by means of least square optimization on the training set. The trained homogeneous ANN has been tested on images, which are scaled (zoomed) to $30\% + i10\%$ with $i = 0, 1, \dots, 17$. Notice that the decrease of the size of the nominal image implies a degradation of the image quality (see Figure 2). 17 scaled images have been successfully recognized.



Figure 2: Illustration of the degradation of the image quality after the zoom 30%

To test the robustness with respect to the noise, a normally distributed noise has been added to the nominal images. About 50% of pixels have been corrupted by the noise (see Figure 3). The homogeneous ANN successfully recognized the noisy images as well.



Figure 3: The characters corrupted by the noise

The construed homogeneous ANN have good capabilities of pattern recognition with relatively small set of parameters to be tuned (just 8×8 matrix C of real numbers).

3.2 Homogenization of ANN with known homogeneity parameters

By the universal approximation theorem, the ANN is a local approximator of a function, but the homogeneous ANN is a global approximator of a homogeneous function. The transformation of the ANN to a homogeneous ANN can be interpreted as global homogeneous extrapolation of the ANN. In this subsection, we consider the simplest case, when the homogeneity parameters ($G_{\mathbf{d}}$ and ν) are assumed to be known. In the next section, the problem of homogeneous extrapolation is studied under parametric uncertainty.

We assume that a conventional ANN (13) is already well trained such that the approximation error (12) is somehow minimized. The linear dilation and the homogeneity degree are assumed to be known. We need to transform the classical ANN (13) to the homogeneous ANN (15).

Proposition 2 Let $\Omega \subset \mathbb{R}^n$ be a compact set. Let g_ϵ be ANN (13) approximating a function $h \in C(\mathbb{R}^n \setminus \{\mathbf{0}\}, \mathbb{R})$ on the compact $\Omega \subset \mathbb{R}^n$ with an error $\epsilon > 0$, i.e., the inequality (12) is fulfilled for $g = h$. Let the function h be \mathbf{d} -homogeneous of degree $\nu \in \mathbb{R}$. If there exists a norm $\|\cdot\|$ in \mathbb{R}^n such that

$$S = \{x \in \mathbb{R}^n : \|x\| = 1\} \subset \text{int } \Omega \quad (18)$$

and the dilation \mathbf{d} be monotone with respect to $\|\cdot\|$, then the homogeneous ANN (15) with parameters A, b, C taken from the conventional ANN (13) satisfies the error estimate (14), where $\|\cdot\|_{\mathbf{d}}$ is the canonical homogeneous norm induced by $\|\cdot\|$.

Proof. By construction the homogeneous ANN h_ϵ coincides with the original ANN h_ϵ on the unit sphere, i.e.,

$$h_\epsilon(y) = g_\epsilon(y), \quad \forall y \in S.$$

This means that

$$|g(y) - g_\epsilon(y)| = |h(y) - h_\epsilon(y)| \leq \epsilon, \quad \forall y \in S.$$

Since $\|\cdot\|_{\mathbf{d}}$ is a canonical homogeneous norm induced by the norm $\|\cdot\|$ then $y = \mathbf{d}(-\ln \|x\|_{\mathbf{d}})x \in S$ for all $x \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and using the homogeneity we derive

$$\frac{|h(x) - h_\epsilon(x)|}{\|x\|_{\mathbf{d}}^\nu} \leq |h(\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x) - h_\epsilon(\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x)| \leq \epsilon.$$

The case of homogeneous vector field can be analyzed similarly. ■

The above proposition shows that the conventional ANN can be straightforwardly upgraded to the homogeneous ANN provided that the information about the dilation symmetry of the function h is available (e.g., from physics). The idea of the transformation/upgrade is very simple: we just assign a unit sphere $S \subset \Omega$ and expand the values of the g_ϵ from the unit sphere to the whole space \mathbb{R}^n by means of the dilation.

Domain	ANN	hANN $\nu = 2$	hANN $\nu = \nu_\epsilon$	hANN $\nu = 1, G_{\mathbf{d}} = G_\epsilon$
Ω_{-4}	0.0864	0.0010	0.0134	0.0284
Ω_{-3}	0.0821	0.0024	0.0168	0.0286
Ω_{-2}	0.0583	0.0038	0.0170	0.0213
Ω_{-1}	0.0274	0.0063	0.0135	0.0122
$\Omega_0 = \Omega$	0.0110	0.0074	0.0082	0.0094
Ω_1	0.0612	0.0090	0.0193	0.0293
Ω_2	0.2257	0.0115	0.0417	0.0787
Ω_3	0.5056	0.0153	0.0773	0.1551
Ω_4	1.0409	0.0192	0.1160	0.3097

Table 1: Approximation precision of the conventional and homogeneous ANN

Notice that, with such a homogeneous extrapolation, the local error estimate (12) (being uniform on Ω) is converted to the global $\|x\|_{\mathbf{d}}$ -dependent estimate (14) on \mathbb{R}^n . The similar upgrade can be proposed for a function $h : \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$ defined on a \mathbf{d} -homogeneous cone² \mathcal{D} . The condition (18) becomes $S \cap \mathcal{D} \subset \Omega \cap \mathcal{D}$ in this case.

Example 2 Let us consider now the rotation dynamics (1) of the rigid body with $J = \begin{pmatrix} 1 & 0.3 & 0.1 \\ 0.3 & 1.2 & 0.2 \\ 0.1 & 0.2 & 0.8 \end{pmatrix}$. First, we approximate the right-hand side of (1)

$$h(x) = -J^{-1}(\omega \times J\omega + \tau), \quad x = (\omega^\top, \tau^\top)^\top \in \mathbb{R}^6$$

by means of the conventional ANN with the sigmoid activation function

$$g_\epsilon(x) = C\sigma(Ax + b),$$

where $A \in \mathbb{R}^{N \times 6}$, $b \in \mathbb{R}^N$ are randomly selected for $N = 500$ and $C \in \mathbb{R}^{3 \times N}$ is obtained by means of the least square error minimization for $M = 20000$ measurements $y_i = h(x_i)$ of the function h at the points $x_i \in \mathbb{R}^6$ randomly uniformly distributed in the set

$$\Omega = \{x \in \mathbb{R}^6 : 0.95 \leq |x| \leq 1.05\}.$$

On Ω the obtained ANN has the following approximation precision:

$$|g_\epsilon(x) - h(x)| \leq \epsilon \approx 0.0110, \quad x \in \Omega,$$

which is evaluated by taking 20000 random points in Ω .

By Proposition 2, the conventional ANN g_ϵ can be upgraded to a homogeneous ANN h_ϵ by the formula (15), where $G_{\mathbf{d}} = \text{diag}(1, 1, 1, 2, 2, 2)$, $\nu = 2$ and the canonical homogeneous norm $\|\cdot\|_{\mathbf{d}}$ is induced by the usual Euclidean norm $|\cdot|$. Such an upgrade slightly improve the approximation precision (L^∞ -norm of the error) on the set Ω . Moreover, due to homogeneity of h , the approximation precision of the homogeneous ANN (hANN), on the domains $\{x \in \mathbb{R}^6 : |x| \leq 0.95\}$ and $\{x \in \mathbb{R}^6 : |x| \geq 1.05\}$ is better (in times) than the precision of the classical ANN. Table 1 presents the comparison of the approximation precision for the domains

$$\begin{aligned} \Omega_{-4} &= \{x \in \mathbb{R}^6 : 0 < |x| \leq 0.25\}, & \Omega_4 &= \{x \in \mathbb{R}^6 : 1.75 < |x| \leq 2\}, \\ \Omega_{-3} &= \{x \in \mathbb{R}^6 : 0.25 < |x| \leq 0.5\}, & \Omega_3 &= \{x \in \mathbb{R}^6 : 1.5 < |x| \leq 1.75\}, \\ \Omega_{-2} &= \{x \in \mathbb{R}^6 : 0.5 < |x| \leq 0.75\}, & \Omega_2 &= \{x \in \mathbb{R}^6 : 1.25 < |x| \leq 1.5\}, \\ \Omega_{-1} &= \{x \in \mathbb{R}^6 : 0.75 < |x| \leq 1\}, & \Omega_1 &= \{x \in \mathbb{R}^6 : 1 < |x| \leq 1.25\}. \end{aligned}$$

The table contains the estimates of the norms $\|g_\epsilon(x) - h(x)\|_{L^\infty(\Omega_k, \mathbb{R}^6)}$ calculated using 20000 randomly selected points in Ω_k , $k = -4, \dots, 4$. Below we show that the dilation symmetry and/or the homogeneity degree ν of the function can be first identified using some already well-trained conventional ANN, but, next, this ANN can be upgraded to a homogeneous one. The results of such a data-driven upgrade are also presented in Table 1 for comparison.

