Lifted Model Construction under Approximate Commutativity

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

[1] Institute for Humanities-Centered Artificial Intelligence, University of Hamburg
{malte.luttermann,marcel.gehrke,ralf.moeller}@uni-hamburg.de
[2] Data Science Group, University of Münster
{jan.speller,tanya.braun}@uni-muenster.de

Abstract. Lifted inference algorithms enable scalable probabilistic inference even for large object domains by leveraging the indistinguishability of objects in a probability distribution. An essential prerequisite for constructing a lifted representation is to identify commutative factors, i.e., functions whose output values are invariant under permutations of a subset of their input values, in a potential-based factorisation. In practice, however, parameters learned from data inevitably deviate even if associated objects are indistinguishable, causing their corresponding factors to be only approximately commutative instead of being exactly commutative. We address this problem by introducing the concept of ε-commutativity, a relaxation of commutativity where output values are only approximately invariant under permutations of input values. Specifically, we show how ε-commutativity can be exploited for lifted model construction, downstream probabilistic inference, and prove strict bounds on the induced approximation error, thereby ensuring the practical applicability of lifted model construction while maintaining highly accurate query results. These theoretical guarantees are confirmed empirically, demonstrating comparable query accuracy at lower runtime.

1 Introduction

Uncertain reasoning is a fundamental task in artificial intelligence, where uncertainty is commonly represented through probability distributions over random variables. Probabilistic graphical models such as factor graphs (FGs) provide a compact way to encode joint distributions over a set of random variables (randvars) by decomposing a distribution into products of local functions, called factors. Even with a factorised representation, the problem of probabilistic inference (i.e., computing marginal or conditional distributions of randvars) is intractable in general as the computational cost grows exponentially with the number of randvars [3]. To allow for tractable probabilistic inference whose complexity grows polynomially in the number of objects in the underlying domains, lifted inference algorithms exploit the indistinguishability of objects by using a

[*] These authors contributed equally to this work.

representative for computations [13]. Performing lifted inference, however, requires a lifted representation in which indistinguishable objects are grouped together. The construction of such a lifted representation includes identifying factors that exhibit symmetries and can therefore be grouped together. One important class are *commutative factors* (that is, factors whose output values are invariant under permutations of a subset of their input values). Whether a factor is commutative depends on the precise output values it maps its input values to, and herein lies a subtle but pervasive practical issue: When numbers are estimated from observational data, two values that should ideally coincide hardly ever do so up to the last digit. Even in simple cases such as counted data, learning can always yield a distribution of values, which makes an otherwise commutative factor fail the strict definition of commutativity, even though the commutativity is clearly present from a modelling perspective. Current state-of-the-art algorithms for detecting commutative factors have no means to tolerate such discrepancies. In this paper, we address this problem by softening the underlying equality requirement, thereby enabling the detection of approximate commutativity and allowing for further compression of FGs.

Previous work. When dealing with noisy distributions, typical approximate techniques control the deviation on the level of the full (global) probability distribution based on classes of prior distributions [1], as done, e.g., by distortion models such as the constant odds ratio model, which bounds the multiplicative deviation from a nominal reference distribution [12]. In contrast, we consider a factorised distribution and bound individual potentials directly via a local constant-odds-ratio-type neighbourhood on the level of a single factor. Structurally, however, the overall global approximation follows the same guarantees as distortion models. There has also been research on *robustness* in graph learning against false measurements and outliers [6,14], whereas our approach relaxes exact structural symmetry in factors for lifted inference. The line of work on lifted probabilistic inference goes back to Poole [15], who introduced parametric factor graphs (PFGs) as a representation that interleaves first-order logic with probabilistic modelling and lifted variable elimination (LVE) as an inference algorithm operating on PFGs. A substantial body of subsequent research has refined and extended LVE in various directions [18]. One of these refinements includes the introduction of so-called counting randvars (CRVs) [11], which count over the occurrences of range values in the input of a factor instead of explicitly listing possible combinations of input values, thereby exploiting the commutativity of factors. The current state-of-the-art to construct a lifted representation such as a PFG is the advanced colour passing (ACP) algorithm [7]. A core ingredient of ACP is to determine, for every factor, a maximum sized subset of commutative arguments such that factors can be compressed by using CRVs to count over the commutative arguments. To determine maximum sized subsets of commutative arguments, the detection of commutative factors (DECOR) algorithm [8,9] is the state-of-the-art. Such lifted model construction techniques have so far focused on approximate factors rather than approximate commutativity [10,16].

Our contributions. We introduce ε-*commutativity*, an approximate variant of commutativity in which output values are allowed to deviate depending on a hyperparameter ε. On the basis of ε-commutativity, we show how approximately commutative factors can be compressed by replacing groups of similar values with representative values, thereby simplifying the underlying distribution by giving events of a symmetry set the same value. We further establish a strict bound on the approximation error induced by such a compression, guaranteeing that subsequent inference on the resulting lifted representation remains accurate. Finally, we empirically demonstrate that the approximation error remains well below the theoretical bounds in practice, while the proposed algorithm substantially reduces runtime across a range of parameter settings.

Structure of this paper. The remainder of the paper is organised as follows. We first introduce the necessary background on FGs and commutative factors therein. We then formalise the concept of ε-commutativity, develop a compression scheme based on ε-commutative representatives, and prove important properties leading to a strict bound on the resulting approximation error. Afterwards, we embed the compression of ε-commutative factors into the ACP algorithm for lifted model construction and empirically show that downstream probabilistic inference still produces highly accurate results. Due to space constraints, detailed proofs are relegated to App. A and further experimental results to App. B.

2 Background

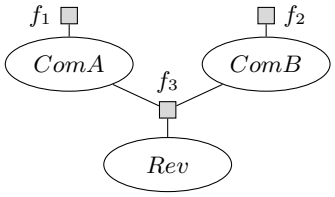
We briefly introduce the notion of an FG and then formally define what a commutative factor is. An FG compactly encodes a joint probability distribution over randvars as a product of factors [4,5]. In the following, we write $\mathrm{range}(R)$ to denote the values a randvar R can take (called range) and $\mathcal{X}_{\boldsymbol{R}} := \times_{R \in \boldsymbol{R}} \mathrm{range}(R)$ to denote the Cartesian product over the ranges of a set of randvars $\boldsymbol{R}$.

