Towards Explainability of Approximate Lifted Model Construction: A Geometric Perspective

Jan Speller[1,*], Malte Luttermann[2,3], Marcel Gehrke[3] and Tanya Braun[1]

[1]University of Münster, Computer Science Department, Einsteinstraße 62, 48149 Münster, Germany
[2]German Research Center for Artificial Intelligence (DFKI), Ratzeburger Allee 160, 23562 Lübeck, Germany
[3]University of Hamburg, Institute for Humanities-Centered AI (CHAI), Warburgstraße 28, 20354 Hamburg, Germany

Abstract

Advanced colour passing (ACP) is the state-of-the-art algorithm for lifting a propositional probabilistic model to a first-order level by combining exchangeable factors, enabling the use of lifted inference algorithms to allow for tractable probabilistic inference with respect to domain sizes. More recently, an approximate version of ACP, called ε-ACP, ensures the practical applicability of ACP by accounting for inaccurate estimates of underlying distributions. ε-ACP permits underlying distributions, encoded as potential-based factorisations, to slightly deviate depending on a hyperparameter ε while maintaining a bounded approximation error. To navigate through different levels of compression versus accuracy, a hierarchical version of ε-ACP has emerged that builds a hierarchy of ε values. In a drive towards interpretability of results, this paper looks at geometric properties of ε-equivalence, a central notion employed by ε-ACP and its hierarchical version to quantify the maximum allowed deviation between potentials. Specifically, we present a unified view on the results for ε-ACP and its hierarchical version and provide a geometric interpretation of ε-equivalence in $\mathcal{L}^p$, thereby making results more interpretable.

Keywords

Lifted Inference, Geometric Interpretability, Approximate Model, Factor Graph, Colour Passing

1. Introduction

Intelligent systems tasked with modelling its environment have to contend with uncertainty in the world as well as objects and relations among them. The former is canonically modelled using probabilistic graphical models, encoding features as random variables (randvars) and relations between them in a factorised probability distribution. The latter is often represented using some form of higher-order logic, describing a world through hard constraints on objects and their relations. Combining both perspectives has lead to various probabilistic relational formalisms: Parametric factor graphs (PFGs) [1] emerge from the uncertainty perspective with logical constructs added to model objects and relations among them under uncertainty. Markov logic networks [2] originate from the logic perspective with weights added to the constraints to denote how likely they are to hold. Both formalisms follow grounding semantics [3, 4], yielding a full joint probability distribution over indistinguishable grounded (propositional) randvars. Over the years, many researchers have focused on developing efficient inference algorithms in such models, exploiting these first-order structures under the name of lifting [1] for episodic reasoning [5, 6, 7, 8, 9, 10], queries [11, 12], evidence [9, 13], temporal reasoning [14, 15], and decision making [16, 17, 18, 19, 20, 21, 22], allowing for tractable inference in the number of objects [23].

While these works document the impressive progress made over the years, these works usually assume a first-order model lifted from the ground level to start with. How to lift a propositional model has been a more overlooked research question. Colour Passing (CP) is the state-of-the-art algorithm to turn a propositional model into a first-order one, specifically lifting a factor graph (FG)

Perspectives on Humanities-Centred AI and Formal & Cognitive Reasoning Workshop 2025 (CHAI 2025 & FCR 2025), Joint Workshop at the 48th German Conference on Artificial Intelligence, September 16, 2025, Potsdam, Germany
*Corresponding author.
jan.speller@uni-muenster.de (J. Speller); malte.luttermann@dfki.de (M. Luttermann); marcel.gehrke@uni-hamburg.de (M. Gehrke); tanya.braun@uni-muenster.de (T. Braun)
0000-0003-1106-8177 (J. Speller); 0009-0005-8591-6839 (M. Luttermann); 0000-0001-9056-7673 (M. Gehrke); 0000-0003-0282-4284 (T. Braun)

to a PFG, grouping factors with identical potentials [5], based on the Weisfeiler-Leman algorithm for graphs [24]. CP has been revisited over ten years later to further optimise the algorithm, considering commutativity among the randvars in factors [25, 26, 27] as well as scaling of potentials [28]. To deal with the fact that identical potentials are highly unlikely to emerge naturally in a ground FG, even if the underlying objects behave almost indistinguishably, ε-Advanced Colour Passing (ε-ACP) constitutes an approximate extension that groups factors whose potentials differ by a factor of at most $(1 \pm \varepsilon)$, with ε as a hyperparameter [29]. In a step towards making the algorithm hyperparameter free, Hierarchical Advanced Colour Passing (HACP) proceeds to build a hierarchy of ε values that lead to increasingly more compact encodings at the expense of a higher approximation error [30], while keeping previous compressions in tact to improve explainability.

This paper takes the most recent advances in approximating Advanced Colour Passing (ACP) and builds toward making the resulting compressed representation more interpretable by considering a geometric perspective. Specifically, this paper contributes

(i) a unified overview of the results for ε-ACP and HACP,

(ii) proofs of novel properties of ε-equivalence in $\mathcal{L}^p$,

(iii) a discussion of their value for interpretability, and

(iv) a glimpse of potential advances for approximate lifted model construction.

The remaining part of this paper is structured as follows. We start with basic definitions and a recap of the results for ε-ACP and HACP. The main part then presents the geometric view of ε-equivalence in the $\mathcal{L}^p$ space, including a closer look at the special case of the Euclidean space. The paper ends with a discussion and conclusion about the potential benefits of the new view and the resulting possibilities for further generalisations of ACP and its approximate variants.

2. Factor Graphs and Approximate Colour Passing

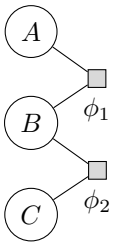
To establish a formal foundation for lifted probabilistic inference under approximation, we use the same notation and definitions of Luttermann et al. [29] and its continuation [30]. The following sequence introduces FGs, probabilistic queries, and then progresses towards the central notion of ε-equivalent factors, which builds the backbone of the initial approximate lifted model construction approaches (i.e., ε-ACP and HACP) and the foundation of a geometric perspective developed in Section 3. First, properties of the introduced concepts are followed by a summary of the algorithms ε-ACP and HACP for approximate lifted model construction and their asymptotic behaviour in Section 2.1 and Section 2.2.

Definition 1 (Factor graph [29, Def. 1])**.** *An FG $M = (\boldsymbol{V}, \boldsymbol{E})$ is an undirected bipartite graph consisting of a node set $\boldsymbol{V} = \boldsymbol{R} \cup \boldsymbol{\Phi}$, where $\boldsymbol{R} = \{R_1, \ldots, R_n\}$ is a set of randvars and $\boldsymbol{\Phi} = \{\phi_1, \ldots, \phi_m\}$ is a set of factors (functions), as well as a set of edges $\boldsymbol{E} \subseteq \boldsymbol{R} \times \boldsymbol{\Phi}$. There is an edge between a randvar $R_i \in \boldsymbol{R}$ and a factor $\phi_j \in \boldsymbol{\Phi}$ in $\boldsymbol{E}$ if R_i appears in the argument list of ϕ_j. A factor $\phi_j(\mathcal{R}_j)$ defines a function $\phi_j \colon \times_{R \in \mathcal{R}_j} \mathrm{range}(R) \mapsto \mathbb{R}_{>0}$ that maps the ranges of its arguments $\mathcal{R}_j$ (a sequence of randvars from $\boldsymbol{R}$) to a positive real number, called potential. The term $\mathrm{range}(R)$ denotes the possible values a randvar R can take. We further define the joint potential for an assignment $\boldsymbol{r}$ (with $\boldsymbol{r}$ being a shorthand for $\boldsymbol{R} = \boldsymbol{r}$) as*

$$\psi(\boldsymbol{r}) = \prod_{j=1}^{m} \phi_j(\boldsymbol{r}_j), \tag{1}$$

where $\boldsymbol{r}_j$ is a projection of $\boldsymbol{r}$ to the argument list of ϕ_j. With $Z = \sum_{\boldsymbol{r}} \prod_{j=1}^{m} \phi_j(\boldsymbol{r}_j)$ as the normalisation constant, the full joint probability distribution encoded by M is then given by

$$P_M(\boldsymbol{r}) = \frac{1}{Z} \prod_{j=1}^{m} \phi_j(\boldsymbol{r}_j) = \frac{1}{Z} \psi(\boldsymbol{r}). \tag{2}$$



A	B	$\phi_1(A,B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

C	B	$\phi_2(C,B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

Figure 1: An example for an FG that represents a probability distribution over three randvars A, B, and C as a product over two factors ϕ_1 and ϕ_2.