4 Data-Driven Dilation Symmetry

In this section we continue the study of a possibility of transformation of an existing artificial neural network to a homogeneous one. We assume that an ANN is given by $g_\epsilon : \mathbb{R}^n \mapsto \mathbb{R}^m$ such that

$$g_\epsilon(x) = C\sigma(Ax + b)$$

approximates a function $g \in C(\mathbb{R}^n \setminus \{\mathbf{0}\}, \mathbb{R})$ on a compact $\Omega \subset \mathbb{R}^n \setminus \{\mathbf{0}\}$, where $A \in \mathbb{R}^{N \times n}$, $b \in \mathbb{R}^N$, $C \in \mathbb{R}^{1 \times N}$. The index $\epsilon > 0$ indicates that the approximation precision of ANN (see Theorem 2) is given by

$$\sup_{x \in \Omega} |g(x) - g_\epsilon(x)| \leq \epsilon, \quad \forall x \in \Omega.$$

²A set \mathcal{D} is said to be a \mathbf{d} -homogeneous cone if $\mathbf{d}(s)\mathcal{D} \subset \mathcal{D}$ for all $s \in \mathbb{R}$.

The ANN is supposed to be well trained to approximate a homogeneous function $g = h$ on a compact Ω , i.e., $\epsilon > 0$ is assumed to be small enough. But now we consider the case when *some (or any) information about dilation symmetry of the function h is unavailable*, so, first, the parameters of the dilation group and/or the homogeneity degree have to be identified using g_ϵ , next, a homogeneous ANN has to be designed.

4.1 Practical homogeneity degree

Let us study the problem of identification of an *unknown* homogeneity degree of the function h under the assumption that the linear dilation $\mathbf{d}(s) = e^{sG_{\mathbf{d}}}$, $s \in \mathbb{R}$ is *known*.

By Definition 2, the scalar-valued function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is \mathbf{d} -homogeneous of degree $\nu \in \mathbb{R}$ if

$$h(e^{sG_{\mathbf{d}}}x) = e^{\nu s}h(x), \quad \forall x \in \mathbb{R}^n, \quad \forall s \in \mathbb{R}, \quad (19)$$

where $\mathbf{d}(s) = e^{sG_{\mathbf{d}}}$ is a linear dilation in \mathbb{R}^n . If g_ϵ approximates a \mathbf{d} -homogeneous function $g = h$ on Ω with a high enough precision then we should have

$$g_\epsilon(\mathbf{d}(s)x) \approx e^{\nu s}g_\epsilon(x), \quad x \in \Omega, \quad s \in \mathbb{R} : \mathbf{d}(s)x \in \Omega.$$

Let a norm $\|\cdot\|$ in \mathbb{R}^n and the set Ω be such that $S = \{x \in \mathbb{R}^n : \|x\| = 1\} \subset \Omega$ and the dilation \mathbf{d} be monotone with respect to $\|\cdot\|$. In this case, due to $\mathbf{d}(-\ln\|x\|_{\mathbf{d}})x \in S$, the above approximate identity implies

$$g_\epsilon(\mathbf{d}(-\ln\|x\|_{\mathbf{d}})x) \approx \|x\|_{\mathbf{d}}^{-\nu}g_\epsilon(x), \quad \forall x \in \Omega, \quad (20)$$

where $\|\cdot\|_{\mathbf{d}}$ is the canonical homogeneous norm induced by the norm $\|\cdot\|$. Hence, for $x \neq \mathbf{0}$ we have

$$\nu \approx \frac{1}{\ln\|x\|_{\mathbf{d}}} \ln \frac{g_\epsilon(x)}{g_\epsilon(\mathbf{d}(-\ln\|x\|_{\mathbf{d}})x)} \quad (21)$$

provided that

$$0 < \frac{g_\epsilon(x)}{g_\epsilon(\mathbf{d}(-\ln\|x\|_{\mathbf{d}})x)} < +\infty. \quad (22)$$

Notice that the implication

$$h(x) \neq 0 \quad \Rightarrow \quad 0 < \frac{h(x)}{h(\mathbf{d}(s)x)} < +\infty, \quad \forall s \in \mathbb{R}$$

is necessary for a function h to be homogeneous. Indeed, if $h(x) \neq 0$ and $\frac{h(x)}{h(\mathbf{d}(-\ln\|x\|_{\mathbf{d}})x)} < 0$ then the identity (19) is impossible and h is surely non-homogeneous. Therefore, the condition (22) can be utilized for detection of the non-homogeneity of g_ϵ . To decrease an impact of the approximation error on a decision about non-homogeneity of g_ϵ , we restrict the domain in (22) to the compact set

$$K_\delta = \left\{ x \in \Omega : \min_{\{\|g_\epsilon(x)\|, \|g_\epsilon(\pi_{\mathbf{d}}(x))\| \}} \leq \delta \leq \max_{\{\|g_\epsilon(x)\|, \|g_\epsilon(\pi_{\mathbf{d}}(x))\| \}} \right\}, \quad (23)$$

where $\pi_{\mathbf{d}}(x) = \mathbf{d}(-\ln\|x\|_{\mathbf{d}})x$ is the \mathbf{d} -homogeneous projector of $x \neq \mathbf{0}$ on the unit sphere S and $\delta > 0$ is small enough. In this case, if

$$\frac{g_\epsilon(x)}{g_\epsilon(\pi_{\mathbf{d}}(x))} > 0, \quad \forall x \in K_\delta \quad (24)$$

then the *practical homogeneity degree of the function g_ϵ* can be defined as

$$\nu_\epsilon = \frac{1}{M_\delta} \int_{K_\delta} \frac{1}{\ln\|x\|_{\mathbf{d}}} \ln \frac{g_\epsilon(x)}{g_\epsilon(\pi_{\mathbf{d}}(x))} dx \quad (25)$$

provided that $\epsilon > 0$ is small enough, where $M_\delta = \int_{K_\delta} dx$. If the condition (24) is not fulfilled then g_ϵ is not an approximation of homogeneous function or the approximation precision $\epsilon > 0$ is not enough to detect \mathbf{d} -homogeneity.

Proposition 3 *Let g_ϵ be ANN (13) approximating a non-zero continuous function $h \in C(\mathbb{R}^n \setminus \{\mathbf{0}\}, \mathbb{R})$ on a compact $\Omega \subset \mathbb{R}^n$ with an error $\epsilon > 0$, i.e., the inequality (12) is fulfilled for $g = h$. Let the function h be \mathbf{d} -homogeneous of degree $\nu \in \mathbb{R}$. If there exists a norm $\|\cdot\|$ in \mathbb{R}^n such that*

$$S = \{x \in \mathbb{R}^n : \|x\| = 1\} \subset \text{int } \Omega \quad (26)$$

and the dilation \mathbf{d} is monotone with respect to $\|\cdot\|$, then for sufficiently small $\delta > 0$ it holds

$$\nu_\epsilon \rightarrow \nu \quad \text{as} \quad \epsilon \rightarrow 0 \quad (27)$$

where $\nu_\epsilon \in \mathbb{R}$ is given by (25).