Definition 1 (Factor Graph). *An* FG *is a tuple* $M = (\boldsymbol{R}, \boldsymbol{F}, \boldsymbol{E}, \boldsymbol{\Phi})$ *where* $\boldsymbol{R} = \{R_1, \ldots, R_p\}$ *is a set of randvars,* $\boldsymbol{F} = \{f_1, \ldots, f_m\}$ *is a set of factor nodes, and* $\boldsymbol{\Phi} = \{\phi_1, \ldots, \phi_m\}$ *is a set of function definitions (called factors). Each* $\phi_j(\boldsymbol{R}_{\phi_j}) \in \boldsymbol{\Phi}$ *defines a function* $\phi_j \colon \mathcal{X}_{\boldsymbol{R}_{\phi_j}} \to \mathbb{R}_{\geq 0}$ *over the Cartesian product of the ranges of its argument sequence of randvars* $\boldsymbol{R}_{\phi_j} \subseteq \boldsymbol{R}$*. The image of* ϕ_j *consists of non-negative real numbers (potentials), and each factor is non-trivial in the sense that it attains at least one positive value. For each pair of variable node* $R_i \in \boldsymbol{R}$ *and factor node* $f_j \in \boldsymbol{F}$*, there is an edge* $\{R_i, f_j\} \in \boldsymbol{E} \subseteq \{\{R, f\} \mid R \in \boldsymbol{R} \wedge f \in \boldsymbol{F}\}$ *if* $R_i \in \boldsymbol{R}_{\phi_j}$*. The full joint probability distribution encoded by* M *for an assignment* $(R_1 = r_1, \ldots, R_p = r_p)$ *to the randvars in* $\boldsymbol{R}$*, abbreviated as* $\boldsymbol{R} = \boldsymbol{r}$*, is the normalised product over all factors in* M*:*

$$P_M(\boldsymbol{R} = \boldsymbol{r}) = \frac{1}{Z}\psi(\boldsymbol{r}) = \frac{1}{Z}\prod_{j=1}^{m} \phi_j(\boldsymbol{R}_{\phi_j} = \boldsymbol{r}_j),$$

where $\boldsymbol{r}_j$ *is a projection of* $\boldsymbol{r}$ *to the argument list of* ϕ_j *and* Z *is the normalisation constant, defined as* $Z = \sum_{\boldsymbol{r} \in \mathcal{X}_{\boldsymbol{R}}} \prod_{j=1}^{m} \phi_j(\boldsymbol{R}_{\phi_j} = \boldsymbol{r}_j)$.

Example 1. Fig. 1a shows an FG M modelling the interplay between the competences $ComA$ and $ComB$ of two employees *Alice* and *Bob*, respectively, and the revenue Rev of the company they work for. The ranges of the randvars are $\text{range}(ComA) = \text{range}(ComB) = \text{range}(Rev) = \{\text{high}, \text{low}\}$ and Fig. 1b shows the function definition of the factor $\phi_3(ComA, ComB, Rev)$, illustrated as a potential table, i.e., $\phi_3(ComA = \text{high}, ComB = \text{high}, Rev = \text{high}) = \varphi_1$, and so on, such that $\varphi_1, \ldots, \varphi_6 \in \mathbb{R}_{\geq 0}$ are non-negative real numbers with $\varphi_i \neq \varphi_j$ for all $i \neq j$. We omit the potential tables of $\phi_1(ComA)$ and $\phi_2(ComB)$ for brevity.



(a)

$ComA$	$ComB$	Rev	$\phi_3(ComA, ComB, Rev)$
high	high	high	φ_1
high	high	low	φ_2
high	low	high	φ_3
high	low	low	φ_4
low	high	high	φ_3
low	high	low	φ_4
low	low	high	φ_5
low	low	low	φ_6

(b)

Fig. 1: (a) An FG modelling the interplay between the competences of two employees and the revenue of the company they work for. (b) The function definition of ϕ_3, where $\varphi_1, \ldots, \varphi_6 \in \mathbb{R}_{\geq 0}$ with $\varphi_i \neq \varphi_j$ for all $i \neq j$.

Ex. 1 illustrates that the output of $\phi_3(ComA, ComB, Rev)$ depends only on the number of highly competent employees, not on their identities. For instance, $\phi_3(ComA = \text{high}, ComB = \text{low}, Rev = \text{high}) = \phi_3(ComA = \text{low}, ComB = \text{high}, Rev = \text{high}) = \varphi_3$. Thus, permuting $ComA$ and $ComB$ leaves the output unchanged and only the count of competences set to high matters. This invariance under permutations is a key property exploited by lifted inference algorithms and motivates the following formal definition of commutativity.

Definition 2 (Commutative Factor [7]). *Let $\phi(R_1, \ldots, R_n)$ denote a factor and let $\boldsymbol{C}_\phi \subseteq \boldsymbol{R}_\phi := \{R_1, \ldots, R_n\}$ be a subset of ϕ's arguments with $|\boldsymbol{C}_\phi| \geq 2$. Then, ϕ is* commutative *with respect to $\boldsymbol{C}_\phi$ if for all assignments $\boldsymbol{r} = (r_1, \ldots, r_n) \in \mathcal{X}_{\boldsymbol{R}_\phi}$ it holds that $\phi(r_1, \ldots, r_n) = \phi(r_{\pi(1)}, \ldots, r_{\pi(n)})$ for all permutations π of $\{1, \ldots, n\}$ with $\pi(i) = i$ for all $R_i \notin \boldsymbol{C}_\phi$.*

We refer to the arguments in $\boldsymbol{C}_\phi$ as commutative arguments. Note that arguments can only be commutative if they have the same range because only arguments with identical ranges can be permuted. Thus, when searching for commutative arguments, only subsets with the same range need to be considered. Consider again $\phi_3(ComA, ComB, Rev)$ from Fig. 1b. According to Def. 2,

ϕ_3 is commutative with respect to $\{ComA, ComB\}$. The above definition relies on exact equality of potentials. In practice, however, small estimation errors often break this equality even when the underlying symmetry is still present. The next section therefore introduces approximate commutativity.

3 Approximate Commutativity

So far, the definition of a commutative factor requires strict equality among potentials, which we relax here, introducing the notion of ε-commutativity based on the idea of ε-equivalence [10], establish its fundamental properties, and derive the bounds needed for its practical application in lifted model construction.

We begin to introduce the more general notion of *ε-equivalence* between two potentials $\varphi_1, \varphi_2 \in \mathbb{R}_{\geq 0}$, written as $\varphi_1 =_\varepsilon \varphi_2$, which holds if

$$\varphi_1 \in [\varphi_2 \cdot (1-\varepsilon), \varphi_2 \cdot (1+\varepsilon)] \text{ and } \varphi_2 \in [\varphi_1 \cdot (1-\varepsilon), \varphi_1 \cdot (1+\varepsilon)]. \quad (1)$$

Two factors are *ε-equivalent* if there exists a permutation of arguments such that Eq. (1) holds for all pairwise row comparisons of potentials. Combining ε-equivalence and commutativity naturally leads to the notion of ε-commutativity.

Definition 3 (ε-Commutative Factor). *Let $\phi(R_1, \ldots, R_n)$ denote a factor, let $\boldsymbol{C}_\phi \subseteq \boldsymbol{R}_\phi$ be a subset of its arguments with $|\boldsymbol{C}_\phi| \geq 2$, and let $\varepsilon \in [0, 1]$. Then, ϕ is* ε-commutative *with respect to $\boldsymbol{C}_\phi$ if for all assignments $(r_1, \ldots, r_n) \in \mathcal{X}_{\boldsymbol{R}_\phi}$ and all permutations*

$$\pi \in \Pi_{\boldsymbol{C}_\phi} := \{\pi \in S_n \mid \pi(i) = i \text{ for all } R_i \notin \boldsymbol{C}_\phi\},$$

where S_n is the symmetric group over $\{1, \ldots, n\}$, the potentials $\phi(r_1, \ldots, r_n)$ and $\phi(r_{\pi(1)}, \ldots, r_{\pi(n)})$ are ε-equivalent, i.e.,

$$\phi(r_1, \ldots, r_n) \in [\phi(r_{\pi(1)}, \ldots, r_{\pi(n)}) \cdot (1-\varepsilon), \phi(r_{\pi(1)}, \ldots, r_{\pi(n)}) \cdot (1+\varepsilon)] \text{ and}$$
$$\phi(r_{\pi(1)}, \ldots, r_{\pi(n)}) \in [\phi(r_1, \ldots, r_n) \cdot (1-\varepsilon), \phi(r_1, \ldots, r_n) \cdot (1+\varepsilon)].$$

We refer to the arguments in $\boldsymbol{C}_\phi$ as ε-commutative arguments. Note that, for the choice of $\varepsilon = 0$, ε-commutativity reduces to strict commutativity.