Example 1 (Factor Graph). *Take a look at the FG depicted in Fig. 1. For the sake of the example, it holds that* $\boldsymbol{R} = \{A, B, C\}$, $\boldsymbol{\Phi} = \{\phi_1, \phi_2\}$, $\boldsymbol{E} = \{\{A, \phi_1\}, \{B, \phi_1\}, \{B, \phi_2\}, \{C, \phi_2\}\}$, *and* $\mathit{range}(A) = \mathit{range}(B) = \mathit{range}(C) = \{\text{true}, \text{false}\}$. *The potential tables (i.e., function definitions) of the factors* ϕ_1 *and* ϕ_2 *are shown next to the graph on the right. Specifically, it holds that* $\phi_1(\mathit{true}, \mathit{true}) = \varphi_1$, $\phi_1(\mathit{true}, \mathit{false}) = \varphi_2$, *and so on, where* $\varphi_i \in \mathbb{R}_{>0}$, $i = 1, \ldots, 4$, *are arbitrary positive real numbers.*

An FG enables an efficient representation of complex distributions by decomposing them into factors over subsets of variables. Each factor captures an aspect of the joint distribution, giving rise to natural queries that we will define next to determine the objectives of probabilistic inference.

Definition 2 (Query [29, Def. 2]). *A query* $P(Q \mid E_1 = e_1, \ldots, E_k = e_k)$ *consists of a query term* Q *and a set of events* $\{E_j = e_j\}_{j=1}^k$ *where* Q *and all* E_j, $j = 1, \ldots, k$, *are randvars. To query a specific probability instead of a distribution, the query term is an event* $Q = q$.

Example 2 (Query). *The query* $P(A \mid B = \mathit{true})$ *asks for the probability distribution of* A *given that the event* $B = \mathit{true}$ *is observed.*

To enable approximate inference and structure compression, the notion of ε-*equivalence* has been introduced by Luttermann et al. [29]. ε-equivalence relaxes exact equality between potentials and enables coarser, more efficient representations under bounded deviation.

Definition 3 (ε-Equivalence [29, Def. 3]). *Let* $\varepsilon \in \mathbb{R}_{>0}$ *be a positive real number. Two potentials* $\varphi_1, \varphi_2 \in \mathbb{R}_{>0}$ *are* ε-equivalent, *denoted as* $\varphi_1 =_\varepsilon \varphi_2$, *if* $\varphi_1 \in [\varphi_2 \cdot (1-\varepsilon), \varphi_2 \cdot (1+\varepsilon)]$ *and* $\varphi_2 \in [\varphi_1 \cdot (1-\varepsilon), \varphi_1 \cdot (1+\varepsilon)]$. *Further, two factors* $\phi_1(R_1, \ldots, R_n)$ *and* $\phi_2(R'_1, \ldots, R'_n)$ *are* ε-equivalent, *denoted as* $\phi_1 =_\varepsilon \phi_2$, *if there exists a permutation* π *of* $\{1, \ldots, n\}$ *such that for all assignments* $(r_1, \ldots, r_n) \in \times_{i=1}^n \mathit{range}(R_i)$, *where* $\phi_1(r_1, \ldots, r_n) = \varphi_1$ *and* $\phi_2(r_{\pi(1)}, \ldots, r_{\pi(n)}) = \varphi_2$, *it holds that* $\varphi_1 =_\varepsilon \varphi_2$.

Example 3 (ε-Equivalence). *Consider the potentials* $\varphi_1 = 0.49$, $\varphi_2 = 0.5$, *and let* $\varepsilon = 0.1$. *Due to* $\varphi_2 = 0.5 \in [\varphi_1 \cdot (1-\varepsilon) = 0.441, \varphi_1 \cdot (1+\varepsilon) = 0.539]$ *and* $\varphi_1 = 0.49 \in [\varphi_2 \cdot (1-\varepsilon) = 0.45, \varphi_2 \cdot (1+\varepsilon) = 0.55]$, *it holds that* φ_1 *and* φ_2 *are* ε-*equivalent (for* $\varepsilon = 0.1$).

In general, it might happen that indistinguishable randvars are located at different positions in the argument list of their respective factors, which is the reason the definition of ε-equivalence involves permutations of arguments. For simplicity, in this paper, we stipulate that π is the identity function (that is, we assume that for two ε-equivalent factors, all potentials in their potential tables are row-wise ε-equivalent). However, all results of this paper also apply to any other choice of π [29].

While the above definition captures only indirectly the smallest possible value ε_0 for which ε-equivalence holds, Speller et al. [30] introduced an easier access to the concept via a vector-based distance measure called one-dimensional ε-equivalence distance (1DEED), which enables us to quantify deviations pairwise between factors and decide ε-equivalence in a scalable and computationally more accessible manner. In the following, we denote the potential table of a factor ϕ as a vector in $\mathbb{R}_{>0}^n$, where $\phi(k)$ denotes the k-th entry, i.e., the potential associated with the k-th row, in the potential table of ϕ.

Definition 4 (One-dimensional ε-equivalence distance and ε-equivalence [30, Def. 4])**.** 1DEED*, defined as the mapping* $d_\infty \colon \mathbb{R}^n_{>0} \times \mathbb{R}^n_{>0} \to \mathbb{R}$ *for two n-dimensional vectors* $\phi_1, \phi_2 \in \mathbb{R}^n_{>0}$*, is given by*

$$\begin{aligned} d_\infty(\phi_1, \phi_2) &:= \max_{k=1,\dots,n} \left\{ \left| \frac{\phi_1(k) - \phi_2(k)}{\phi_1(k)} \right|, \left| \frac{\phi_1(k) - \phi_2(k)}{\phi_2(k)} \right| \right\} \\ &= \max_{k=1,\dots,n} \left\{ \frac{|\phi_1(k) - \phi_2(k)|}{\min\{|\phi_1(k)|, |\phi_2(k)|\}} \right\}. \end{aligned} \tag{3}$$

With Def. 4, the concept of ε-equivalence can be re-expressed in terms of 1DEED, which Speller et al. [30] demonstrate to be equivalent to the original definition of ε-equivalence (Def. 3) via the upcoming Theorem.

Theorem 1 ([30, Thm. 2])**.** *Two vectors* $\phi_1, \phi_2 \in \mathbb{R}^n_{>0}$ *are* ε*-equivalent (Definition 3) if and only if* $d_\infty(\phi_1, \phi_2) \leq \varepsilon$ *holds.*

The next lemma provides a brief overview of the main characteristics of 1DEED and ε-equivalence.

Lemma 2 ([30, Cor. 1, Prop. 9])**.** *The following properties hold.*

(i) 1DEED is non-negative and symmetric.

(ii) It holds that $d_\infty(\phi_1, \phi_2) = 0$ *if and only if* $|\phi_1(k) - \phi_2(k)| = 0$ *for all* $k = 1, \dots, n$*, which holds only if* $\phi_1 = \phi_2$*.*

(iii) ε*-equivalence is not transitive.*

Since transitivity of ε-equivalence does not hold (Lemma 2, *iii*), ε-equivalence cannot be considered an equivalence relation and thus is not suitable for defining equivalence classes. This is the main reason why the consideration of a hierarchical variant makes sense, because there is not necessarily a unique ordering of groups for all possible ε-values.

For more comprehensive examples of the defined concepts and the proofs of their first properties, we refer the reader to the detailed descriptions given by Luttermann et al. [29] and Speller et al. [30].

2.1. The Algorithms ε-ACP and HACP

Both the ε-ACP algorithm and the HACP algorithm are primarily based on the ACP algorithm [25]. The ACP algorithm aims to identify symmetries within an FG, specifically those arising from exactly matching factors (that is, factors whose potentials are strictly equal). In this context, scalar multiples of factors can also be treated as equivalent, as a rescaling to the same magnitude can be achieved via normalisation [31]. In case the number of detected symmetries is insufficient or an FG with potentials estimated from data is given, for which scalability is desired, the concept of ε-equivalence (Def. 3 and Def. 4) has been introduced to approximate the construction of a lifted (compressed) model. The concept of ε-equivalence forms the foundation for the development of the ε-ACP algorithm.