Proof. On the one hand, since h is \mathbf{d} -homogeneous, then, for a sufficiently small $\delta > 0$, the compact set $K_\delta \subset \Omega \setminus \{\mathbf{0}\}$ is non-empty. Since h is \mathbf{d} -homogeneous then $\nu = \frac{1}{\|x\|_{\mathbf{d}}} \ln \frac{h(x)}{h(\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x)}$ for all $x \in K_\delta$ and

$$\nu = \frac{1}{M_\delta} \int_{K_\delta} \frac{1}{\ln \|x\|_{\mathbf{d}}} \ln \frac{h(x)}{h(\pi_{\mathbf{d}}(x))} dx.$$

On the other hand, by definition ν_ϵ , we have

$$\nu_\epsilon = \frac{1}{M_\delta} \int_{K_\delta} \frac{1}{\ln \|x\|_{\mathbf{d}}} \ln \frac{h(x) + o(x, \epsilon)}{h(\pi_{\mathbf{d}}(x)) + o(\pi_{\mathbf{d}}(x), \epsilon)} dx,$$

where $\sup_{y \in \Omega} |o(y, \epsilon)| \leq \epsilon$ (see (12)). Since h is assumed to be nonzero and $\pi_{\mathbf{d}}(x) = \mathbf{d}(-\ln \|x\|_{\mathbf{d}})x \in S$ then $|h(\pi_{\mathbf{d}}(x))| \geq \underline{h} = \inf_{y \in S \cap K_\delta} h(y) > 0$. This implies that

$$\frac{1}{\|x\|_{\mathbf{d}}} \ln \frac{g_\epsilon(x)}{g_\epsilon(\pi_{\mathbf{d}}(x))} \rightarrow \frac{1}{\|x\|_{\mathbf{d}}} \ln \frac{h(x)}{h(\pi_{\mathbf{d}}(x))} \quad \text{as } \epsilon \rightarrow 0 \quad (28)$$

uniformly on any compact from $\Omega \cap K_\delta \setminus \{\mathbf{0}\}$, so $\nu_\epsilon \rightarrow \nu$ as $\epsilon \rightarrow 0$. ■

The practical homogeneity degree is still well-defined by the formula (25) provided that the practical “necessary” condition of homogeneity (24) is fulfilled. The formula (25) defines the practical homogeneity degree as a mean value of some function. This value can also be approximated as follows

$$\nu_\epsilon \approx \frac{1}{M} \sum_{i=1}^M \frac{1}{\ln \|x_i\|_{\mathbf{d}}} \ln \frac{g_\epsilon(x_i)}{g_\epsilon(\pi_{\mathbf{d}}(x_i))}, \quad (29)$$

where the points $x_i \in K_\delta \cap \Omega$ are assumed to be uniformly distributed in $K_\delta \cap \Omega$ and $M \in \mathbb{N}$ is a large enough number.

Example 3 Let us consider the function h studied in Example 2 and assume that the homogeneity degree ν of h is unknown, but the dilation $G_{\mathbf{d}}$ is known. By Proposition 3, we can identify a practical homogeneity ν_ϵ using the conventional ANN, which is already trained on the set Ω . We compute ν_ϵ by the formula (29) for randomly selected points $x_i \in \Omega$, where $i = 1, 2, \dots, M = 2000$. This gives the practical estimate of the homogeneity degree $\nu_\epsilon = 1.9647$ being rather close to the real value $\nu = 2$. The approximation precision of the hANN with the practical homogeneity degree is given in Table 1. The use of homogeneous ANN (even with the practical homogeneity degree) essentially improves the extrapolation capabilities of the ANN.

4.2 Practical linear dilation

Let us study the identification problem of a dilation symmetry (if any) for the ANN g_ϵ which approximates a continuous function g . It is easy to see that if a function g is \mathbf{d} -homogeneous of degree ν (i.e., $g = h$) then, for any $\gamma > 0$ this function is $\tilde{\mathbf{d}}$ -homogeneous of the degree $\tilde{\nu} = \gamma\nu$ with respect to the dilation $\tilde{\mathbf{d}}(s) = \mathbf{d}(\gamma s)$, $s \in \mathbb{R}$. So, the identification of the dilation symmetry of g_ϵ can be reduced to identification of the generator $G_{\mathbf{d}} \in \mathbb{R}^{n \times n}$ only, while the homogeneity degree can always be assigned to one of the following constants $\{-1, 0, 1\}$. In other words, if the sign of the homogeneity degree is unknown, then the three identification problems with $\nu = -1$, $\nu = 0$ and $\nu = 1$ has to be solved in order to identify a dilation symmetry of g_ϵ . Without loss of generality we assume below that the homogeneity degree ν of g_ϵ is known or fixed to one of the following numbers $\{-1, 0, 1\}$. We just need to identify an unknown generator $G_{\mathbf{d}} \in \mathbb{R}^{n \times n}$ of the dilation \mathbf{d} .

By Theorem 1, for any locally Lipschitz continuous (on $\mathbb{R}^n \setminus \{\mathbf{0}\}$) \mathbf{d} -homogeneous function h , we have

$$\int_{x^*}^x (\text{trace}(G_{\mathbf{d}}) + \nu) h(y) - \sum_{i: x_i \neq x_i^*} \Delta_i(h, y, x, x^*, G_{\mathbf{d}}) dy = 0 \quad (30)$$

for all $x, x^* \in \mathbb{R}^n$ such that $R(x, x^*) \in \mathbb{R}^n \setminus \{\mathbf{0}\}$.

If g_ϵ approximates a \mathbf{d} -homogeneous function h of degree ν with a rather high precision ϵ on a compact Ω then the identity (30) implies that

$$\int_{x^*}^x (\text{trace}(G) + \nu) g_\epsilon(y) - \sum_{i: x_i \neq x_i^*} \Delta_i(g_\epsilon, y, x, x^*, G_{\mathbf{d}}) dy \approx 0. \quad (31)$$

The only unknown parameter $G_{\mathbf{d}} \in \mathbb{R}^{n \times n}$ is involved in above approximate identity in the affine manner. Its identification can be based on a quadratic optimization.

Let Ω have a non-empty interior. For $\delta \in (0, 1)$ let $\Omega_\delta \subset \text{int } \Omega$ be such that

$$\Omega_\delta \dot{+} B_\delta \subset \Omega \quad \text{and} \quad \Omega_\delta \cap B_\delta = \emptyset, \quad (32)$$

where $B_\delta = \{x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n : |x_i| \leq \delta, i = 1, \dots, n\}$. Homogeneous functions may have singularity at zero, so the condition $\Omega_\delta \cap B_\delta = \emptyset$ exclude this potential singularity point. Let us introduce

$$G_\epsilon \in \text{argmin}_{G \in \Xi} J_\epsilon(G), \quad (33)$$

where $\Xi \subset \mathbb{R}^{n \times n}$ is a set of admissible $n \times n$ matrices and

$$J_\epsilon(G) = \frac{1}{M_\delta} \int_{\Omega_\delta} \int_{B_\delta} \mathcal{J}_\epsilon^2(x^* + z, x^*, G) dz dx^*, \quad (34)$$

$$\mathcal{J}_\epsilon(x, x^*, G) = \int_{x^*}^x (\text{trace}(G) + \nu) g_\epsilon(y) - \sum_{i: x_i \neq x_i^*} \Delta_i(g_\epsilon, y, x, x^*, G_{\mathbf{d}}) dy,$$

and $M_\delta = \int_{\Omega_\delta} dx \cdot \int_{B_\delta} dz$. If G_ϵ is anti-Hurwitz then, in the view of the following theorem, it can be interpreted as a generator of the *practical dilation in \mathbb{R}^n* .

Theorem 4 *Let $\Xi \subset \mathbb{R}^{n \times n}$ be a class of admissible generators of linear dilations. Let g_ϵ be an ANN (13) approximating a function $h: \mathbb{R}^n \rightarrow \mathbb{R}$ on a compact $\Omega \subset \mathbb{R}^n$ with an error $\epsilon > 0$, i.e., the inequality (12) is fulfilled for $g = h$. If h is locally Lipschitz continuous on $\mathbb{R}^n \setminus \{\mathbf{0}\}$ and \mathbf{d} -homogeneous of degree $\nu \in \mathbb{R}$ with respect to a linear dilation $\mathbf{d}(s) = e^{G_{\mathbf{d}} s}$, $s \in \mathbb{R}^n$ with $G_{\mathbf{d}} \in \Xi$, then*

$$\min_{G \in \Xi} J_\epsilon(G) \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0.$$

Proof. By Theorem 1, we have

$$\int_{x^*}^x (\text{trace}(G_{\mathbf{d}}) + \nu) h(y) - \sum_{i: x_i \neq x_i^*} \Delta_i(h, y, x, x^*, G_{\mathbf{d}}) dy = 0$$

for all $x, x^* \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $R(x, x^*) \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. Since $g_\epsilon(x) \rightarrow h(x)$ as $\epsilon \rightarrow 0$ uniformly on compacts from Ω then

$$\int_{x^*}^x (\text{trace}(G_{\mathbf{d}}) + \nu) g_\epsilon(y) - \sum_{i: x_i \neq x_i^*} \Delta_i(g_\epsilon, y, x, x^*, G_{\mathbf{d}}) dy \rightarrow 0$$

as $\epsilon \rightarrow 0$ for all $x, x^* \in \Omega_\delta$. This means that

$$\min_{G \in \Xi} J_\epsilon(G) \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0.$$

The proof is complete \blacksquare

Notice that the above theorem implies that

$$G_{\mathbf{d}} \in \mathcal{P} := \lim_{\epsilon \rightarrow 0} \operatorname{argmin}_{G \in \Xi} J_\epsilon(G).$$

provided that $G_{\mathbf{d}} \in \Xi$. However, the set \mathcal{P} is not a singleton in the general case and the problem of identification of the generator $G_{\mathbf{d}}$ is ill-posed even if $\Xi = \Xi_{\text{in}}$, where

$$\Xi_{\text{in}} = \{G \in \mathbb{R}^{n \times n} : G \text{ is anti-Hurwitz}\}. \quad (35)$$

Indeed, for example, for any $\gamma > 0$, the function $h: \mathbb{R}^2 \rightarrow \mathbb{R}$ given by

$$h(x) = x^\top x, \quad x = (x_1, x_2)^\top \in \mathbb{R}^2 \quad (36)$$

is \mathbf{d}_γ -homogeneous of degree 2 with respect to the linear dilation $\mathbf{d}_\gamma(s) = e^s \begin{pmatrix} \cos(\gamma s) & -\sin(\gamma s) \\ \sin(\gamma s) & \cos(\gamma s) \end{pmatrix}$, $s \in \mathbb{R}$. So, the linear dilation symmetry of the function h given above cannot be uniquely identified.