As illustrated in Table 1b, the commutativity of a factor can be exploited to obtain a more compact representation of the factor by introducing a CRV that groups the commutative arguments such that the original potentials are replaced by aggregated potentials. Here, $ComA$ and $ComB$ are not listed individually but have been replaced by a CRV $\#_E[Com(E)]$ that counts the number of employees with a certain competence (i.e., $[2, 0]$ stands for two employees with a high competence and none with a low competence, $[1, 1]$ stands for one employee with a high competence and one with a low competence, and $[0, 2]$ stands for two employees with a low competence and none with a high competence). Given a commutative factor such as ϕ_3 from Fig. 1b, all assignments that differ only by a permutation of the range values of the commutative arguments share the

same potential, so choosing $\varphi_i^* = \varphi_i$ preserves the semantics exactly. However, for an ε-commutative factor such as ϕ from Table 1a, potentials corresponding to assignments only differ by a permutation on the commutative arguments may deviate by a factor of up to $(1 \pm \varepsilon)$, and hence collapsing them into a single potential φ_i^* inevitably discards information (e.g., φ_3^* must be chosen to represent both φ_3 and φ_3' at the same time). We therefore seek to choose φ_i^* such that its deviation from the original potentials is minimised.

$ComA$	$ComB$	Rev	$\phi(ComA, ComB, Rev)$
high	high	high	φ_1
high	high	low	φ_2
high	low	high	φ_3
high	low	low	φ_4
low	high	high	φ_3'
low	high	low	φ_4'
low	low	high	φ_5
low	low	low	φ_6

(a)

$\#_E[Com(E)]$	Rev	$\phi(\#_E[Com(E)], Rev)$
$[2, 0]$	high	φ_1^*
$[2, 0]$	low	φ_2^*
$[1, 1]$	high	φ_3^*
$[1, 1]$	low	φ_4^*
$[0, 2]$	high	φ_5^*
$[0, 2]$	low	φ_6^*

(b)

Table 1: (a) An exemplary factor $\phi(ComA, ComB, Rev)$ with ε-commutative arguments $\{ComA, ComB\}$ provided that φ_3 and φ_3' as well as φ_4 and φ_4' are ε-equivalent. (b) A compressed representation of $\phi(ComA, ComB, Rev)$ from (a) using a CRV to represent the ε-commutative arguments $\{ComA, ComB\}$.

Example 2. Consider the factor $\phi(ComA, ComB, Rev)$ from Table 1a. Given that φ_3 and φ_3' as well as φ_4 and φ_4' are ε-equivalent for some $\varepsilon \in [0, 1]$, ϕ is ε-commutative with respect to $\{ComA, ComB\}$.

In the following, we aim to find an optimal representative factor to compress an ε-commutative factor such that the underlying semantics is preserved as much as possible. To do so, we need a few theoretical properties induced by the notion of ε-commutativity. The following lemma, which has been proven only across different factors [10, Lemma 6], establishes a useful consequence of the symmetric definition of ε-commutativity and serves as a tool to derive subsequent bounds.

Lemma 1. *Let ϕ denote a factor that is ε-commutative with respect to a subset of its arguments $\boldsymbol{C}_\phi$. Then, for all permutations $\pi \in \Pi_{\boldsymbol{C}_\phi}$, it holds that*

$$\phi(r_1, \ldots, r_n) \in [\phi(r_{\pi(1)}, \ldots, r_{\pi(n)}) \cdot \frac{1}{1+\varepsilon}, \phi(r_{\pi(1)}, \ldots, r_{\pi(n)}) \cdot (1+\varepsilon)] \text{ and}$$
$$\phi(r_{\pi(1)}, \ldots, r_{\pi(n)}) \in [\phi(r_1, \ldots, r_n) \cdot \frac{1}{1+\varepsilon}, \phi(r_1, \ldots, r_n) \cdot (1+\varepsilon)].$$

The proof follows the same structural argument as [10, Lemma 6], using the symmetric definition of ε-commutativity. Details are given in App. A.

Before we turn our attention to the compression of an ε-commutative factor, we briefly state two important structural properties of ε-commutativity, which are algorithmically relevant when it comes to detecting ε-commutativity of a factor (which needs to be done before the compression can take place). A natural question is whether ε-commutativity with respect to a set is inherited by its subsets (as it is the case for the exact version of commutativity [9, Prop. 10]).

Proposition 1. *Let ϕ be ε-commutative with respect to $\boldsymbol{C}_\phi$. Then, ϕ is ε-commutative with respect to every subset $\boldsymbol{C}'_\phi \subseteq \boldsymbol{C}_\phi$ satisfying $|\boldsymbol{C}'_\phi| \geq 2$.*

Proof. Since, $\boldsymbol{C}'_\phi \subseteq \boldsymbol{C}_\phi$, we have $\Pi_{\boldsymbol{C}'_\phi} \subseteq \Pi_{\boldsymbol{C}_\phi}$. Hence, by Def. 3, for all $\pi \in \Pi_{\boldsymbol{C}_\phi}$ and assignments $\boldsymbol{r} = (r_1, \ldots, r_n)$, it holds that

$$\begin{aligned} \phi(\boldsymbol{r}) &\in [\phi(r_{\pi(1)}, \ldots, r_{\pi(n)}) \cdot (1-\varepsilon), \phi(r_{\pi(1)}, \ldots, r_{\pi(n)}) \cdot (1+\varepsilon)] \text{ and} \\ \phi(r_{\pi(1)}, \ldots, r_{\pi(n)}) &\in [\phi(\boldsymbol{r}) \cdot (1-\varepsilon), \phi(\boldsymbol{r}) \cdot (1+\varepsilon)]. \end{aligned}$$

Since $\Pi_{\boldsymbol{C}'_\phi} \subseteq \Pi_{\boldsymbol{C}_\phi}$, the same relations hold for all $\pi' \in \Pi_{\boldsymbol{C}'_\phi}$. □

Prop. 1 shows that ε-commutativity is preserved when moving from a set of randvars to its subsets. Hence, a factor that is commutative with respect to $\boldsymbol{C}'_\phi$ may be ε-commutative with respect to a strict superset $\boldsymbol{C}_\phi \supsetneq \boldsymbol{C}'_\phi$, thereby enabling further compression of the original factor. An important follow-up question is whether local ε-commutativity of smaller subsets implies this also for their union, which is indeed the case for exact commutativity [9, Theorem 11] and gives rise to an efficient algorithm to compute subsets of commutative arguments [9, Algorithm 2]. The following theorem shows that the corresponding property does not hold for ε-commutativity, implying that computing subsets of ε-commutative arguments is more sophisticated than computing subsets of strictly commutative arguments.