Given $\varepsilon \geq 0$, groups $\boldsymbol{G}_l = \{\phi^l_1, \dots, \phi^l_{m_l}\}$ of pairwise ε-equivalent factors $\phi^l_i \in \mathbb{R}^n_+$ are determined. A representative factor is chosen for every group $\boldsymbol{G}_l$ by computing the optimal solution that minimises the loss function between the representative and every factor in $\boldsymbol{G}_l$, where the loss function between two factors ϕ^l_i and ϕ_j is defined as the sum of squared deviations

$$Err(\phi^l_i, \phi_j) = \sum_{r_1,\dots,r_n} \Big(\phi^l_i(r_1, \dots, r_n) - \phi_j(r_1, \dots, r_n) \Big)^2, \tag{4}$$

with $\boldsymbol{r} = r_1, \dots, r_n$ denoting all possible assignments to the arguments of ϕ^l_i and ϕ_j. The representative ϕ^{l*} of each group is chosen as

$$\phi^{l*} = \arg\min_{\phi_j} \sum_{\phi^l_i \in \boldsymbol{G}_l} Err(\phi^l_i, \phi_j), \tag{5}$$

to minimise the total deviation between ϕ^{l*} and the group. All $\phi_i^l \in \boldsymbol{G}_l$ are replaced by the representative ϕ^{l*}, which is computed as the arithmetic mean over all factors in the group $\boldsymbol{G}_l$ [29, Thm. 1]:

$$\phi^{l*}(\boldsymbol{r}) = \frac{1}{m_l} \sum_{i=1}^{m_l} \phi_i^l(\boldsymbol{r}). \tag{6}$$

Replacing a group $\boldsymbol{G}_l$ of factors by a single representative significantly reduces the storage requirements and thus also drastically speeds up run times for probabilistic inference depending on the size of ε, provided sufficient symmetries are present. For more details, see the work by Luttermann et al. [29].

Building on the sorted groups of pairwise ε-equivalent factors, a hierarchical algorithm (HACP) has been introduced [30, Alg. 2] to ensure a consistent comparison of various approximate compressed FGs and to guide the choice of suitable ε values. To this end, an ordering of group memberships is first established based on 1DEED [30, Alg. 1], which can be interpreted as an agglomerative clustering algorithm using 1DEED as a base distance. Only completely pairwise ε-equivalent groups are allowed to be merged (complete linkage within maximal deviation). The design of the implementation of the ε-ACP algorithm and the HACP algorithm guarantees the following useful properties.

Corollary 3 ([29, Cor. 2])**.** *If $\varepsilon = 0$, ε-ACP returns the same PFG as ACP.*

Corollary 4. *If $\varepsilon = 0$, HACP returns the same PFG as ACP.*

Proof. Follows directly from [30, Prop. 4], with $\varepsilon = 0$. □

2.2. Asymptotic Properties

A key advantage of the HACP algorithm lies in its ability to retain the same upper bound on the deviation of probabilistic queries as established for ε-ACP. For a fixed $\varepsilon > 0$, ε-ACP ensures that the deviation of any query result in the approximate compressed FG from the original distribution remains within a boundary. HACP, although introducing a hierarchical structure, builds upon the same guarantee. To quantify the change in query results between the original and modified (compressed) model, we adopt the symmetric divergence measure introduced by Chan and Darwiche [32], which provides a tight bound on the maximal deviation between two distributions P_M and $P_{M'}$ with respect to any assignment $\boldsymbol{r}$ via

$$D_{CD}(P_M, P_{M'}) := \ln \max_{\boldsymbol{r}} \frac{P_{M'}(\boldsymbol{r})}{P_M(\boldsymbol{r})} - \ln \min_{\boldsymbol{r}} \frac{P_{M'}(\boldsymbol{r})}{P_M(\boldsymbol{r})}. \tag{7}$$

The next results are proven asymptotic properties for the ε-ACP and HACP algorithm shown by Luttermann et al. [29] and Speller et al. [30], respectively.

Theorem 5 ([29, Thm. 7], [30, Prop. 4])**.** *Let $M = (\boldsymbol{R} \cup \boldsymbol{\Phi}, \boldsymbol{E})$ be an FG and let M' be the output of ε-ACP or the output of HACP when run on M. With P_M and $P_{M'}$ being the underlying full joint probability distributions encoded by M and M', respectively, and $m = |\boldsymbol{\Phi}|$, it holds that*

$$D_{CD}(P_M, P_{M'}) \leq \ln \left(\frac{\left(1 + \frac{m-1}{m}\varepsilon\right)\left(1+\varepsilon\right)}{1 + \frac{1}{m}\varepsilon} \right)^m \tag{8}$$

$$< \ln \left(1+\varepsilon\right)^{2m} \tag{9}$$

$$< \ln \left(\frac{1+\varepsilon}{1-\varepsilon} \right)^m, \tag{10}$$

where the bound given in Eq. (8) is optimal (i.e., sharp).

Even though the worst-case bound can be attained in constructed examples, such cases are rare in practice. Nonetheless, the need to control this bound persists, as the selection of a specific ε for either algorithm can be guided by a prior assessment of the maximum permissible deviation. This enables a more informed decision before executing the actual lifted probabilistic inference algorithm and, consequently, before executing the ε-ACP or HACP algorithm.

Theorem 6 ([30, Thm. 5]). *The maximal absolute deviation between any initial probability $p = P_M(r \mid \boldsymbol{e})$ of r given $\boldsymbol{e}$ in model M and the probability $p' = P_{M'}(r \mid \boldsymbol{e})$ in the modified model M' resulting from running ε-ACP or HACP on M can be bounded by*

$$p_{\max\Delta} := \max_{\text{for any } r|\boldsymbol{e}} |p - p'| \leq \frac{\sqrt{e^d} - 1}{\sqrt{e^d} + 1} \text{ with } d = D_{CD}(P_M, P'_M).$$

Due to the monotonicity of $\frac{\sqrt{e^d}-1}{\sqrt{e^d}+1}$ in d, Theorem 6 can easily be converted into the next corollary.

Corollary 7 ([30, Cor. 6]). *The change in any probabilistic query in an initial model M and a modified model M' obtained by running ε-ACP or HACP is bounded by*

$$\begin{aligned} p_{\max\Delta} &\leq \frac{\sqrt{e^{d_1}} - 1}{\sqrt{e^{d_1}} + 1} \text{ with } d_1 = D_{CD}(P_M, P'_M) \\ &\leq \frac{\sqrt{e^{d_2}} - 1}{\sqrt{e^{d_2}} + 1} \text{ with } d_2 = \ln\left(\frac{\left(1 + \frac{m-1}{m}\varepsilon\right)\left(1+\varepsilon\right)}{1 + \frac{1}{m}\varepsilon}\right)^m. \end{aligned}$$

While we have calculated the deviation that follows from a specific choice of ε so far, we can also calculate how large ε can be chosen to get a maximal deviation of p^*_Δ we want to allow.

Theorem 8 ([30, Thm. 7]). *For any given $p^*_\Delta \in (0, \frac{1}{2}]$, the output of ε-ACP and HACP guarantees for any $\varepsilon \in (0, 1)$, which is smaller or equal to*

$$\varepsilon_1 = -\frac{1 + \frac{m-1}{m} - \frac{1}{m}\sqrt[m]{e^d}}{2\,\frac{m-1}{m}} + \sqrt{\left(-\frac{1 + \frac{m-1}{m} - \frac{1}{m}\sqrt[m]{e^d}}{2\,\frac{m-1}{m}}\right)^2 - \frac{1 - \sqrt[m]{e^d}}{\frac{m-1}{m}}}$$

with $d = \ln\left(\frac{p^*_\Delta + 1}{1 - p^*_\Delta}\right)^2$ the bound $p_{\max\Delta} \leq p^*_\Delta$.

Therefore, a calculation of $\varepsilon_1(p^*_\Delta)$ bounds the maximal deviation p^*_Δ of ε-ACP and HACP, respectively, before the algorithm has been started.