To reduce the uncertainty in the dilation identification, the cost functional can be modified as follow

$$G_\epsilon = \operatorname{argmin}_{G \in \Xi} J_\epsilon(G) + \xi \text{trace}(G^\top G), \quad (37)$$

where $\Xi = \mathbb{R}^{n \times n}$ and $\xi > 0$ is a small regularization parameter. The second term in the above formula is a strictly convex function (being a norm) in the space of square matrices. The quadratic optimization problem becomes strictly convex with respect to G , so, for any fixed $\epsilon > 0$, the matrix G_ϵ is uniquely defined by (37). Another possible way to reduce the uncertainty is to restrict the set of admissible generators Ξ , for example, to diagonalizable (or simply diagonal) matrices. However, this introduces additional a-priori assumption about the function.

The functional J_ϵ admits the following approximation

$$J_\epsilon(G) \approx \frac{1}{ML} \sum_{k=1}^M \sum_{j=1}^L \mathcal{J}_\epsilon^2(x_k + z_j, x_i, G) dz \quad (38)$$

where the points $x_k \in \Omega_\delta$ are uniformly distributed in Ω_δ , $z_j \in B_\delta$ are uniformly distributed in B_δ and $M, L \in \mathbb{N}$. Finally, denoting $f(x, x^*, y, G) = (\text{trace}(G) + \nu) g_\epsilon(y) - \sum_{i: x_i \neq x_i^*} \Delta_i(g_\epsilon, y, x, x^*, G)$ we derive the

following approximation of the functional \mathcal{J}_ϵ^2 for a sufficiently small $\delta \in (0, 1)$:

$$\mathcal{J}_\epsilon(x_i + z_j, x_i, G) \approx \prod_{i=1}^n e_i^\top z_j \sum_{\lambda \in \{0,1\}^n} f(x_k + z_j, x_k, x_k + \text{diag}(\lambda) z_j, G). \quad (39)$$

The presented discretizations of the functionals J_ϵ and \mathcal{J}_ϵ can be utilized for identification of the generator G_ϵ of the practical dilation, however, in this case, an anti-Hurwitz solution of the optimization problem (33) may exist even if the ANN g_ϵ approximates a non-homogeneous function. So, the practical necessary condition of homogeneity (24) has to be utilized in order to check if the ANN indeed approximates a \mathbf{d}_ϵ -homogeneous (at least on Ω) function. In the view of Proposition 4, a correctness of this conclusion depends on approximation precision of ANN and increases as $\epsilon \rightarrow 0$.

Example 4 Let us consider again Example 2. Now we assume that the dilation is unknown and we need to approximate $G_{\mathbf{d}}$ using the already well-trained ANN g_ϵ . For the data-driven identification of the dilation symmetry, we consider the set $\Omega_\delta = \{x \in \mathbb{R}^n : 0.98 \leq |x| \leq 1.02\}$ and optimize the approximate functional (38) with $M = 4000$, $L = 1$, $z_1 = \delta(1, \dots, 1)^\top$, $\delta = 0.01$. We also restrict the class of admissible generators to diagonal matrices. For $\nu = 1$ the obtained matrix $G_\epsilon = \text{diag}(0.5199, 0.5150, 0.4972, 1.0021, 0.9965, 1.0027)$, which, obviously, corresponds to $G_{\mathbf{d}}$ up to the scaling of G_ϵ and ν by the multiplier $\gamma = 2$ (please see explanations in the beginning of Section 4.2 for more details). The homogeneous ANN constructed in this way is a homogeneous extrapolation of the ANN g_ϵ . In all domain Ω_k except the training set Ω_0 , the approximation precision of hANN is better in times comparing with the original ANN (see Table 1).

Notice that for $\nu = 0$ and $\nu = -1$, the optimization problem does not provide anti-Hurwitz solutions. This implies impossibility of non-positive homogeneity degree of the function.

5 Example: explicit homogeneous stabilization of LTI plant

In this section we present an possible application of homogeneous ANN to the design of the so-called generalized homogeneous control system [10], [4], [1]. Generalized homogeneous control may provide a better regulation quality (such as faster response [35], [23], better robustness [33], [1], and smaller overshoots [28, Chapter 1]) comparing with the linear control.

Let us consider the problem of generalized homogeneous stabilization of LTI plant

$$\dot{x} = \hat{A}x + \hat{B}u + g(t, x), \quad t > 0, \quad x(0) = x_0 \in \mathbb{R}^n, \quad (40)$$

where $x(t) \in \mathbb{R}^n$ is the system state, $u(t) \in \mathbb{R}^m$ is the control input and $\hat{A} \in \mathbb{R}^{n \times n}$, $\hat{B} \in \mathbb{R}^{n \times m}$ is a controllable pair of the system matrices. The unknown vector field $g \in C(\mathbb{R} \times \mathbb{R}^n, \mathbb{R}^n)$ models system uncertainties and disturbances.

In [30], [34] it is shown that, under a proper selection of parameters, the feedback law of the form

$$u = K_0x + \|x\|_{\mathbf{d}}^{\mu+1} K \mathbf{d}(-\ln \|x\|_{\mathbf{d}})x, \quad K_0, K \in \mathbb{R}^{m \times n} \quad (41)$$

stabilizes the system (40) to zero such that the closed-loop system is \mathbf{d} -homogeneous of degree $\mu \in \mathbb{R}$ (in the sense of [16]), where $\mathbf{d}(s) = e^{sG_{\mathbf{d}}}$ is a linear dilation in \mathbb{R}^n and $\|x\|_{\mathbf{d}}$ is the canonical homogeneous norm induced by the weighted Euclidean norm $\|x\| = \sqrt{x^\top P x}$ in \mathbb{R}^n with $0 \prec P = P^\top \in \mathbb{R}^{n \times n}$. The nonlinear feedback law (41) admits the simple tuning rules [28, Chapter 9], [34]:

- 1) solve the linear algebraic equation

$$\hat{A}G_0 - G_0\hat{A} + \hat{B}Y_0 = \hat{A}, \quad G_0\hat{B} = \mathbf{0} \quad (42)$$

with respect to $Y_0 \in \mathbb{R}^{m \times n}$, $G_0 \in \mathbb{R}^{n \times n}$ and define $\mu \in [-1, \frac{1}{n-1}]$, $G_{\mathbf{d}} = I_n + \mu G_0$, $K_0 = Y_0(G_0 - I_n)^{-1}$;

- 2) solve the system of Linear Matrix Inequalities (LMIs)

$$(\hat{A} + \hat{B}K_0)X + (\hat{A} + \hat{B}K_0)^\top X + \hat{B}Y + Y^\top \hat{B}^\top + \rho(G_{\mathbf{d}}X + XG_{\mathbf{d}}^\top) \prec 0, \quad G_{\mathbf{d}}X + XG_{\mathbf{d}}^\top \succ 0, \quad X \succ 0 \quad (43)$$

with respect to $X \in \mathbb{R}^{n \times n}$, $Y \in \mathbb{R}^{m \times n}$ for some given $\rho > 0$ and define $K = YX^{-1}$, $X = P^{-1}$;