Theorem 1. *ε-commutativity is not closed under unions of sets with respect to which a factor ϕ is ε-commutative.*

Proof. Let $(r_1, r_2, r_3) \in \times_{i=1}^{3} \{\text{high}, \text{medium}, \text{low}\}$ denote any assignment to the randvars R_1, R_2, R_3. Let ε be 0.1. The potentials of the factor ϕ are given by

R_1	R_2	R_3	$\phi(R_1, R_2, R_3)$
high	medium	low	1
high	low	medium	1.1
medium	high	low	1.1
medium	low	high	1.21
low	high	medium	1.21
low	medium	high	1.1

All non-listed assignments are mapped to the potential 1. ϕ is ε-commutative with respect to all individual pairs $\{R_1, R_2\}$, $\{R_1, R_3\}$, and $\{R_2, R_3\}$, but not

with respect to the set $\{R_1, R_2, R_3\}$. Consider the assignment $(r_1, r_2, r_3) = (\text{high}, \text{medium}, \text{low})$ and the permutation $\pi(1) = 3, \pi(2) = 1, \pi(3) = 2$, (cycle of two transpositions), which maps $(\text{high}, \text{medium}, \text{low})$ to $(r_{\pi(1)}, r_{\pi(2)}, r_{\pi(3)}) = (r_3, r_1, r_2) = (\text{low}, \text{high}, \text{medium})$. However, $\phi(\text{high}, \text{medium}, \text{low}) = 1$ is not an element of the interval $[\phi(\text{low}, \text{high}, \text{medium}) \cdot (1-0.1), \phi(\text{low}, \text{high}, \text{medium}) \cdot (1+0.1)] = [1.089, 1.331]$. Thus, ϕ is not ε-commutative with respect to $\{R_1, R_2, R_3\}$. □

Theorem 1 implies that the algorithm presented in [9, Algorithm 2] to efficiently compute a maximum sized subset of commutative arguments cannot directly be transferred to compute a subset of ε-commutative arguments instead. We next turn our attention to the compression of an ε-commutative factor, which is performed by replacing it with a smaller representative factor.

3.1 Replacement of Potentials by Representative

The construction developed in this subsection formalises the aggregation of potentials over symmetry sets induced by permutations of ε-commutative arguments. We first introduce symmetrised representative factors as weighted representatives over symmetry sets and then study the algebraic structure induced by the permutation group $\Pi_{\boldsymbol{C}_\phi}$, which partitions the assignment space into equivalence classes of pairwise ε-equivalent potentials.

Definition 4 (Symmetrised Representative Factor). *Let $\phi(R_1, \ldots, R_n)$ be a factor, which is ε-commutative with respect to $\boldsymbol{C}_\phi \subseteq \boldsymbol{R}_\phi = \{R_1, \ldots, R_n\}$. For an assignment $\boldsymbol{r} = (r_1, \ldots, r_n) \in \mathcal{X}_{\boldsymbol{R}_\phi}$, the* symmetrised representative factor ϕ^* *associated with ϕ is defined by*

$$\phi^*(\boldsymbol{r}) := \sum_{\boldsymbol{r}' \in \mathfrak{S}_\phi(\boldsymbol{r})} \omega_{\boldsymbol{r}'} \cdot \phi(\boldsymbol{r}'),$$

where $\mathfrak{S}_\phi(\boldsymbol{r}) := \left\{ (r_{\pi(1)}, \ldots, r_{\pi(n)}) \;\middle|\; \pi \in \Pi_{\boldsymbol{C}_\phi} \right\}$ denotes the symmetry set of $\boldsymbol{r}$ under permutations of the ε-commutative arguments, and where the weights satisfy $\omega_{\boldsymbol{r}'} \geq 0$ and $\sum_{\boldsymbol{r}' \in \mathfrak{S}_\phi(\boldsymbol{r})} \omega_{\boldsymbol{r}'} = 1$.

To better understand why such a construction is possible, we first investigate the structure underlying these symmetry classes. This leads us to examine $\Pi_{\boldsymbol{C}_\phi}$, which is shown to be closed under composition and inversion in the following.

Lemma 2. *For a set $\boldsymbol{C}_\phi \subseteq \boldsymbol{R}_\phi = \{R_1, \ldots, R_n\}$, $\Pi_{\boldsymbol{C}_\phi}$ is a subgroup of S_n.*

Proof. By definition $\pi_{\text{id}} \in \Pi_{\boldsymbol{C}_\phi}$. If $\pi_1, \pi_2 \in \Pi_{\boldsymbol{C}_\phi}$, then $(\pi_1 \circ \pi_2)(i) = \pi_1(\pi_2(i)) = \pi_1(i) = i$ for all $R_i \notin \boldsymbol{C}_\phi$, resulting in closed group operations $\pi_1 \circ \pi_2 \in \Pi_{\boldsymbol{C}_\phi}$. If $\pi \in \Pi_{\boldsymbol{C}_\phi}$, then $\pi \in S_n$ due to $\Pi_{\boldsymbol{C}_\phi} \subseteq S_n$. S_n is a group and therefore $\pi^{-1} \in S_n$. This means that for all i with $R_i \notin \boldsymbol{C}_\phi$ it holds that $i = \pi_{\text{id}}(i) = (\pi^{-1} \circ \pi)(i) = \pi^{-1}(\pi(i)) = \pi^{-1}(i)$, which means that the inverse π^{-1} is also in $\Pi_{\boldsymbol{C}_\phi}$. □

This subgroup naturally induces a group operation of $\Pi_{\boldsymbol{C}_\phi}$ on the assignment space $\mathcal{X}_{\boldsymbol{R}_\phi}$ via $(r_1, \ldots, r_n) \mapsto (r_{\pi(1)}, \ldots, r_{\pi(n)})$ for $\pi \in \Pi_{\boldsymbol{C}_\phi}$, whose properties are inherited by the corresponding symmetry sets $\mathfrak{S}_\phi(\boldsymbol{r})$.

Hence, symmetry sets are precisely the orbits of this group operation.

Corollary 1. *Let $\phi(R_1, \dots, R_n)$ be a factor, which is ε-commutative with respect to $\boldsymbol{C}_\phi \subseteq \boldsymbol{R}_\phi = \{R_1, \dots, R_n\}$. If $\boldsymbol{r}' \in \mathfrak{S}_\phi(\boldsymbol{r})$, then $\mathfrak{S}_\phi(\boldsymbol{r}') = \mathfrak{S}_\phi(\boldsymbol{r})$.*

Proof. As $\boldsymbol{r}' \in \mathfrak{S}_\phi(\boldsymbol{r})$, there exists $\pi_1 \in \Pi_{\boldsymbol{C}_\phi}$ such that $\boldsymbol{r}' = (r_{\pi_1(1)}, \dots, r_{\pi_1(n)})$. Let $\boldsymbol{r}'' \in \mathfrak{S}_\phi(\boldsymbol{r}')$. Then, there is $\pi_2 \in \Pi_{\boldsymbol{C}_\phi}$ such that $\boldsymbol{r}'' = (r'_{\pi_2(1)}, \dots, r'_{\pi_2(n)}) = (r_{\pi_1(\pi_2(1))}, \dots, r_{\pi_1(\pi_2(n))})$. Since $\Pi_{\boldsymbol{C}_\phi}$ is a subgroup of S_n, it is closed under composition, hence $\pi_1 \circ \pi_2 \in \Pi_{\boldsymbol{C}_\phi}$. Therefore $\boldsymbol{r}'' \in \mathfrak{S}_\phi(\boldsymbol{r})$, which shows $\mathfrak{S}_\phi(\boldsymbol{r}') \subseteq \mathfrak{S}_\phi(\boldsymbol{r})$. The reverse inclusion follows analogously for $\pi_1^{-1}, \pi_2^{-1} \in \Pi_{\boldsymbol{C}_\phi}$. □

While the previous result describes how ε-commutativity behaves on the level of argument subsets, the following lemma establishes the induced equivalence on the level of potentials via ε-equivalence.