3. Geometric View of ε-Equivalence

Although ε-equivalence works well in practice, it introduces a conceptual barrier due to its reduction of the relation between two factors to a single relative distance via 1DEED. This section aims to enhance the interpretability of the concept and to enable potential generalisations for ε-ACP and HACP. We investigate the geometric implications of ε-equivalence and the behaviour of pairwise ε-equivalent groups. All factors are assumed to lie in $\mathbb{R}^n_+$, which corresponds to an orthant in the Euclidean space. However, to allow for a more general setting, we consider the general $\mathcal{L}^p$ norm in Section 3.1 rather than sticking to the Euclidean setting.

Before we start, we summarise the proven results on the relative ordering of lengths directly derived from the original definition of ε-equivalence (Def. 3). Two ε-equivalent factors may deviate maximally in the positive and negative direction, respectively, and due to symmetry, the expression $(1 - \varepsilon)$ can more precisely be replaced by the expression $\frac{1}{1+\varepsilon}$.

Lemma 9 ([29, Lem. 6]). *For two ε-equivalent factors $\phi_1, \phi_2 \in \mathbb{R}^n_+$, it holds that $\phi_1(k) \in [\phi_2(k) \cdot \frac{1}{1+\varepsilon}, \phi_2(k) \cdot (1+\varepsilon)]$ and $\phi_2(k) \in [\phi_1(k) \cdot \frac{1}{1+\varepsilon}, \phi_1(k) \cdot (1+\varepsilon)]$ for $k = 1, \ldots, n$.*

From Lemma 9, we derive a corollary reformulating the same statement in terms of quotients.

Corollary 10. *If $\phi_1 =_\varepsilon \phi_2$ holds for two factors, then it also holds that*

$$\frac{\phi_1(k)}{\phi_2(k)} \in \left[\frac{1}{1+\varepsilon}, 1+\varepsilon\right] \text{ and } \frac{\phi_2(k)}{\phi_1(k)} \in \left[\frac{1}{1+\varepsilon}, 1+\varepsilon\right] \text{ for all } k \in \{1, \ldots, n\}. \tag{11}$$

Corollary 10 implies that the length deviations per dimension remain relatively small, and that deviations can be interpreted around the multiplicative identity 1.

3.1. Geometric Perspective in $\mathcal{L}^p$

Considering distances between two ε-equivalent, normalised factors yields the following result.

Lemma 11. *Let ϕ_1, ϕ_2 be two factors in $\mathbb{R}^n_+ \cap S_p^{n-1}$ with $S_p^{n-1} := \{\phi \in \mathbb{R}^n : \|\phi\|_p = 1\}$ being the unit $n-1$-sphere in $\mathcal{L}^p$ for $p \geq 1$. If $\phi_1 =_\varepsilon \phi_2$ holds, then we have*

(i) $\|\phi_1 - \phi_2\|_p \leq \varepsilon$, *and*

(ii) $\phi_{3-i} \in \overline{B_\varepsilon^p}(\phi_i) \cap S_p^{n-1}$ *for* $i = 1, 2$ *with* $\overline{B_\varepsilon^p}(\phi_i) := \{\phi \in \mathbb{R}^n : \|\phi - \phi_i\|_p \leq \varepsilon\}$ *being the closed p-ball with centre ϕ_i and radius $\varepsilon > 0$.*

Proof. By definition of ε-equivalence, we get for (i) the property

$$\phi_1(k) \in [(1-\varepsilon)\phi_2(k), (1+\varepsilon)\phi_2(k)] \text{ for all } k = 1, \ldots, n, \tag{12}$$

which is equivalent to

$$\phi_1(k) - \phi_2(k) \in [-\varepsilon\phi_2(k), \varepsilon\phi_2(k)] \text{ for all } k = 1, \ldots, n, \tag{13}$$

or in shorter form $|\phi_1(k) - \phi_2(k)| \leq \varepsilon\phi_2(k)$ for all $k = 1, \ldots, n$. Substituting this into the $\mathcal{L}^p$ norm taking into account the normalisation $\|\phi_2\|_p = 1$, we obtain the desired result:

$$\begin{aligned} \|\phi_1 - \phi_2\|_p^p &= \sum_{k=1}^n |\phi_1(k) - \phi_2(k)|^p && (14)\\ &\leq \sum_{k=1}^n |\varepsilon\phi_2(k)|^p && (15)\\ &= \varepsilon^p \sum_{k=1}^n |\phi_2(k)|^p = \varepsilon^p \|\phi_2\|_p^p = \varepsilon^p. && (16) \end{aligned}$$

Property (ii) follows directly from property (i). □

The previous proof also contains a result for non-normalised factors, presented in the next corollary.

Corollary 12. *Let ϕ_1, ϕ_2 be two ε-equivalent factors in $\mathbb{R}^n_+$. Then, $\|\phi_1 - \phi_2\|_p \leq \varepsilon\|\phi_i\|_p$ holds for $i = 1, 2$.*

Proof. Follows from Eq. (16). By symmetry, the claim holds for $i = 2$ and also $i = 1$. □

In contrast, when viewing the problem from the opposite direction, we can only establish a relatively weak statement about ε-equivalence.

Lemma 13. *Let ϕ_1, ϕ_2 be two factors in $\mathbb{R}^n_+$. If $\|\phi_1 - \phi_2\|_p \leq \varepsilon$, then $\phi_1 =_{\varepsilon'} \phi_2$ holds with*

$$\varepsilon' := \frac{\varepsilon}{\min\limits_{\substack{k=1,\ldots,n,\\ i=1,2}} \{\phi_i(k)\}}. \tag{17}$$

Proof. Consider

$$\|\phi_1 - \phi_2\|_p^p = \sum_{k=1}^n |\phi_1(k) - \phi_2(k)|^p \leq \varepsilon^p. \tag{18}$$

Since $|\phi_1(k) - \phi_2(k)|^p \geq 0$ for all k, Eq. (18) also holds for every k individually: $|\phi_1(k) - \phi_2(k)| \leq \varepsilon$. Therefore, we get

$$\phi_1(k) \leq \varepsilon + \phi_2(k) = \left(1 + \frac{\varepsilon}{\phi_2(k)}\right)\phi_2(k) \tag{19}$$

and

$$\phi_1(k) \geq -\varepsilon + \phi_2(k) = \left(1 - \frac{\varepsilon}{\phi_2(k)}\right)\phi_2(k), \tag{20}$$

which implies

$$\phi_1(k) \in \left[\left(1 - \frac{\varepsilon}{\phi_2(k)}\right)\phi_2(k), \left(1 + \frac{\varepsilon}{\phi_2(k)}\right)\phi_2(k)\right]. \tag{21}$$

Analogously, we get the opposite inequality and interval inclusion for $\phi_2(k)$

$$\phi_2(k) \in \left[\left(1 - \frac{\varepsilon}{\phi_1(k)}\right)\phi_1(k), \left(1 + \frac{\varepsilon}{\phi_1(k)}\right)\phi_1(k)\right]. \tag{22}$$

This is true for each dimension k. To guarantee

$$\phi_{3-i}(k) \in \left[\left(1 - \varepsilon'\right)\phi_i(k), \left(1 + \varepsilon'\right)\phi_i(k)\right] \tag{23}$$

for $i = 1, 2$, for $k = 1, \ldots, n$, and a fixed ε' and therefore to guarantee ε-equivalence, we need to choose the largest $\frac{\varepsilon}{\phi_i(k)}$ value for ε'. It is easy to see that the broadest interval is generated by the smallest denominator among all possibilities, which is given by $\min_{k=1,\ldots,n,i=1,2}\{\phi_i(k)\}$. □

For a given $\varepsilon > 0$, we can construct an example that hits the same bound as mentioned in Lemma 13 independently of the choice of the p-norm, because the worst deviation may occur in a single dimension:

Example 4. *Let p be any value in $[1, \infty)$ and $\varepsilon = 0.01$, then the $\mathcal{L}^p$ norm of the two factors*

$$\phi_1 = \begin{pmatrix} 0.01 \\ 1.0 \end{pmatrix}, \phi_2 = \begin{pmatrix} 0.02 \\ 1.0 \end{pmatrix} \tag{24}$$

is given by

$$\|\phi_1 - \phi_2\|_p^p = \sum_{k=1}^{2} |\phi_1(k) - \phi_2(k)|^p = 0.01^p + 0^p = 0.01^p = \varepsilon^p \tag{25}$$

and therefore fulfils the conditions of Lemma 13. Still, it leads to a comparably high value for ε-equivalence via $\phi_2(1) = 2.0 \cdot \phi_1(1) = (1 + \varepsilon')\phi_1(1)$ and

$$\varepsilon' = \frac{\varepsilon}{\min\limits_{\substack{k=1,\ldots,n,\\ i=1,2}} \{\phi_i(k)\}} = \frac{0.01}{0.01} = 1. \tag{26}$$

This example can be extended with identical values $\phi_1(k) = \phi_2(k)$ for all $k > 1$ to arbitrary large dimensions n. When normalising the two vectors $\|\phi_i\|_p = 1$ for $i = 1, 2$, the differences in one dimension will affect the others, resulting in an ε' that depends on p and the number of dimensions n. Moreover, even normalisation cannot improve upon the result of Lemma 13, as entries of a single dimension k can be arbitrarily small, leading to a large discrepancy between $\phi_1(k)$ and $\phi_2(k)$ in relative terms.