In the case of controllable pair $\{\hat{A}, \hat{B}\}$ the linear algebraic systems (42) and (43) are always feasible and can be easily solved in MATLAB. The feedback (41) robustly stabilizes the system (40) rejecting the uncertainties/disturbances, which satisfy the inequality

$$\frac{x^\top \mathbf{d}^\top(-\ln \|x\|_{\mathbf{d}}) X^{-1} \mathbf{d}(-\ln \|x\|_{\mathbf{d}}) g(t, x)}{x^\top \mathbf{d}^\top(-\ln \|x\|_{\mathbf{d}}) X^{-1} G_{\mathbf{d}} \mathbf{d}(-\ln \|x\|_{\mathbf{d}}) x} \leq \kappa \|x\|_{\mathbf{d}}^\mu, \quad \forall x \in \mathbb{R}^n \setminus \{\mathbf{0}\}, \quad \forall t \in \mathbb{R}, \quad (44)$$

for some $\kappa \in (0, \rho)$. However, the feedback (41) has one obvious drawback, which complicates its practical application in multi-dimensional case. Namely, the canonical homogeneous norm $\|x\|_{\mathbf{d}}$ is defined implicitly (see Definition 1) as a unique solution of the nonlinear algebraic equation

$$x^\top \mathbf{d}^\top(-\ln \|x\|_{\mathbf{d}}) P \mathbf{d}(-\ln \|x\|_{\mathbf{d}}) x = 1. \quad (45)$$

To implement the implicit homogeneous control (41) in a real system, the above equation has to be solved numerically in a real time by a digital device (computer) realizing the feedback law (41). In some cases (e.g., for autonomous vehicles) the performance of the on-board computed may be insufficient for practical application of the suggested control law. The use of homogeneous ANN (15) essentially simplifies this task. Indeed, in the view of Theorem 3 the canonical homogeneous norm $\|\cdot\|_{\mathbf{d}}$ can be approximated by the homogeneous ANN

$$\|x\|_{\mathbf{d}, \epsilon} := \|x\|_{\bar{r}} C \sigma(\mathbf{A} \mathbf{d}(-\|x\|_{\bar{r}}) x + b), \quad x \in \mathbb{R}^n \setminus \{\mathbf{0}\} \quad (46)$$

with an arbitrary high precision $\epsilon > 0$ such that

$$\|x\|_{\mathbf{d}} - \|x\|_{\mathbf{d},\epsilon} \leq \epsilon, \quad \forall x \in S = \left\{x \in \mathbb{R}^n : x^\top P x = 1\right\}, \quad (47)$$

where $\|\cdot\|_r$ is any explicit \mathbf{d} -homogeneous norm (e.g., given by the formula (4)), $\sigma : \mathbb{R} \mapsto \mathbb{R}$ is an activation function and the parameters A, b, C are as before. The following theorem shows that the system (40) can be stabilized by a control of the form (41), where the implicit norm $\|\cdot\|_{\mathbf{d}}$ is replaced with its explicit approximation $\|x\|_{\mathbf{d},\epsilon}$.

Theorem 5 *Let the pair $\{\hat{A}, \hat{B}\}$ be controllable and $\mu, G_{\mathbf{d}}, K_0$ be defined as shown above. Given $\epsilon \in (0, 1)$, let the system of linear matrix inequalities*

$$\begin{pmatrix} \tilde{A}X + X\tilde{A}^\top + \hat{B}Y + Y^\top \hat{B}^\top - X & \hat{B}Y & X\mathbf{d}(\ln \frac{1+\epsilon}{1-\epsilon}) \\ Y^\top \hat{B}^\top & -\frac{G_{\mathbf{d}}^{-1}X + XG_{\mathbf{d}}^{-\top}}{\omega} & \mathbf{0} \\ \mathbf{d}(\ln \frac{1+\epsilon}{1-\epsilon})X & \mathbf{0} & -X \end{pmatrix} \prec 0, \quad (48)$$

be feasible with respect to the variable $X \in \mathbb{R}^{n \times n}$, $Y \in \mathbb{R}^{m \times n}$ for some $0 < \delta_{\min} \leq \delta_{\max}$ and $\rho > 0$, where

$$\tilde{A} = \left(1 + (1 + \mu) \ln \left(\frac{1+\epsilon}{1-\epsilon}\right)\right) (\hat{A} + \hat{B}K_0) + \rho \left(\frac{1+\epsilon}{1-\epsilon}\right)^{1+\mu} G_{\mathbf{d}}$$

and

$$\omega = \int_{1-\epsilon}^{1+\epsilon} \sigma^{2\mu+1} d\sigma > 0.$$

Let $\|\cdot\|_{\mathbf{d}}$ be the canonical homogeneous norm induced by the norm $\|x\| = \sqrt{x^\top P x}$ with $P = X^{-1}$. If the homogeneous ANN (46) approximates $\|\cdot\|_{\mathbf{d}}$ with the precision $\epsilon > 0$ on the unit sphere (see the formula (47)) then the closed-loop system (40) with the control

$$u(x) = K_0 x + \|x\|_{\mathbf{d},\epsilon}^{1+\mu} K \mathbf{d}(-\ln \|x\|_{\mathbf{d},\epsilon}) x, \quad K = Y X^{-1} \mathbf{d}(\ln(1+\epsilon)) \quad (49)$$

and with the function g satisfying (44) for some $\kappa \in [0, \rho]$ is globally asymptotically stable. Moreover, the canonical homogeneous norm is a Lyapunov function of the closed-loop system (40), (49):

$$\frac{d}{dt} \|x\|_{\mathbf{d}} \leq -(\rho - \kappa) \|x\|_{\mathbf{d}}^{1+\mu}, \quad \forall x \neq \mathbf{0}. \quad (50)$$

Proof. 1) Let us denote $\delta_{\min} = 1 - \epsilon$, $\delta_{\max} = 1 + \epsilon$ and $A_0 = \hat{A} + \hat{B}K_0$. Notice that $\omega > 0$ and for any solution (X, Y) of the linear matrix inequality (48) we have $G_{\mathbf{d}}^{-1}X + XG_{\mathbf{d}}^{-\top} \succ 0$ or, equivalently, $PG_{\mathbf{d}}^{-1} + G_{\mathbf{d}}^{-\top}P \succ 0$ and $G_{\mathbf{d}}^\top P + PG_{\mathbf{d}} \succ 0$, where $P = X^{-1} \succ 0$. Let us show that

$$2z^\top P \left(A_0 + \delta^{1+\mu} \hat{B}K \mathbf{d}(-\ln \delta)\right) z < -\rho z^\top (PG_{\mathbf{d}} + G_{\mathbf{d}}P) z, \quad (51)$$

$$\forall z \in \mathbb{R}^n \setminus \{\mathbf{0}\}, \quad \forall \delta \in [\delta_{\min}, \delta_{\max}].$$

Let us fix an arbitrary $z \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and consider the smooth function $q : [\delta_{\min}, \delta_{\max}] \rightarrow \mathbb{R}$ defined as

$$q(\delta) = z^\top P \left(A_0 + \delta^{1+\mu} \hat{B}K \mathbf{d}(-\ln \delta)\right) z.$$

To estimate its derivative $q'(\delta)$ we use the inequality $2y^\top v \leq y^\top y + v^\top v$, which holds for all $y, v \in \mathbb{R}^n$ and, in particular, for

$$y = \delta^{0.5+0.5\mu} \Gamma^{-\frac{1}{2}} \mathbf{d}(-\ln \delta_{\max}) K^\top \hat{B}^\top P z$$

and

$$v = \delta^{-0.5-0.5\mu} \Gamma^{\frac{1}{2}} G_{\mathbf{d}} \mathbf{d}\left(-\ln \frac{\delta}{\delta_{\max}}\right) z,$$

where $\Gamma = G_{\mathbf{d}}^{-\top}P + PG_{\mathbf{d}}^{-1}$. Since $\frac{d}{ds} \mathbf{d}(s) = G_{\mathbf{d}} \mathbf{d}(s) = \mathbf{d}(s) G_{\mathbf{d}}$ for $s \in \mathbb{R}$ then

$$\begin{aligned} (1 + \mu) \frac{q(\delta) - z^\top P A_0 z}{\delta} - \frac{dq}{d\delta} &= \delta^\mu z^\top P \hat{B}K G_{\mathbf{d}} \mathbf{d}(-\ln \delta) z = \delta^\mu y^\top v \leq \\ & \frac{\delta^{2\mu+1} z^\top P \hat{B}K \mathbf{d}^\top(\ln \delta_{\max}^{-1}) \Gamma^{-1} \mathbf{d}(\ln \delta_{\max}^{-1}) K^\top \hat{B}^\top P z + \frac{1}{2} z^\top \mathbf{d}^\top(\ln \frac{\delta_{\max}}{\delta}) (PG_{\mathbf{d}} + G_{\mathbf{d}}P) \mathbf{d}(\ln \frac{\delta_{\max}}{\delta}) z}{2} = \\ & \frac{\delta^{2\mu+1} z^\top P \hat{B}K \mathbf{d}^\top(\ln \delta_{\max}^{-1}) \Gamma^{-1} \mathbf{d}(\ln \delta_{\max}^{-1}) K^\top \hat{B}^\top P z - \frac{d}{d\delta} (z^\top \mathbf{d}^\top(\ln \frac{\delta_{\max}}{\delta}) P \mathbf{d}(\ln \frac{\delta_{\max}}{\delta}) z)}{2}. \end{aligned}$$