Lemma 3. *Let $\phi(R_1, \dots, R_n)$ be ε-commutative with respect to $\boldsymbol{C}_\phi$. Then, all potentials induced by assignments within the same symmetry set are pairwise ε-equivalent, i.e., if $\boldsymbol{r}', \boldsymbol{r}'' \in \mathfrak{S}_\phi(\boldsymbol{r})$, then $\phi(\boldsymbol{r}') =_\varepsilon \phi(\boldsymbol{r}'')$.*

Proof. Let $\boldsymbol{r}', \boldsymbol{r}'' \in \mathfrak{S}_\phi(\boldsymbol{r})$. Then there exist permutations $\pi_1, \pi_2 \in \Pi_{\boldsymbol{C}_\phi}$ such that $\boldsymbol{r}' = (r_{\pi_1(1)}, \dots, r_{\pi_1(n)})$ and $\boldsymbol{r}'' = (r_{\pi_2(1)}, \dots, r_{\pi_2(n)})$. Since $\Pi_{\boldsymbol{C}_\phi}$ is a subgroup of S_n (Lemma 2), it is closed under composition and taking inverses. Hence, $\tau := \pi_2 \circ \pi_1^{-1} \in \Pi_{\boldsymbol{C}_\phi}$. Therefore, $\boldsymbol{r}'' = (r'_{\tau(1)}, \dots, r'_{\tau(n)})$. Because ϕ is ε-commutative with respect to $\boldsymbol{C}_\phi$, it follows from Def. 3 that $\phi(\boldsymbol{r}') \in [\phi(\boldsymbol{r}'') \cdot (1-\varepsilon), \phi(\boldsymbol{r}'') \cdot (1+\varepsilon)]$ and $\phi(\boldsymbol{r}'') \in [\phi(\boldsymbol{r}') \cdot (1-\varepsilon), \phi(\boldsymbol{r}') \cdot (1+\varepsilon)]$. Hence, it holds that $\phi(r'_1, \dots, r'_n) =_\varepsilon \phi(r''_1, \dots, r''_n)$. □

Thus, we can also view $\mathfrak{S}_\phi(\boldsymbol{r}) =: [\boldsymbol{r}]_\phi$ as the equivalence class of the equivalence relation $\sim$ on $\mathcal{X}_{\boldsymbol{R}_\phi}$ induced by the group operation of $\Pi_{\boldsymbol{C}_\phi}$ via

$$\boldsymbol{r} \sim \boldsymbol{r}' \quad :\Longleftrightarrow \quad \exists \pi \in \Pi_{\boldsymbol{C}_\phi} \colon (r'_1, \dots, r'_n) = (r_{\pi(1)}, \dots, r_{\pi(n)}).$$

Consequently, the quotient space $\mathcal{X}_{\boldsymbol{R}_\phi} / \sim$ is well-defined and can be identified with the set of all symmetry sets. Hence, there exists a representative set $\mathcal{X}_{\mathcal{R}_\phi} \subset \mathcal{X}_{\boldsymbol{R}_\phi}$ containing exactly one element per equivalence class, inducing a bijection

$$\mathcal{X}_{\boldsymbol{R}_\phi} / \sim \;\cong\; \{[\boldsymbol{r}]_\phi \mid \boldsymbol{r} \in \mathcal{X}_{\mathcal{R}_\phi}\} = \{\mathfrak{S}_\phi(\boldsymbol{r}) \mid \boldsymbol{r} \in \mathcal{X}_{\mathcal{R}_\phi}\}.$$

Different symmetry sets (equivalence classes) of the same factor may induce different weighting schemes. Within each equivalence class, however, the weights must be chosen consistently with the quotient structure $\mathcal{X}_{\boldsymbol{R}_\phi} / \sim$ so that every class contributes exactly once to the representative potential (for the size of the symmetry set, see Appendix A.1). This raises the question whether replacing potentials by symmetrised representatives preserves ε-equivalence of factors.

Theorem 2. *Let ϕ be ε-commutative with respect to $\boldsymbol{C}_\phi$ and let ϕ^* be its symmetrised representative factor for $\varepsilon \in [0, 1]$. Then, ϕ and ϕ^* are ε-equivalent.*

Proof. By Lemma 3, potentials $\phi(\boldsymbol{r}')$ with $\boldsymbol{r}' \in \mathfrak{S}_\phi(\boldsymbol{r})$ are pairwise ε-equivalent. Hence, by [17, Theorem 15], any weighted mean $\sum_{\boldsymbol{r}' \in \mathfrak{S}_\phi(\boldsymbol{r})} \omega_{\boldsymbol{r}'} \cdot \phi(\boldsymbol{r}')$ is also ε-equivalent to $\phi(\boldsymbol{r}')$. Applied row-wise to all assignments and potentials relative to their equivalence classes, this yields ε-equivalent factors. □

Choosing the arithmetic mean $\overline{\phi}$ as the symmetrised representative factor arises as a trivial special case of Thm. 2 by assigning uniform weights $\omega_{\boldsymbol{r}'} := \frac{1}{|\mathfrak{S}_\phi(\boldsymbol{r})|}$ on $\mathfrak{S}_\phi(\boldsymbol{r})$ for which $\sum_{\boldsymbol{r}' \in \mathfrak{S}_\phi(\boldsymbol{r})} \omega_{\boldsymbol{r}'} = 1$ holds, yielding $\phi^*(\boldsymbol{r}) = \overline{\phi}(\boldsymbol{r})$.

Corollary 2. *Let ϕ be ε-commutative with respect to $\boldsymbol{C}_\phi$ and let the mean*

$$\phi^*(\boldsymbol{r}) := \overline{\phi}(\boldsymbol{r}) := \frac{1}{|\mathfrak{S}_\phi(\boldsymbol{r})|} \sum_{\boldsymbol{r}' \in \mathfrak{S}_\phi(\boldsymbol{r})} \phi(\boldsymbol{r}'), \tag{2}$$

be defined with weights $\omega_{\boldsymbol{r}'} := \frac{1}{|\mathfrak{S}_\phi(\boldsymbol{r})|}$ for $\boldsymbol{r}' \in \mathfrak{S}_\phi(\boldsymbol{r})$ for a given $\varepsilon \in [0, 1]$. Then ϕ^* is a symmetrised representative factor of ϕ and is ε-equivalent to ϕ.