3.2. Groups of Pairwise ε-Equivalent Factors

The current implementation of ε-ACP and HACP uses the arithmetic mean over all factors in a group of pairwise ε-equivalent factors. An important property is that the factor defined as the arithmetic mean over a group of pairwise ε-equivalent factors itself is again ε-equivalent to all factors in the group.

Lemma 14 ([29, Lem. 5])**.** *Let $\boldsymbol{G} = \{\phi_1, \ldots, \phi_k\}$ denote a group of pairwise ε-equivalent factors in $\mathbb{R}_+^n$ and let $\phi^*(\boldsymbol{r}) = \frac{1}{k}\sum_{i=1}^{k} \phi_i(\boldsymbol{r})$ for all assignments $\boldsymbol{r}$. Then, $\boldsymbol{G}^* = \{\phi_1, \ldots, \phi_k, \phi^*\}$ is a group of pairwise ε-equivalent factors.*

According to Luttermann et al. [29, Thm. 1], the arithmetic mean is the optimal choice for the squared loss objective used for minimisation within the error function $Err(\phi_i^l, \phi_j)$ (see Eq. (4)). However, one may wonder whether the arithmetic mean is always the optimal choice for any given task. In principle, any weighted mean can be used as a representative ε-equivalent factor, as shown in the next theorem.

Theorem 15. *Let $\boldsymbol{G} = \{\phi_1, \ldots, \phi_k\}$ denote a group of pairwise ε-equivalent factors and let $\phi_\omega(\boldsymbol{r}) = \sum_{i=1}^k \omega_i \phi_i(\boldsymbol{r})$ be the weighted mean with weights $\omega_i \geq 0$ and $\sum_{i=1}^k \omega_i = 1$ for all assignments $\boldsymbol{r}$. Then, $\boldsymbol{G}^* = \{\phi_1, \ldots, \phi_k, \phi_\omega\}$ is a group of pairwise ε-equivalent factors.*

Proof. We show the claim in two directions by proving that $\phi_\omega(\boldsymbol{r}) \in [\phi_i(\boldsymbol{r}) \cdot (1-\varepsilon), \phi_i(\boldsymbol{r}) \cdot (1+\varepsilon)]$ and $\phi_i(\boldsymbol{r}) \in [\phi_\omega(\boldsymbol{r}) \cdot (1-\varepsilon), \phi_\omega(\boldsymbol{r}) \cdot (1+\varepsilon)]$ hold for any assignment $\boldsymbol{r}$ and $\phi_i \in \boldsymbol{G}$.

For the first direction, let $\boldsymbol{r}$ be an arbitrary assignment and let $\phi_i \in \boldsymbol{G}$. As all factors in $\boldsymbol{G}$ are pairwise ε-equivalent, it holds that

$$\phi_i(\boldsymbol{r}) \cdot (1-\varepsilon) \leq \min_{\phi_j \in \boldsymbol{G}} \phi_j(\boldsymbol{r}) \tag{27}$$

$$= \min_{\phi_j \in \boldsymbol{G}} \phi_j(\boldsymbol{r}) \sum_{j=1}^k \omega_j = \sum_{j=1}^k \omega_j \cdot \min_{\phi_j \in \boldsymbol{G}} \phi_j(\boldsymbol{r}) \tag{28}$$

$$\leq \sum_{j=1}^k \omega_j \cdot \phi_j(\boldsymbol{r}) = \phi_\omega(\boldsymbol{r}) \tag{29}$$

and analogously also for the upper bound

$$\phi_i(\boldsymbol{r}) \cdot (1+\varepsilon) \geq \max_{\phi_j \in \boldsymbol{G}} \phi_j(\boldsymbol{r}) \tag{30}$$

$$= \max_{\phi_j \in \boldsymbol{G}} \phi_j(\boldsymbol{r}) \sum_{j=1}^k \omega_j = \sum_{j=1}^k \omega_j \cdot \max_{\phi_j \in \boldsymbol{G}} \phi_j(\boldsymbol{r}) \tag{31}$$

$$\geq \sum_{j=1}^k \omega_j \cdot \phi_j(\boldsymbol{r}) = \phi_\omega(\boldsymbol{r}). \tag{32}$$

For the second direction, it holds that for any assignment $\boldsymbol{r}$, every $\phi_i \in \boldsymbol{G}$ is contained in the interval $[\phi_j(\boldsymbol{r}) \cdot (1-\varepsilon), \phi_j(\boldsymbol{r}) \cdot (1+\varepsilon)]$ for any $j \in \{1, \ldots, k\}$. Therefore, we can get the lower bound for a specific $j \in \{1, \ldots, k\}$ and obtain

$$(1-\varepsilon)\phi_\omega(\boldsymbol{r}) = (1-\varepsilon) \sum_{i=1}^k \omega_i \cdot \phi_i(\boldsymbol{r}) \tag{33}$$

$$= \sum_{i=1}^k \omega_i \cdot (1-\varepsilon)\phi_i(\boldsymbol{r}) \tag{34}$$

$$\leq \sum_{i=1}^k \omega_i \phi_j(\boldsymbol{r}) = \phi_j(\boldsymbol{r}) \sum_{i=1}^k \omega_i = \phi_j(\boldsymbol{r}) \tag{35}$$

as well as the upper bound for a specific j:

$$(1+\varepsilon)\phi_\omega(\boldsymbol{r}) = (1+\varepsilon) \sum_{i=1}^k \omega_i \cdot \phi_i(\boldsymbol{r}) \tag{36}$$

$$= \sum_{i=1}^k \omega_i \cdot (1+\varepsilon)\phi_i(\boldsymbol{r}) \tag{37}$$

$$\geq \sum_{i=1}^{k} \omega_i \phi_j(\boldsymbol{r}) = \phi_j(\boldsymbol{r}) \sum_{i=1}^{k} \omega_i = \phi_j(\boldsymbol{r}). \tag{38}$$

Therefore, we get for every j the property $\phi_j(\boldsymbol{r}) \in [(1-\varepsilon)\phi_\omega(\boldsymbol{r}), (1+\varepsilon)\phi_\omega(\boldsymbol{r})]$. □

Geometrically, this property can be viewed as follows: The convex hull of a group of pairwise ε-equivalent factors ϕ_i consists entirely of points that are ε-equivalent to each other:

$$\operatorname{conv}(\phi_1, \ldots, \phi_m) := \left\{ \sum_{i=1}^{m} \omega_i \phi_i \;\middle|\; \omega_i \geq 0, \sum_{i=1}^{m} \omega_i = 1 \right\}. \tag{39}$$

Besides the arithmetic mean, this also includes the geometric mean of the original factors, which can most easily be shown by an appropriate choice of weights.