Using the Newton-Leibniz identity, we derive

$$\begin{aligned} q(\delta) &= q(\delta_{\max}) - \int_{\delta}^{\delta_{\max}} q'(s) ds \leq z^\top P \left(A_0 + \delta_{\max}^{1+\mu} \hat{B}K \mathbf{d}(-\ln \delta_{\max})\right) z \\ &+ (1 + \mu) z^\top P A_0 z \ln \frac{\delta_{\max}}{\delta_{\min}} + \frac{z^\top P \hat{B}K \mathbf{d}^\top(\ln \delta_{\max}^{-1}) \Gamma^{-1} \mathbf{d}(\ln \delta_{\max}^{-1}) K^\top \hat{B}^\top P z}{2} \int_{\delta_{\min}}^{\delta_{\max}} \sigma^{2\mu+1} d\sigma \\ &+ \frac{z^\top \mathbf{d}^\top(\ln \frac{\delta_{\max}}{\delta_{\min}}) P \mathbf{d}(\ln \frac{\delta_{\max}}{\delta_{\min}}) z}{2} - \frac{z^\top P z}{2} - \int_{\delta}^{\delta_{\max}} (1 + \mu) \frac{q(\sigma)}{\sigma} d\sigma \end{aligned}$$

for all $\delta \in [\delta_{\min}, \delta_{\max}]$. Hence, using Grönwall-Bellman inequality we derive

$$q(\delta) \leq \frac{z^\top W z}{2} e^{-(1+\mu) \int_{\delta_{\min}}^{\delta_{\max}} \frac{d\sigma}{\sigma}} = \frac{z^\top W z}{2} \left(\frac{\delta_{\min}}{\delta_{\max}} \right)^{1+\mu}, \quad \forall \delta \in [\delta_{\min}, \delta_{\max}],$$

where

$$\begin{aligned} W = & P \left(A_0 + \delta_{\max}^{1+\mu} \hat{B} K \mathbf{d}(\ln \delta_{\max}^{-1}) \right) + \left(A_0 + \delta_{\max}^{1+\mu} \hat{B} K \mathbf{d}(\ln \delta_{\max}^{-1}) \right)^\top P \\ & + \omega P \hat{B} K \mathbf{d}^\top(\ln \delta_{\max}^{-1}) \Gamma^{-1} \mathbf{d}(\ln \delta_{\max}^{-1}) K^\top \hat{B}^\top P \\ & + \mathbf{d}^\top \left(\ln \frac{\delta_{\max}}{\delta_{\min}} \right) P \mathbf{d} \left(\ln \frac{\delta_{\max}}{\delta_{\min}} \right) - P + (1+\mu) P A_0 \ln \frac{\delta_{\max}}{\delta_{\min}} \end{aligned}$$

Therefore, the matrix inequality

$$W \left(\frac{\delta_{\min}}{\delta_{\max}} \right)^{1+\mu} \prec -\rho(PG_{\mathbf{d}} + G_{\mathbf{d}}P)$$

implies (51). Applying Schur complement³ we derive that the LMI (48) is equivalent to the system of matrix inequalities

$$W + \rho \left(\frac{\delta_{\max}}{\delta_{\min}} \right)^{1+\mu} (PG_{\mathbf{d}} + G_{\mathbf{d}}P) \prec 0, \quad P \succ 0, \quad PG_{\mathbf{d}} + G_{\mathbf{d}}^\top P \succ 0$$

with $X = P^{-1}$ and $Y = K \mathbf{d}(-\ln \delta_{\max}) P^{-1}$.

2) Since $\|\cdot\|_{\mathbf{d},\epsilon}$ and $\|\cdot\|_{\mathbf{d}}$ are \mathbf{d} -homogeneous functions of the degree 1, $y = \mathbf{d}(-\ln \|x\|_{\mathbf{d}})x \in S$ and $\|y\| = 1 \Leftrightarrow \|y\|_{\mathbf{d}} = 1$ then the inequality (47) implies that

$$0 < \delta_{\min} \leq \frac{\|x\|_{\mathbf{d},\epsilon}}{\|x\|_{\mathbf{d}}} = \|\mathbf{d}(-\ln \|x\|_{\mathbf{d}})x\|_{\mathbf{d},\epsilon} < \delta_{\max}, \quad \forall x \neq \mathbf{0}.$$

Let us consider the case $g = \mathbf{0}$. Notice (see, e.g., [34]) that K_0 is selected such that $A_0 \mathbf{d}(s) = e^{\mu s} \mathbf{d}(s) A_0$ and $\mathbf{d}(s) \hat{B} = e^s \hat{B}$. Using the formula (7), let us compute the time derivative of the canonical homogeneous norm $\|\cdot\|_{\mathbf{d}}$ along the trajectories of the system (40),(49)

$$\begin{aligned} \frac{d\|x\|_{\mathbf{d}}}{dt} &= \|x\|_{\mathbf{d}} \frac{x^\top \mathbf{d}^\top(-\ln \|x\|_{\mathbf{d}}) X^{-1} \mathbf{d}(-\ln \|x\|_{\mathbf{d}}) (A_0 x + \|x\|_{\mathbf{d},\epsilon}^{1+\mu} \hat{B} K \mathbf{d}(-\ln \|x\|_{\mathbf{d},\epsilon}) x)}{x^\top \mathbf{d}^\top(-\ln \|x\|_{\mathbf{d}}) X^{-1} G_{\mathbf{d}} \mathbf{d}(-\ln \|x\|_{\mathbf{d}}) x} \\ &= \|x\|_{\mathbf{d}}^{1+\mu} \frac{x^\top \mathbf{d}^\top(-\ln \|x\|_{\mathbf{d}}) X^{-1} \left(A_0 + \left(\frac{\|x\|_{\mathbf{d},\epsilon}}{\|x\|_{\mathbf{d}}} \right)^{1+\mu} \hat{B} K \mathbf{d} \left(-\ln \frac{\|x\|_{\mathbf{d},\epsilon}}{\|x\|_{\mathbf{d}}} \right) \right) \mathbf{d}(-\ln \|x\|_{\mathbf{d}}) x}{x^\top \mathbf{d}^\top(-\ln \|x\|_{\mathbf{d}}) X^{-1} G_{\mathbf{d}} \mathbf{d}(-\ln \|x\|_{\mathbf{d}}) x} \\ &\leq -\rho \|x\|_{\mathbf{d}}^{1+\mu}, \quad \forall x \neq \mathbf{0}, \end{aligned}$$

where the inequality (51) is utilized on the last step.

For $g \neq \mathbf{0}$ the same computations give

$$\frac{d\|x\|_{\mathbf{d}}}{dt} \leq -\rho \|x\|_{\mathbf{d}}^{1+\mu} + \|x\|_{\mathbf{d}} \frac{x^\top \mathbf{d}^\top(-\ln \|x\|_{\mathbf{d}}) X^{-1} \mathbf{d}(-\ln \|x\|_{\mathbf{d}}) g(t,x)}{x^\top \mathbf{d}^\top(-\ln \|x\|_{\mathbf{d}}) X^{-1} G_{\mathbf{d}} \mathbf{d}(-\ln \|x\|_{\mathbf{d}}) x} \leq -(\rho - \kappa) \|x\|_{\mathbf{d}}^{1+\mu}$$

provided that g satisfies the inequality (44). The proof is complete. \blacksquare

Since $\omega \rightarrow 0$ and $\ln \frac{1+\epsilon}{1-\epsilon} \rightarrow 0$ as $\epsilon \rightarrow 0$, then, using Schur complement, we conclude that the LMI (48) in the limit case is equivalent to (43). Therefore, the LMI (48) is feasible at least for a sufficiently small $\epsilon > 0$.