The mean is particularly attractive as it treats all symmetric assignments uniformly avoiding bias towards specific permutations and minimises the squared deviation [10]. It is also optimal with respect to the squared deviation in the sense that, when all values $\phi(\boldsymbol{r}')$ are replaced by the same representative within the group $\boldsymbol{r} \in \mathfrak{S}_\phi(\boldsymbol{r})$, the resulting deviation from the original factor is minimised. These properties make it a natural choice for the symmetrised representative factor in practical implementations (see Sec. 4 for experiments). However, when prior knowledge indicates a bias toward specific arguments in the underlying data, or when domain expertise is available, the weights assigned to individual arguments of the original factor can always be adapted accordingly.

3.2 Asymptotic Bounds

To quantify the approximation error induced by symmetrisation, we next analyse the resulting deviation between the original model and the model obtained by symmetrising ε-commutative factors. We make use of the following symmetric distance measure, which allows us to bound the change in query results [2,10].

Definition 5 ([2]). *The symmetric distance $D_{CD}(P_M, P_{M'})$ between two distributions P_M and $P_{M'}$ introduced in [2] (see [10] for details) is defined as*

$$D_{CD}(P_M, P_{M'}) := \ln \max_{\boldsymbol{r}} \frac{P_{M'}(\boldsymbol{r})}{P_M(\boldsymbol{r})} - \ln \min_{\boldsymbol{r}} \frac{P_{M'}(\boldsymbol{r})}{P_M(\boldsymbol{r})} \tag{3}$$

$$= \ln \max_{\boldsymbol{r}} \frac{\psi'(\boldsymbol{r})}{\psi(\boldsymbol{r})} - \ln \min_{\boldsymbol{r}} \frac{\psi'(\boldsymbol{r})}{\psi(\boldsymbol{r})}. \tag{4}$$

With this measure at hand, we first derive a general bound for arbitrary weighting schemes.

Theorem 3. *Let M be an FG with factors $\phi_1, \ldots, \phi_m$, of which $\phi_1, \ldots, \phi_k$ with $0 \leq k \leq m$ are ε-commutative with respect to $\boldsymbol{C}_{\phi_1}, \ldots, \boldsymbol{C}_{\phi_k}$, respectively. Let M' be the FG with factors $\phi_1^*, \ldots, \phi_k^*, \phi_{k+1}, \ldots, \phi_m$, where ε-commutative factors of M are replaced by their symmetrised representative factors. Further, let P_M and $P_{M'}$ denote the underlying full joint probability distributions encoded by M and M', respectively. Then, it holds that*

$$D_{CD}(P_M, P_{M'}) \leq \ln\left(1 + \varepsilon\right)^{2k}.$$

Proof. According to Def. 3, every updated potential $\phi_i^*(\boldsymbol{r}_i)$ in M' differs from its original potential $\phi_i(\boldsymbol{r}_i)$ in M by at most a factor $(1 \pm \varepsilon)$ for $i = 1, \ldots, k$, while all remaining factors $i > k$ are left unchanged. Since $\psi(\boldsymbol{r}) = \prod_{j=1}^{m} \phi_j(\boldsymbol{r}_j)$ for any assignment $\boldsymbol{r}$, using Lemma 1 we obtain

$$\psi'(\boldsymbol{r}) \geq \prod_{j=1}^{k} \phi_j(\boldsymbol{r}_j) \cdot \frac{1}{1+\varepsilon} \cdot \prod_{j=k+1}^{m} \phi_j(\boldsymbol{r}_j) = \frac{1}{(1+\varepsilon)^k} \cdot \prod_{j=1}^{m} \phi_j(\boldsymbol{r}_j), \text{ and}$$
$$\psi'(\boldsymbol{r}) \leq \prod_{j=1}^{k} \phi_j(\boldsymbol{r}_j) \cdot (1+\varepsilon) \cdot \prod_{j=k+1}^{m} \phi_j(\boldsymbol{r}_j) = (1+\varepsilon)^k \cdot \prod_{j=1}^{m} \phi_j(\boldsymbol{r}_j).$$

Consequently,

$$\min_{\boldsymbol{r}} \frac{\psi'(\boldsymbol{r})}{\psi(\boldsymbol{r})} \geq \frac{\frac{1}{(1+\varepsilon)^k} \cdot \prod_{j=1}^{m} \phi_j(\boldsymbol{r}_j)}{\prod_{j=1}^{m} \phi_j(\boldsymbol{r}_j)} = \frac{1}{(1+\varepsilon)^k}, \text{ and}$$
$$\max_{\boldsymbol{r}} \frac{\psi'(\boldsymbol{r})}{\psi(\boldsymbol{r})} \leq \frac{(1+\varepsilon)^k \cdot \prod_{j=1}^{m} \phi_j(\boldsymbol{r}_j)}{\prod_{j=1}^{m} \phi_j(\boldsymbol{r}_j)} = (1+\varepsilon)^k,$$

where $\boldsymbol{r}_j \in \mathcal{X}_{\boldsymbol{R}_{\phi_j}}$ denotes the projection of $\boldsymbol{r} \in \mathcal{X}_{\boldsymbol{R}}$. Substituting these bounds into Eq. (4) yields $D_{CD}(P_M, P_{M'}) \leq \ln(1+\varepsilon)^k - \ln 1 \,/\, (1+\varepsilon)^k = \ln(1+\varepsilon)^{2k}$. □

For the special case of arithmetic averaging, we can sharpen this bound.

Lemma 4. *Let ϕ be an ε-commutative factor with respect to $\boldsymbol{C}_\phi$, let $\mathcal{X}_{\mathcal{R}_\phi} = \dot{\bigcup}_{j=1}^{l} \{\boldsymbol{r}^j\}$ be a set of disjoint representatives $\boldsymbol{r}^j$ for all disjoint equivalence classes $\dot{\bigcup}_{j=1}^{l} [\boldsymbol{r}^j]_\phi = \dot{\bigcup}_{j=1}^{l} \mathfrak{S}_\phi(\boldsymbol{r}^j) = \mathcal{X}_{\boldsymbol{R}_\phi}$ ordered by size such that $m_1 := |[\boldsymbol{r}^1]_\phi| \geq \ldots \geq m_l := |[\boldsymbol{r}^l]_\phi| \geq 1$, and let ϕ^* be the symmetrised representative factor via the mean as in Eq.* (2)*. Then,*

$$1 \;\leq\; \frac{\max_{\boldsymbol{r}} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})}}{\min_{\boldsymbol{r}} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})}} \;\leq\; \frac{\left(1 + \frac{m_2 - 1}{m_2}\varepsilon\right)(1+\varepsilon)}{1 + \frac{1}{m_1}\varepsilon}.$$