Corollary 16. *Let $\boldsymbol{G} = \{\phi_1, \ldots, \phi_m\}$ denote a group of pairwise ε-equivalent factors and let $\phi_{gm}(\boldsymbol{r}) := \sqrt[m]{\prod_{i=1}^{m} \phi_i(\boldsymbol{r})}$ be the geometric mean for all assignments $\boldsymbol{r}$. Then, $\boldsymbol{G}^* = \{\phi_1, \ldots, \phi_m, \phi_{gm}\}$ is a group of pairwise ε-equivalent factors.*

Proof. Since $\phi_{gm}(\boldsymbol{r}) \in [\min_{i=1,\ldots,m} \phi_i(\boldsymbol{r}), \max_{i=1,\ldots,m} \phi_i(\boldsymbol{r})]$, we can use

$$1 \geq \omega_1 := \frac{\phi_{gm}(\boldsymbol{r}) - \min_{i=1,\ldots,m} \phi_i(\boldsymbol{r})}{\max_{i=1,\ldots,m} \phi_i(\boldsymbol{r}) - \min_{i=1,\ldots,m} \phi_i(\boldsymbol{r})} \geq 0 \tag{40}$$

$$1 \geq \omega_2 := \frac{\max_{i=1,\ldots,m} \phi_i(\boldsymbol{r}) - \phi_{gm}(\boldsymbol{r})}{\max_{i=1,\ldots,m} \phi_i(\boldsymbol{r}) - \min_{i=1,\ldots,m} \phi_i(\boldsymbol{r})} \geq 0, \tag{41}$$

which summarises to $\omega_1 + \omega_2 = 1$ and with Theorem 15, the proof can be completed:

$$\phi_{gm}(\boldsymbol{r}) = \omega_1 \cdot \min_{i=1,\ldots,m} \phi_i(\boldsymbol{r}) + \omega_2 \cdot \max_{i=1,\ldots,m} \phi_i(\boldsymbol{r}).$$ □

Note 1. *For positive real numbers $\phi_1(\boldsymbol{r}), \ldots, \phi_m(\boldsymbol{r})$ the geometric mean is always smaller or equal to the arithmetic mean*

$$\phi_{gm}(\boldsymbol{r}) = \sqrt[m]{\prod_{i=1}^{m} \phi_i(\boldsymbol{r})} \leq \frac{1}{k} \sum_{k=1}^{m} \phi_k(\boldsymbol{r}) = \phi^*(\boldsymbol{r}), \tag{42}$$

and more robust against outliers.

Results previously shown in this subsection required no $\mathcal{L}^p$ normalisation, assuming that the weighted average localises on a spherical shell.

Lemma 17. *Let $\boldsymbol{G} = \{\phi_1, \ldots, \phi_m\}$ denote a group of pairwise ε-equivalent factors with $\phi_i \in S^p_{n-1}$ for $i = 1, \ldots, m$ and let $\phi_\omega(\boldsymbol{r}) = \sum_{i=1}^{m} \omega_i \phi_i(\boldsymbol{r})$ be the weighted mean with weights $\omega_i \geq 0$ and $\sum_{i=1}^{m} \omega_i = 1$ for all assignments $\boldsymbol{r}$. Then, the length of ϕ_ω is bounded by $\|\phi_\omega\|_p \in \left[\frac{1}{1+\varepsilon}, 1\right]$.*

In other words, $\phi_\omega \in \overline{B^p_1}(0) \setminus B^p_{1/(1+\varepsilon)}(0) = \left\{\phi \in \mathbb{R}^n : \frac{1}{1+\varepsilon} \leq \|\phi\|_p \leq 1\right\}$.

Proof. For $p \geq 1$ we can use the Jensen-inequality for the convex function $f_p(x) := |x|^p$ (also called *convexity of $\mathcal{L}^p$* norm, [33]) and get

$$\|\phi_\omega\|_p^p = \left\| \sum_{i=1}^{m} \omega_i \phi_i \right\|_p^p = \sum_{k=1}^{n} \left(\sum_{i=1}^{m} \omega_i \phi_i(k) \right)^p \tag{43}$$

$$\leq \sum_{k=1}^{n} \sum_{i=1}^{m} \omega_i \phi_i^p(k) = \sum_{i=1}^{m} \omega_i \sum_{k=1}^{n} \phi_i^p(k) \tag{44}$$

$$= \sum_{i=1}^{m} \omega_i \|\phi_i\|_p^p = \sum_{i=1}^{m} \omega_i = 1. \tag{45}$$

By the pairwise ε-equivalence of ϕ_ω, from Theorem 15, we get $\phi_\omega(k) \geq \max_{j=1,\dots,m}\{\phi_j(k)\} \,/\, (1+\varepsilon)$ for $k = 1, \dots, n$. We therefore obtain

$$\|\phi_\omega\|_p^p = \sum_{k=1}^{n} \left(\sum_{i=1}^{m} \omega_i \phi_i(k) \right)^p \tag{46}$$

$$\geq \sum_{k=1}^{n} \left(\sum_{i=1}^{m} \omega_i \max_{j=1,\dots,m} \{\phi_j(k)\} \,/\, (1+\varepsilon) \right)^p \tag{47}$$

$$= \sum_{k=1}^{n} \max_{j=1,\dots,m} \{\phi_j(k)\}^p \,/\, (1+\varepsilon)^p \left(\sum_{i=1}^{m} \omega_i \right)^p \tag{48}$$

$$= 1 \,/\, (1+\varepsilon)^p \sum_{k=1}^{n} \max_{j=1,\dots,m} \{\phi_j(k)\}^p \tag{49}$$

$$\geq 1 \,/\, (1+\varepsilon)^p \sum_{k=1}^{n} \phi_1^p(k) \tag{50}$$

$$= 1 \,/\, (1+\varepsilon)^p \|\phi_1\|_p^p = 1 \,/\, (1+\varepsilon)^p. \tag{51}$$

The last inequality is a reduction of the maximum to one arbitrary, but specific factor (in this case ϕ_1), to be able to take its scaling into account. Taking the p-th root results in the desired lower bound. In the final inequality, we reduce the maximum to a single factor ϕ_1, allowing us to use its unit property in $\mathcal{L}^p$. This completes the proof, as taking the p-th root yields the desired lower bound. □

Consequently, we can naturally combine the idea of the convex hull from Theorem 15 with the normalisation from Lemma 17 to obtain an intuitive visual perspective. Given a set of factors $\phi_1, \dots, \phi_k \in \mathbb{R}_+^n$, we normalise them to ensure fair starting conditions such that $\phi_i^{\text{norm}} := \phi_i \,/\, \|\phi_i\|_p \in \mathbb{R}_+^n \cap S_{n-1}^p$, i.e., each lies on the positive orthant of the p-unit n-sphere. If they are pairwise ε-equivalent, then any weighted mean in their convex hull remains pairwise ε-equivalent to them. Additionally, the $\mathcal{L}^p$ norm of any such mean lies between $\frac{1}{1+\varepsilon}$ and 1. However, this norm is in general strictly less than 1 (except boundary factors), implying that normalisation does not preserve the normalised property under convex combinations. This leads to the subsequent observation, which prevents the geometric interpretation from being further simplified to a convex cone.

Note 2. *Let $C := \text{cone}(G) = \{\phi \in \mathbb{R}_+^n : \phi = \sum_{i=1}^m \omega_i \phi_i$ with $\omega_i \geq 0, \phi_i \in G\}$ be the finitely generated convex cone from a group of factors $G = \{\phi_1, \dots, \phi_m\}$, whose normalised set $G^{norm} := \{\phi_i^{norm} = \phi_i \,/\, \|\phi_i\|_p, i = 1, \dots, m\}$ is pairwise ε-equivalent, contains normalised factors $\phi^{norm} \in C$ with $\|\phi^{norm}\|_p = 1$ that are not pairwise ε-equivalent to all elements of G^{norm}. Formally, there exists $\phi^{norm} \in C$ with $\|\phi^{norm}\|_p = 1$ such that for some $i \in \{1, \dots, m\}$ we have $\phi^{norm} \neq_\varepsilon \phi_i^{norm}$. Example 5 provides a counterexample that invalidates this geometric interpretation in general.*

Example 5. *When normalising any factor within the generated cone, it is not necessarily ε-equivalent to the original generating normalised factors. This is illustrated by the following example with*

$$\phi_1 = \begin{pmatrix} 0.10 \\ 1.0 \\ 1.0 \end{pmatrix}, \phi_2 = \begin{pmatrix} 0.11 \\ 1.05 \\ 1.0 \end{pmatrix}, \phi_3 = \begin{pmatrix} 0.11 \\ 1.0 \\ 1.05 \end{pmatrix}. \tag{52}$$

We begin by determining the ε-equivalence of the original factors:

$$d_\infty(\phi_1, \phi_2) = 0.1 = d_\infty(\phi_1, \phi_3) \text{ and } d_\infty(\phi_2, \phi_3) = 0.05. \tag{53}$$