Example 5 As a numerical example, let us consider the controlled planar system

$$\dot{x}_1 = x_2 \quad \dot{x}_2 = u + \gamma(t, x), \quad t > 0, \quad (52)$$

where $x_1(t), x_2(t) \in \mathbb{R}$ are components of the state vector $x = (x_1, x_2)^\top$, $u(t) \in \mathbb{R}$ is the scalar control input and $\gamma : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a continuous unknown function. This system can be represented in the form (40) with $\hat{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\hat{B} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $g = \hat{B} \gamma$. The only solution of (42) is $G_0 = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$ and $Y_0 = \mathbf{0}$. So, $K_0 = \mathbf{0}$ and $G_{\mathbf{d}} = \begin{pmatrix} 1-\mu & 0 \\ 0 & 1 \end{pmatrix}$ with $\mu \in [-1, 1)$. Taking $\mu = -0.25$ we derive the dilation $\mathbf{d}(s) = \text{diag}(e^{5/4s}, e^s)$ in \mathbb{R}^2 . We solve the LMI (48) for $\rho > 1$ and $\epsilon = 0.005$ and we derive

$$K = \begin{pmatrix} -0.8770 & -3.0839 \end{pmatrix}, \quad P = \begin{pmatrix} 0.1157 & 0.0194 \\ 0.0194 & 0.1186 \end{pmatrix}.$$

Inspired by (4), the following explicit homogeneous norm is defined

$$\|x\|_{\tilde{r}} = \left(\Psi^\top(x) Q \Psi(x) \right)^{2/5}, \quad Q = \begin{pmatrix} 0.0720 & 0.0176 \\ 0.0176 & 0.1225 \end{pmatrix},$$

where $x = (x_1, x_2)^\top \in \mathbb{R}^2$, $\Psi(x) = \begin{pmatrix} x_1 \\ |x_2|^{5/4} \text{sign}(x_2) \end{pmatrix}$ and the matrix Q is selected such that the difference between $\|\cdot\|_{\mathbf{d}}$ and $\|\cdot\|_{\tilde{r}}$ is minimized [28, Chapter 9]. On the unit sphere $S = \{x \in \mathbb{R}^2 : \|x\| = 1\}$ the difference between norms is about 0.0255. Such approximation error is inadmissible for application of Theorem 5 with $\epsilon = 0.005$. To improve the approximation precision, we use the homogeneous ANN (46) with the sigmoid activation function σ and randomly selected parameters $A \in \mathbb{R}^{N \times 1}$ and $b \in \mathbb{R}^N$. By definition, the canonical

³ $\begin{pmatrix} X_{11} & X_{12} \\ X_{12}^\top & X_{22} \end{pmatrix} \prec 0 \Leftrightarrow X_{22} \prec 0, X_{11} - X_{12} X_{22}^{-1} X_{12}^\top \prec 0 \Leftrightarrow X_{11} \prec 0, X_{22} - X_{12}^\top X_{11}^{-1} X_{12} \prec 0$

homogeneous norm equals 1 on the unit sphere, i.e., $\|x\|_{\mathbf{d}} = 1$ for all $x \in S$. The vector $C \in \mathbb{R}^{1 \times N}$ is optimized using least square criterion

$$\sum_{i=1}^M (1 - \|x_i\|_{\tilde{r}} C \sigma(\text{Ad}(-\ln \|x_i\|_{\tilde{r}})x + b))^2 \rightarrow \min_{C \in \mathbb{R}^{1 \times N}},$$

where the points x_i are randomly distributed on the unit sphere S . For $N = 13$ such homogeneous ANN (46) improves the approximation precision of $\|x\|_{\mathbf{d}}$ on the unit sphere in ten times in comparison with the norm $\|\cdot\|_{\tilde{r}}$. By Theorem 5, the explicit homogeneous control (49) robustly stabilizes the considered planar system provided that $g = \hat{B}\gamma$ satisfies (44). The simulation results of the closed-loop system with $\gamma = \mathbf{0}$ are depicted on Figure 4. The system is simulated using the explicit Euler method with the sampling period 0.001. The stabilization in the perturbed case $\gamma = \cos^2(10t\|x\|_{\mathbf{d},\epsilon})\|x\|_{\mathbf{d},\epsilon}^{0.75}$ is illustrated on Figure 5.

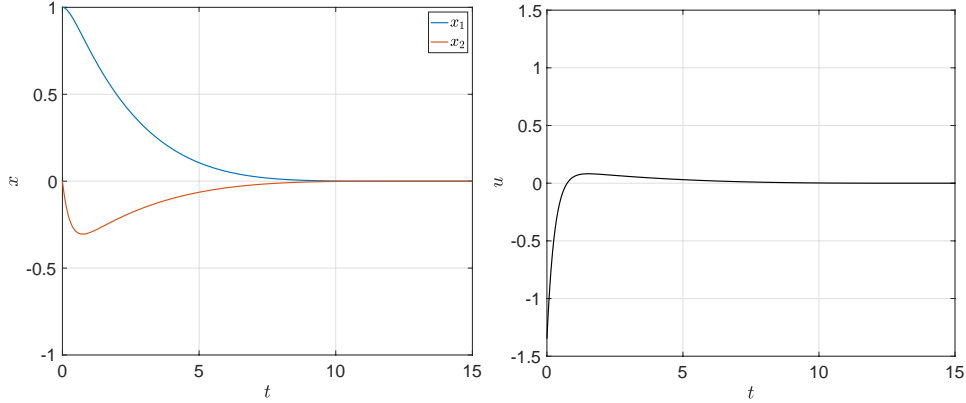


Figure 4: Simulation results of the closed-loop system (52), (49) with $\gamma = \mathbf{0}$

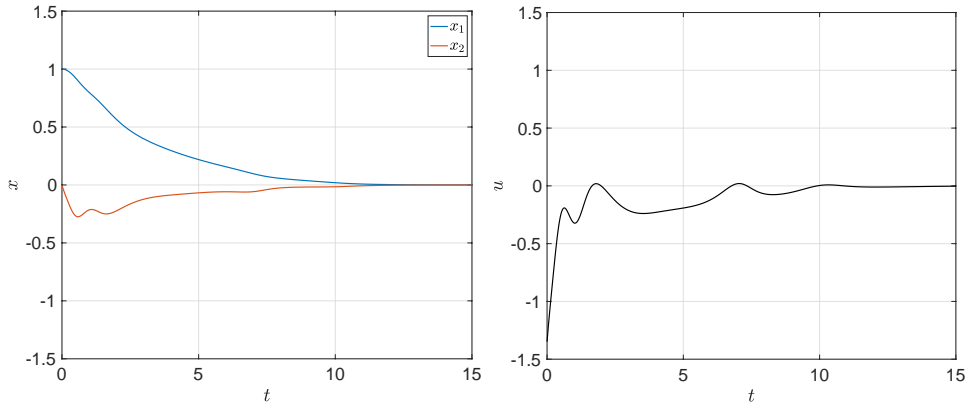


Figure 5: Simulation results of the closed-loop system (52), (49) with $\gamma = \cos^2(10t\|x\|_{\mathbf{d},\epsilon})\|x\|_{\mathbf{d},\epsilon}^{0.75}$

6 Conclusion

The paper introduces a class of ANNs being global approximators of the so-called generalized homogeneous functions. The homogeneous universal approximation theorem is proven. A dilation symmetry of the homogeneous function allows local approximation results to be extended globally.

An information about dilation symmetry of a function approximated by a conventional ANN may be unknown. However, the generalized homogeneous function theorem proven in the paper shows that this information is, in fact, hidden in ANN provided that the ANN is well trained. This important information can be extracted (identified) for the ANN by solving a quadratic optimization problem and, next, the conventional ANN can be easily homogenized using the identified dilation. The obtained homogeneous ANN is a global dilation-symmetry-based extrapolation of the conventional ANN away from the training set.

The potential advantages of homogeneous ANN are demonstrated on academic examples from computer science, systems theory and control engineering. Application of homogeneous ANN to practical problems is an interesting problem of future research. A theoretical study about homogeneous physics-informed ANNs [19] may also be interesting, since many models of mathematical physics are homogeneous in a generalized sense [28].