Condensed Proof (Longer Version in App. A). The mean $\phi^*(\boldsymbol{r})$ is uniquely determined within each equivalence class $[\boldsymbol{r}^j]_\phi$. First, observe that for every equivalence class $[\boldsymbol{r}^j]_\phi$ and all $\boldsymbol{r}', \boldsymbol{r}'' \in [\boldsymbol{r}^j]_\phi$, Lemma 3 implies pairwise ε-equivalence, so the potentials $\phi(\boldsymbol{r}'')$ within one equivalence class $[\boldsymbol{r}^j]_\phi$ can be ordered

$$0 \leq \frac{1}{1+\varepsilon}\phi(\boldsymbol{r}^{j,\max}) \leq \phi(\boldsymbol{r}^{j,\min}) \leq \phi(\boldsymbol{r}'') \leq \phi(\boldsymbol{r}^{j,\max}) \leq \phi(\boldsymbol{r}^{j,\min})(1+\varepsilon),$$

where $\phi(\boldsymbol{r}^{j,\max}) := \max_{\boldsymbol{r}' \in [\boldsymbol{r}^j]_\phi} \phi(\boldsymbol{r}')$ and $\phi(\boldsymbol{r}^{j,\min}) := \min_{\boldsymbol{r}' \in [\boldsymbol{r}^j]_\phi} \phi(\boldsymbol{r}')$. When the maximum and minimum originate from two different equivalence classes

$[\boldsymbol{r}^i]_\phi$ and $[\boldsymbol{r}^j]_\phi$ with $i \neq j$, the worst-case assignments are independent. For the maximum choose $\phi(\boldsymbol{r}) = \phi(\boldsymbol{r}^{i,\min})$, while all remaining potentials in the same equivalence class attain their maximal admissible value $\phi(\boldsymbol{r}') := \phi(\boldsymbol{r}^{i,\max}) = (1+\varepsilon)\phi(\boldsymbol{r}^{i,\min})$ for all $\boldsymbol{r}' \in [\boldsymbol{r}^i]_\phi \setminus \{\boldsymbol{r}\}$. This yields

$$\begin{aligned}\max_{\boldsymbol{r}\in[\boldsymbol{r}^i]_\phi} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})} &= \frac{(\phi(\boldsymbol{r}^{i,\min}) + (|[\boldsymbol{r}^i]_\phi| - 1)(1+\varepsilon)\phi(\boldsymbol{r}^{i,\min})) \,/\, |[\boldsymbol{r}^i]_\phi|}{\phi(\boldsymbol{r}^{i,\min})} \\ &= 1 + \frac{(m_i - 1)\varepsilon}{m_i}.\end{aligned}$$

For the minimum choose $\phi(\boldsymbol{r}) = \phi(\boldsymbol{r}^{j,\max})$, while all remaining values in the same equivalence class attain their minimal admissible value $\phi(\boldsymbol{r}') := \phi(\boldsymbol{r}^{j,\min}) = \frac{1}{1+\varepsilon}\phi(\boldsymbol{r}^{i,\max})$ for all $\boldsymbol{r}' \in [\boldsymbol{r}^j]_\phi \setminus \{\boldsymbol{r}\}$. Consequently,

$$\min_{\boldsymbol{r}\in[\boldsymbol{r}^j]_\phi} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})} = \frac{((|[\boldsymbol{r}^j]_\phi| - 1)\frac{1}{1+\varepsilon}\phi(\boldsymbol{r}^{j,\max}) + \phi(\boldsymbol{r}^{j,\max})) \,/\, |[\boldsymbol{r}^j]_\phi|}{\phi(\boldsymbol{r}^{j,\max})} = \frac{1 + \frac{1}{m_j}\varepsilon}{1+\varepsilon}.$$

Observe, that $1 + \frac{(m_i-1)\varepsilon}{m_i}$ is monotonically increasing in m_i, whereas $\frac{1+\frac{1}{m_j}\varepsilon}{1+\varepsilon}$ is monotonically decreasing in m_j. Hence, the largest equivalence classes by size produce the extremal deviations. To determine which class size appear in the numerator and denominator, consider the monotonically decreasing $g(x) = \left(1 + \frac{x-1}{x}\,\varepsilon\right)\left(1 + \frac{1}{x}\varepsilon\right)$ for $x \geq 2$. Therefore, for $x_1, x_2 \in \mathbb{N}_{\geq 2}$ it follows

$$g(x_1) \geq g(x_2) \quad \Leftrightarrow \quad \frac{1 + \frac{x_1-1}{x_1}\,\varepsilon}{1 + \frac{1}{x_2}\varepsilon} \geq \frac{1 + \frac{x_2-1}{x_2}\,\varepsilon}{1 + \frac{1}{x_1}\varepsilon}.$$

Consequently, the sharpest bound is obtained by using $x_1 = m_2$ and $x_2 = m_1$. When both extrema originate from the same equivalence class, $\phi^*(\boldsymbol{r}) = \phi^*(\boldsymbol{r}')$ for all $\boldsymbol{r}, \boldsymbol{r}' \in [\boldsymbol{r}^j]_\phi$, the quotient is maximised via $\phi(\boldsymbol{r}) = \phi(\boldsymbol{r}^{j,\min}) = \frac{1}{1+\varepsilon}\phi(\boldsymbol{r}^{j,\max})$ and $\phi(\boldsymbol{r}') = \phi(\boldsymbol{r}^{j,\max}) = (1+\varepsilon)\phi(\boldsymbol{r}^{j,\min})$, leading to smaller values

$$\frac{\max_{\boldsymbol{r}\in[\boldsymbol{r}^j]_\phi} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})}}{\min_{\boldsymbol{r}'\in[\boldsymbol{r}^j]_\phi} \frac{\phi^*(\boldsymbol{r}')}{\phi(\boldsymbol{r}')}} = \frac{\max_{\boldsymbol{r}\in[\boldsymbol{r}^j]_\phi} \frac{1}{\phi(\boldsymbol{r})}}{\min_{\boldsymbol{r}'\in[\boldsymbol{r}^j]_\phi} \frac{1}{\phi(\boldsymbol{r}')}} \leq \frac{\frac{1}{\phi(\boldsymbol{r}^{j,\min})}}{\frac{1}{(1+\varepsilon)\phi(\boldsymbol{r}^{j,\min})}} = 1 + \varepsilon. \qquad \square$$

Combining the class-wise extremal bounds yields the final global statement for the representative mean-construction, which we show next.

Theorem 4. *Let M be an FG with factors $\phi_1, \ldots, \phi_m$, of which $\phi_1, \ldots, \phi_k$ with $0 \leq k \leq m$ are ε-commutative with respect to $\boldsymbol{C}_{\phi_1}, \ldots, \boldsymbol{C}_{\phi_k}$, respectively. Let $\mathcal{X}_{\mathcal{R}_{\phi_i}} = \dot{\bigcup}_{j=1}^{l_i}\{\boldsymbol{r}_i^j\}$ be the set of disjoint representatives $\boldsymbol{r}_i^j$ for all equivalence classes $\dot{\bigcup}_{j=1}^{l_i}[\boldsymbol{r}_i^j]_{\phi_i} = \mathcal{X}_{\boldsymbol{R}_{\phi_i}}$ ordered by size with $m_{1_i} := |[\boldsymbol{r}_i^1]_{\phi_i}| \geq \ldots \geq m_{l_i} := |[\boldsymbol{r}_i^l]_{\phi_i}| \geq 1$. Let M' be the FG obtained by M by replacing each factor ϕ_i by its mean-based symmetrised representative factor ϕ_i^* as in Eq. (2). Then, for the induced distributions P_M and $P_{M'}$ and $\varepsilon > 0$, it holds that*

$$D_{CD}(P_M, P_{M'}) \leq \ln \prod_{i=1}^{k} \frac{\left(1 + \frac{m_{2_i}-1}{m_{2_i}}\varepsilon\right)(1+\varepsilon)}{1 + \frac{1}{m_{1_i}}\varepsilon} \tag{5}$$

$$< \ln \left(1+\varepsilon\right)^{2k}.$$

Proof. The statement follows by factorisation of the joint potential, the determination of the extrema on a projected factor-level instead of a global assignment-level, and the class-wise bound from Lemma 4 for the monotonic logarithm.