After normalisation, for instance using the $\mathcal{L}^2$-norm via $\phi_i^{norm} := \phi_i \,/\, \|\phi_i\|_2$ for all i, we obtain:

$$d_\infty(\phi_1^{norm}, \phi_2^{norm}) = 0.07244888 = d_\infty(\phi_1^{norm}, \phi_3^{norm}) \text{ and } d_\infty(\phi_2^{norm}, \phi_3^{norm}) = 0.05. \tag{54}$$

Due to Def. 4, we have pairwise ε-equivalence for

$$\varepsilon_0 := \max_{i,j\in\{1,2,3\}} d_\infty(\phi_i^{norm}, \phi_j^{norm}) = 0.07244888 \tag{55}$$

for this triple of normalised factors. For the convex combination $\phi_\omega := 0 \cdot \phi_1^{norm} + 0.1 \cdot \phi_2^{norm} + 0.9 \cdot \phi_3^{norm}$, pairwise ε_0-equivalence still holds (see also Theorem 15) with $d_\infty(\phi_\omega, \phi_i^{norm}) \leq \varepsilon_0$ for $i = 1, 2, 3$. However, the normalised version $\phi_\omega^{norm} := \phi_\omega / \|\phi_\omega\|_2$ has $d_\infty(\phi_\omega^{norm}, \phi_1^{norm}) = 0.07256301 > \varepsilon_0$ and is in particular not ε_0-equivalent to ϕ_1^{norm} and automatically not pairwise ε-equivalent to the normalised group anymore.

This example shows that the original idea of finding symmetries within the ACP framework can be extended from exact symmetries to normalised and also to ε-equivalent approximate symmetries. However, the latter cannot be interpreted as a finitely generated convex cone. Consequently, scaling assumes a more profound role than previously anticipated, particularly in determining a unique order in the hierarchical version HACP, making its choice significant. Nonetheless, the proven results offer positive prospects by enabling extensions of the ε-ACP and HACP algorithms via alternative loss functions that permit different optimal representatives for a pairwise ε-equivalent group through weighted mean selection. In addition, the convex hull of a group of pairwise ε-equivalent factors is the smallest set containing these weighted elements. If it can be shown that a new factor is also pairwise ε-equivalent to each factor of the given group, the new convex hull is a superset and extends the possible representatives while retaining all previous ones.

3.3. Special Case: Euclidean Perspective

We now focus on the geometric interpretation in the Euclidean space $\mathcal{L}^2$ as a special case of Section 3.1. Although the counterexample from Example 5 still applies, this setting has advantages. In particular, it allows a more interpretable and intuitive understanding of ε-equivalence due to the accessibility of Euclidean norms and angles, which provide a one-dimensional comparison similar to 1DEED; see also [15, 31] for related discussions.

Definition 5. *The Cosine distance $D_{\cos}$ for two non-zero vectors $\phi_1, \phi_2 \in \mathbb{R}^n$ is defined as*

$$D_{\cos}(\phi_1, \phi_2) := 1 - \frac{\phi_1 \cdot \phi_2}{||\phi_1||_2 \cdot ||\phi_2||_2} \tag{56}$$

$$= 1 - \frac{\sum_{i=1}^n \phi_1(i) \cdot \phi_2(i)}{\sqrt{\sum_{i=1}^n \phi_1^2(i)} \cdot \sqrt{\sum_{i=1}^n \phi_2^2(i)}}, \tag{57}$$

which is equal to $1 - \cos(\theta)$ for one $\theta \in [0, \pi]$.

In [31], a refined definition of the Cosine distance for factors is given, which is based more precisely on the assignments $\boldsymbol{r}$. It also includes the notion of *exchangeable* factors, i.e., two factors that are identical after scaling and permutation of potentials.

Theorem 18 (Theorem 2, [31])**.** *Let $\phi_1(R_1, \ldots, R_n)$ and $\phi_2(R_1, \ldots, R_n)$ denote two factors. If ϕ_1 and ϕ_2 are exchangeable, then it holds that $D_{\cos}(\phi_1, \phi_2) = 0$.*

When normalising the factors in $\mathcal{L}^2$, the Cosine distance satisfies the following well-known property.

Lemma 19. *The Cosine distance for two normalised factors $\phi_1, \phi_2 \in \mathbb{R}^n_{>0}$ and the Euclidean distance for two normalised factors $\phi_1, \phi_2 \in \mathbb{R}^n_{>0}$ with $\|\phi_i\|_2 = 1$ for $i = 1, 2$ are the same besides a scaling factor, i.e.,*

$$2D_{\cos}(\phi_1, \phi_2) = ||\phi_1 - \phi_2||_2. \tag{58}$$

Proof. With the normalised vectors $||\phi_i||_2 = 1$ for $i = 1, 2$, consider

$$
\begin{aligned}
||\phi_1 - \phi_2||_2 &= (\phi_1 - \phi_2)(\phi_1 - \phi_2) && (59)\\
&= ||\phi_1||_2 + ||\phi_2||_2 - 2(\phi_1 \cdot \phi_2) && (60)\\
&= 2(1 - (\phi_1 \cdot \phi_2)) && (61)\\
&= 2\left(1 - \frac{\phi_1 \cdot \phi_2}{||\phi_1||_2 \cdot ||\phi_2||_2}\right) && (62)\\
&= 2D_{\cos}(\phi_1, \phi_2). && \square
\end{aligned}
$$

As the Cosine distance for $\varepsilon > 0$ can be seen as an angle between two factors, we obtain a bound on the maximal angle difference between ε-equivalent factors.

Lemma 20. *The angle $\theta \in [0, \frac{\pi}{2}]$ of two ε-equivalent factors $\phi_1, \phi_2 \in \mathbb{R}^n_{>0} \cap S^2_{n-1}$ is at most*

$$
\theta(\varepsilon) \leq \begin{cases} \arccos(1 - \frac{\varepsilon}{2}) & \text{for } \varepsilon \leq 2 \\ \frac{\pi}{2} & \text{for } \varepsilon > 2. \end{cases} \tag{63}
$$

Proof. The angle of two vectors within an orthant in the Eucledian space is always maximal $\frac{\pi}{2}$, which will be the upper bound for $\varepsilon > 2$. For the other case of $0 \leq \varepsilon \leq 2$ we get from Lemma 11 (i) and Lemma 19 the inequality:

$$
\begin{aligned}
\frac{\varepsilon}{2} &\geq \frac{\|\phi_1 - \phi_2\|_2}{2} = D_{\cos}(\phi_1, \phi_2) = 1 - \cos(\theta) && (64)\\
\Leftrightarrow \cos(\theta) &\geq 1 - \frac{\varepsilon}{2} && (65)\\
\Leftrightarrow \quad \theta(\varepsilon) &\leq \arccos\left(1 - \frac{\varepsilon}{2}\right). && (66)
\end{aligned}
$$

In the last step, taking the strictly monotonically decreasing $\arccos$ for values in $[0, 1]$ or $\varepsilon \in [0, 2]$ on both sides switches the direction of Eq. (66). The operation remains well-defined if $\varepsilon \leq 2$. $\square$

4. Discussion

Understanding of ε-equivalence. As an extension of exchangeable factors, the general concept of ε-equivalence and the more manageable 1DEED seem to be a promising option to extend lifted inference to a new level via approximation of *similar* factors. Therefore, the understanding of *similarity* in this context is key. Building on the maximum metric (Chebyshev distance), 1DEED takes relative lengths into account, resulting in the lack of transitivity. However, the property of ε-equivalence is easy to check and leads to multiple practical properties including the guaranteed asymptotic bounds of Section 2.2 and a compressed (compact) and more interpretable model depending on the choice of ε. The hierarchical approach additionally helps to understand the complexity of the given FG by pre-analysing the 1DEED of all factors.

Geometric interpretability. Understanding differences between factors as an angle between vectors within the Euclidean space is not a new concept [31]. However, it is one way to interpret the concept of ε-equivalence. Pairwise ε-equivalent factors can also be seen as set of n-dimensional vectors, which bound and generate a compact set (the convex hull) within the $\mathcal{L}^p$ space. This geometric object is basically the minimal set containing the group of ε-equivalent factors. If another factor that is pairwise ε-equivalent to all factors in a group is added to the group, the resulting set becomes a superset and its newly generated convex hull is again a set of pairwise ε-equivalent factors containing the old set.