References

- [1] V. Andrieu, L. Praly, and A. Astolfi. Homogeneous Approximation, Recursive Observer Design, and Output Feedback. *SIAM Journal of Control and Optimization*, 47(4):1814–1850, 2008.
- [2] F. Bach. Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research*, 18:1–53, 2017.
- [3] M. Ballesteros, A. Polyakov, D. Efimov, I. Chairez, and A. Poznyak. Non-parametric identification of homogeneous dynamical systems. *Automatica*, 129(7):109600, 2021.
- [4] S. P. Bhat and D. S. Bernstein. Geometric homogeneity with applications to finite-time stability. *Mathematics of Control, Signals and Systems*, 17:101–127, 2005.
- [5] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [6] J.-M. Coron, L. Grüne, and K. Worthmann. Model predictive control, cost controllability, and homogeneity. *SIAM Journal on Control and Optimization*, 58(5):2979–2996, 2020.
- [7] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 4(2):303–3014, 1989.
- [8] V. Fischer and M. Ruzhansky. *Quantization on Nilpotent Lie Groups*. Springer, 2016.
- [9] G. Folland. Subelliptic estimates and function spaces on nilpotent Lie groups. *Arkiv for Matematik*, 13(1-2):161–207, 1975.
- [10] L. Grüne. Homogeneous state feedback stabilization of homogeneous systems. *SIAM Journal of Control and Optimization*, 38(4):1288–1308, 2000.
- [11] H. Hermes. Nilpotent approximations of control systems and distributions. *SIAM Journal of Control and Optimization*, 24(4):731, 1986.
- [12] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.
- [13] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.
- [14] L.S. Husch. Topological Characterization of The Dilation and The Translation in Frechet Spaces. *Mathematical Annals*, 190:1–5, 1970.
- [15] Z. Ji and M. Telgarsky. Directional convergence and alignment in deep learning. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 2020.
- [16] M. Kawski. Families of dilations and asymptotic stability. *Analysis of Controlled Dynamical Systems*, pages 285–294, 1991.
- [17] V. V. Khomenuk. On systems of ordinary differential equations with generalized homogenous right-hand sides. *Izvestia vuzov. Mathematica (in Russian)*, 3(22):157–164, 1961.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25 (NeurIPS 2012)*, 2012.
- [19] I. E. Lagaris, A. Likas, and D. I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks*, 9(5):987–1000, 1998.
- [20] M. Leshno, V. Lin, A. Pinkus, and S. Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993.
- [21] A. Levant. Higher-order sliding modes, differentiation and output-feedback control. *International Journal of Control*, 76(9-10):924–941, 2003.
- [22] A. M. Lyapunov. *The general problem of the stability of motion*. Taylor & Francis, 1992.
- [23] H. Nakamura, Y. Yamashita, and H. Nishitani. Smooth Lyapunov functions for homogeneous differential inclusions. In *Proceedings of the 41st SICE Annual Conference*, pages 1974–1979, 2002.
- [24] B. Neyshabur, R. Salakhutdinov, and N. Srebro. Path-sgd: Path-normalized optimization in deep neural networks. In *Advances in Neural Information Processing Systems 28 (NeurIPS 2015)*. The MIT Press, 2015.
- [25] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer, 1983.

- [26] W. Perruquetti, T. Floquet, and E. Moulay. Finite-time observers: application to secure communication. *IEEE Transactions on Automatic Control*, 53(1):356–360, 2008.
- [27] A. Polyakov. Sliding mode control design using canonical homogeneous norm. *International Journal of Robust and Nonlinear Control*, 29(3):682–701, 2019.
- [28] A. Polyakov. *Generalized Homogeneity in Systems and Control*. Springer, 2020.
- [29] A. Polyakov, J.-M. Coron, and L. Rosier. On homogeneous finite-time control for linear evolution equation in hilbert space. *IEEE Transactions on Automatic Control*, 63(9):3143–3150, 2018.
- [30] A. Polyakov, D. Efimov, and W. Perruquetti. Finite-time and fixed-time stabilization: Implicit Lyapunov function approach. *Automatica*, 51(1):332–340, 2015.
- [31] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [32] L. Rosier. Homogeneous Lyapunov function for homogeneous continuous vector field. *Systems & Control Letters*, 19:467–473, 1992.
- [33] E.P. Ryan. Universal stabilization of a class of nonlinear systems with homogeneous vector fields. *Systems & Control Letters*, 26:177–184, 1995.
- [34] K. Zimenko, A. Polyakov, D. Efimov, and W. Perruquetti. Robust feedback stabilization of linear mimo systems using generalized homogenization. *IEEE Transactions on Automatic Control*, 65(12):5429–5436, 2020.
- [35] V. I. Zubov. *Methods of A.M. Lyapunov and Their Applications*. Noordhoff, Leiden, (Translated from V.I. Zubov, *Metody A.M. Lyapunova i ih primeneniye*, Izdatelstvo Leningradskogo universiteta, 1957) edition, 1964.
- [36] V.I. Zubov. On systems of ordinary differential equations with generalized homogeneous right-hand sides. *Izvestia vuzov. Mathematica (available on-line: <http://mi.mathnet.ru/ivm2874>)*, 1:80–88, 1958.

7 Reviews received from the journal ” Mathematics of Control, Signals, and Systems”

Reviewer 1: The paper I have received for refereeing is the first revision of a draft submitted to Mathematics of Control, Signals, and Systems, in which the modifications and additions have been highlighted in blue.

The paper’s main result is the observation that the celebrated Cybenko-Hornik neural networks universality theorem can be adapted in a structure-preserving fashion to generalized homogeneous functions. The proof of this result is relatively simple and follows as a straightforward corollary of the classical theorem. The most important gain that is obtained out of the preservation of the homogeneous structure is that the thesis of the theorem on the possibility of uniform approximation is not only valid on compacta (as in Cybenko’s classical theorem) but globally. The importance of homogeneity is very nicely illustrated in the paper in the newly introduced Examples 1 (in connection with pattern recognition) and 2 (homogeneous rotation dynamics). Another substantial example has been added in the new Section 5 on the stabilization of LTI plants in which taking homogeneity into account has an important positive influence.

The manuscript is well structured and written (the English, especially the use of articles, has to be revised before publication), and the contents are clearly conveyed to the reader. A different question is if the depth of the mathematical contents in the paper suffices for publication in Mathematics of Control, Signals, and Systems, which traditionally accepts papers with substantial mathematical contributions. In that respect, the new examples and applications added during the revision have not changed the paper much in that department. I think it would have been preferable to dwell on how, for example, approximation bounds/rates à la Maurey/Barron/Jones are articulated in the homogeneous setup. This is just one of many other theoretical topics on neural approximation theory that the author could have considered to increase the theoretical density of his contribution.

Having said all that, my impression of the paper is positive, and I believe it deserves publication. I leave it to the editor to decide if the venue should be Mathematics of Control, Signals, and Systems.

Reviewer 2: In this paper, the author presents a curated adaptation of a two-layer neural network which accounts for linear dilations, namely, so that the resulting neural network homogeneous (of a fixed degree) with respect to a fixed dilation.

On one hand, this paper is not carefully written. Not only is it rife with typos, but the presentation is extremely dense and confusing, and critical definitions (such as that of a linear dilation) are given inline, within a paragraph, without any emphasis, instead of an explicit say. In my opinion, the paper in its present form is closer to a draft than to a publishable article, especially for a journal in the vein of MCSS. The author is welcome to follow guidelines, such as those suggested in the AMS guide, on proper form.

On the other hand, I do not truly understand what are the novel insights from the results or setup presented herein. First of all, with regard to the proof of the main result (Theorem 3), which ensures

universal approximation of d -homogeneous continuous functions by means of the novel two-layer architecture presented herein. The main novelty here is the fact that approximation holds globally on the entire euclidean space, instead of locally, as in the celebrated result by G. Cybenko. While I do not strictly dispute this, the proof is not more than a straightforward adaptation of the former result. Furthermore, it has been well understood by now that, Cybenko's result, while seminal, is not quantitative and thus not representative of current trends and practices. The author could comment on whether an adaptation of A. Barron's celebrated quantitative universal approximation result (which is not even cited!) is possible in this particular setting. Extensions to deep architectures, and adding skip connections as in ResNets, should also be discussed at the very least, since these are staples of contemporary architectures. Beyond the main result, the applications, in particular the image processing one, do not appear particularly compelling. Indeed while it is true that a two-layer neural network will a priori not take into account changes in scale, the truth is that not a single image processing task is ever solved with a dense two layer neural network, but rather with much more structured networks that account for various group actions. In the past, convolutional neural networks (CNNs) were (and in fact still are, to a degree) the norm. And for CNNs, there is a series of works initiated by S. Mallat which deal with the various invariance or equivariance with respect to diffeomorphisms. These papers merit a discussion herein, at the very least. These papers have since been cited hundreds or even thousands of times, and I am fairly confident that some of them might have an overlap with the setup presented by the author.

For these reasons, I do not recommend acceptance of this paper. I would encourage the author to delve deeper into the existing literature and compare with past works, and to increase the quality of the presentation, before submitting this article elsewhere.