New potential for future exploration. Using weighted means as representatives for groups of ε-equivalent factors enriches the possibilities of how the algorithms ε-ACP and HACP can be applied. The usage of weighted means opens the whole concept of approximate lifted model construction based on ε-equivalence to possible robustifications and generalisations by changing the original squared loss function and its optimal choice of the arithmetic mean as representatives for groups of ε-equivalent factors. Choosing different loss functions and their new corresponding optimal solution of a representative increases the flexibility of the ε-ACP algorithm and the HACP algorithm.

5. Conclusion

We presented fundamental properties of ε-equivalence related to lifted model construction under a unified view and have proven several additional properties for the $\mathcal{L}^p$ space. By viewing the concept of ε-equivalence from geometric perspective in $\mathcal{L}^p$, we provided a solid foundation for geometric interpretability and gave a new perspective on the concept of pairwise ε-equivalence via the convex hull of groups of ε-equivalent factors. Our geometric interpretation opens up for advancements with respect to the ε-ACP algorithm and its hierarchical version HACP in terms of generalisation and robustification of the previously introduced compression techniques.

Acknowledgments

This work was partially funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia. The research of Malte Luttermann was funded by the BMBF project AnoMed 16KISA057.

Declaration on Generative AI

The authors have not employed any generative AI tools.

References

[1] D. Poole, First-order Probabilistic Inference, in: Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-2003), IJCAI Organization, 2003, pp. 985–991.
[2] M. Richardson, P. Domingos, Markov Logic Networks, Machine Learning 62 (2006) 107–136.
[3] N. Fuhr, Probabilistic Datalog - A Logic for Powerful Retrieval Methods, in: SIGIR-95 Proc. of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 1995, pp. 282–290.
[4] T. Sato, A Statistical Learning Method for Logic Programs with Distribution Semantics, in: Proc. of the 12th International Conference on Logic Programming, MIT Press, 1995, pp. 715–729.
[5] B. Ahmadi, K. Kersting, M. Mladenov, S. Natarajan, Exploiting Symmetries for Scaling Loopy Belief Propagation and Relational Training, Machine Learning 92 (2013) 91–132.
[6] T. Braun, R. Möller, Lifted Junction Tree Algorithm, in: Proceedings of the Thirty-Ninth German Conference on Artificial Intelligence (KI-2016), Springer, 2016, pp. 30–42.
[7] V. Gogate, P. Domingos, Probabilistic Theorem Proving, in: Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI-2011), AUAI Press, 2011, pp. 256–265.
[8] M. Hartwig, R. Möller, T. Braun, An Extended View on Lifting Gaussian Bayesian Networks, Artificial Intelligence 330 (2024) 104082.
[9] N. Taghipour, D. Fierens, J. Davis, H. Blockeel, Lifted Variable Elimination: Decoupling the Operators from the Constraint Language, Journal of Artificial Intelligence Research 47 (2013) 393–439.

[10] G. Van den Broeck, N. Taghipour, W. Meert, J. Davis, L. De Raedt, Lifted Probabilistic Inference by First-order Knowledge Compilation, in: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI-2011), IJCAI Organization, 2011, pp. 2178–2185.
[11] T. Braun, R. Möller, Parameterised Queries and Lifted Query Answering, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-2018), IJCAI Organization, 2018, pp. 4980–4986.
[12] G. Van den Broeck, Lifted Inference and Learning in Statistical Relational Models, Ph.D. thesis, KU Leuven, 2013.
[13] G. Van den Broeck, J. Davis, Conditioning in First-order Knowledge Compilation and Lifted Probabilistic Inference, in: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI-2012), AAAI Press, 2012, pp. 1961–1967.
[14] M. Gehrke, T. Braun, R. Möller, Relational Forward Backward Algorithm for Multiple Queries, in: FLAIRS-19 Proc. of the 32nd International Florida Artificial Intelligence Research Society Conference, AAAI Press, 2019, pp. 464–469.
[15] M. Gehrke, T. Braun, R. Möller, Taming Reasoning in Temporal Probabilistic Relational Models, in: 9th International Workshop on Statistical Relational AI at the 34th AAAI Conference on Artificial Intelligence, 2020.
[16] U. Apsel, R. I. Brafman, Extended Lifted Inference with Joint Formulas, in: UAI-11 Proc. of the 27th Conference on Uncertainty in Artificial Intelligence, AUAI Press, 2011, pp. 74–83.
[17] A. Nath, P. Domingos, A Language for Relational Decision Theory, in: Proc. of the 6th International Workshop on Statistical Relational Learning, 2009.
[18] M. Gehrke, T. Braun, R. Möller, Lifted Temporal Maximum Expected Utility, in: Proc. of the 32nd Canadian Conference on Artificial Intelligence, Canadian AI 2019, Springer, 2019, pp. 380–386.
[19] M. Gehrke, T. Braun, S. Polovina, Restricting the Maximum Number of Actions for Decision Support under Uncertaitny, in: ICCS-20 Proc. of the 25th International Conference on Conceptual Structures, Springer, 2020, pp. 145–160.
[20] S. Sanner, K. Kersting, Symbolic Dynamic Programming for First-order POMDPs, in: AAAI-10 Proc. of the 24th AAAI Conference on Artificial Intelligence, AAAI Press, 2010, pp. 1140–1146.
[21] T. Braun, M. Gehrke, F. Lau, R. Möller, Lifting in Multi-agent Systems under Uncertainty, in: UAI-22 Proc. of the 38th Conference on Uncertainty in Artificial Intelligence, AUAI Press, 2022, pp. 1–8.
[22] N. N. Karabulut, T. Braun, Lifting Partially Observable Stochastic Games, in: SUM-24 Proceedings of the 16th International Conference on Scalable Uncertainty Management, Springer, 2024, pp. 201–216.
[23] M. Niepert, G. Van den Broeck, Tractability through Exchangeability: A New Perspective on Efficient Probabilistic Inference, in: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI-2014), AAAI Press, 2014, pp. 2467–2475.
[24] B. Weisfeiler, A. Leman, A Reduction of a Graph to a Canonical Form and Algebra Arising during this Reduction, Nauchno-Technicheskaya Informatsia 2 (1968) 12–16. Translated from Russian.
[25] M. Luttermann, T. Braun, R. Möller, M. Gehrke, Colour Passing Revisited: Lifted Model Construction with Commutative Factors, in: Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-2024), AAAI Press, 2024, pp. 20500–20507.
[26] M. Luttermann, J. Machemer, M. Gehrke, Efficient Detection of Exchangeable Factors in Factor Graphs, in: FLAIRS-24 Proceedings of the 37th International FLAIRS Conference, volume 37, Florida Online Journals, 2024.
[27] M. Luttermann, J. Machemer, M. Gehrke, Efficient Detection of Commutative Factors in Factor Graphs, in: PGM-24 Proceedings of the 12th International Conference on Probabilistic Graphical Models, volume 246, PMLR, 2024, pp. 38–56.
[28] M. Luttermann, R. Möller, M. Gehrke, Lifted Model Construction without Normalisation: A Vectorised Approach to Exploit Symmetries in Factor Graphs, in: Proceedings of the Third Learning on Graphs Conference (LoG-2024), PMLR, 2025, pp. 46:1–46:17.
[29] M. Luttermann, J. Speller, M. Gehrke, T. Braun, R. Möller, M. Hartwig, Approximate Lifted Model

Construction, in: Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-2025), 2025. Https://arxiv.org/abs/2504.20784.
[30] J. Speller, M. Luttermann, M. Gehrke, T. Braun, Compression versus Accuracy: A Hierarchy of Lifted Models, 2025. Https://arxiv.org/abs/2505.22288.
[31] M. Luttermann, R. Möller, M. Gehrke, Lifted Model Construction without Normalisation: A Vectorised Approach to Exploit Symmetries in Factor Graphs, in: Proceedings of the Third Learning on Graphs Conference (LoG-2024), PMLR, 2025, pp. 46:1–46:17.
[32] H. Chan, A. Darwiche, A Distance Measure for Bounding Probabilistic Belief Change, International Journal of Approximate Reasoning 38 (2005) 149–174.
[33] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, Acta mathematica 30 (1906) 175–193.