$$\begin{aligned}
D_{CD}(P_M, P_{M'}) &\overset{\text{Eq. (4)}}{=} \ln \max_{\boldsymbol{r}} \frac{\psi'(\boldsymbol{r})}{\psi(\boldsymbol{r})} - \ln \min_{\boldsymbol{r}} \frac{\psi'(\boldsymbol{r})}{\psi(\boldsymbol{r})} \\
&= \ln \max_{\boldsymbol{r}} \frac{\prod_{i=1}^{k} \phi_i^*(\boldsymbol{r}_i) \prod_{i=k+1}^{m} \phi_i(\boldsymbol{r}_i)}{\prod_{i=1}^{m} \phi_i(\boldsymbol{r}_i)} - \ln \min_{\boldsymbol{r}} \frac{\prod_{i=1}^{k} \phi_i^*(\boldsymbol{r}_i) \prod_{i=k+1}^{m} \phi_i(\boldsymbol{r}_i)}{\prod_{i=1}^{m} \phi_i(\boldsymbol{r}_i)} \\
&= \ln \max_{\boldsymbol{r}} \prod_{i=1}^{k} \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)} - \ln \min_{\boldsymbol{r}} \prod_{i=1}^{k} \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)} \\
&\leq \ln \prod_{i=1}^{k} \max_{\boldsymbol{r}_i} \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)} - \ln \prod_{i=1}^{k} \min_{\boldsymbol{r}_i} \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)} \\
&= \ln \prod_{i=1}^{k} \frac{\max_{\boldsymbol{r}_i} \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)}}{\min_{\boldsymbol{r}_i} \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)}} \\
&\overset{\text{Lemma 4}}{\leq} \ln \prod_{i=1}^{k} \frac{\left(1 + \frac{m_{2_i}-1}{m_{2_i}}\varepsilon\right)(1+\varepsilon)}{1 + \frac{1}{m_{1_i}}\varepsilon}.
\end{aligned}$$

□

Even more precisely, the bound is sharp in the general non-trivial case.

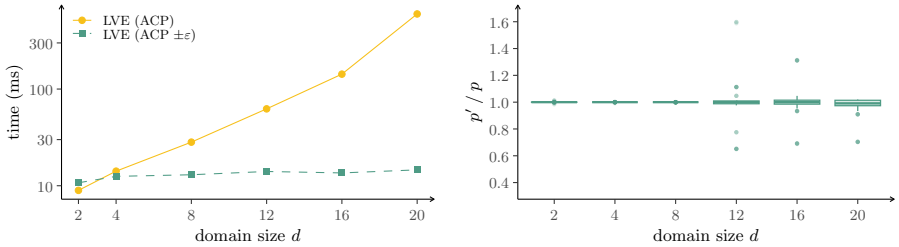
Theorem 5. *The bound given in Thm. 4 is optimal.*

Proof Sketch. Ex. 3 in App. A uses the construction from the proof of Lemma 4 per factor, hits the bound, and is therefore sharp. □

Note that the choice of a multiplicative relaxation in Def. 3 is what makes this bound attainable in the first place, as D_{CD} is based on quotients of potentials and only relies on relative sizes. An additive relaxation, requiring permuted potentials differ by at most a constant, would lead to arbitrary large deviations.

4 Experiments

We complement our theoretical results with an empirical evaluation assessing to what extent ε-commutativity translates into practical benefits for lifted model construction and downstream inference. In particular, we answer the question of how the trade-off between higher compression and accuracy of query results behaves when compressing ε-commutative factors. Concretely, we compare the run time of LVE on the compressed model returned by the exact ACP algorithm (which is only able to compress strictly commutative factors) to the run time of LVE on the compressed model returned by its ε-relaxed variant using the mean



(a) LVE run time (ms). (b) Query result quotient $p' \,/\, p$.

Fig. 2: Downstream lifted inference performance, averaged over all generated instances and queries. (a) Run time of LVE on the compressed model returned by ACP and its ε-relaxed variant (ACP $\pm\varepsilon$). (b) Distribution of the per-query quotient $p' \,/\, p$ between the approximate result p' and the exact result p.

as the symmetrised representative factor (denoted as "ACP $\pm\varepsilon$" in the following) and measure the deviation of the query results, quantified by the per-query quotient $p' \,/\, p$ between the query result (i.e., marginal probability) p' obtained on the ACP $\pm\varepsilon$-compressed model and the query result p obtained on the exact ACP-compressed model. For our experiments, we use the same input FGs as in the original ACP paper [7] and add noise to the potentials of the (strictly) commutative factors to obtain ε-commutative factors. We only manipulate commutative factors and leave the remaining factors unchanged to investigate the effect of ε-commutativity on the compression-accuracy trade-off in isolation (the effect of grouping different factors that are approximately equal has been investigated in [10] and hence, we only consider indistinguishability *within* factors, characterised by ε-commutativity, and no indistinguishability *between* factors). Specifically, each input FG contains between $2d+1$ and $d \cdot \lfloor \log_2(d) \rfloor + 2d + 2$ randvars with Boolean range and between $2d+1$ and $d \cdot \lfloor \log_2(d) \rfloor + d + 2$ factors, where $d \in \{2, 4, 8, 12, 16, 20\}$ controls the size of the FGs. Each input FG contains $k \in \{1, 3, 7\}$ ε-commutative factors, where $\varepsilon \in \{0.001, 0.01, 0.1\}$. The ε-commutative factors are obtained by multiplying the potentials of the (strictly) commutative factors by a uniform sample from $[1, 1+\varepsilon]$.

Fig. 2 reports the results averaged over all generated instances and queries. Fig. 2a shows that the compression of ε-commutative factors substantially accelerates downstream probabilistic inference, where the speedup increases exponentially with the domain size. At the same time, Fig. 2b confirms that the query results (i.e., marginal probabilities) remain extremely close to those obtained with exact ACP: The per-query quotient $p' \,/\, p$ concentrates tightly around the optimal value one across all evaluated configurations (except for very few outliers). In other words, ACP $\pm\varepsilon$ unlocks the speedup of lifted inference at almost no loss of accuracy. To give insights into the influence of k and ε, we provide further experimental results for individual choices in App. B.

5 Conclusion

We have introduced ε-commutativity, a principled relaxation of commutativity that tolerates the minute numerical deviations that inevitably arise in practice, e.g., when factor potentials are estimated from data. On this basis, we have developed a representative-based compression scheme that collapses ε-equivalent potentials into a shared value, integrated the relaxation into the state-of-the-art ACP algorithm, and proved a sharp (optimal) bound on the resulting approximation error of probabilistic queries. Our empirical evaluation confirms that the bound is loose in practice: Compressing ε-commutative factors unlocks substantial lifted inference speedups while the deviation of query results stays well within the theoretical guarantee. As factors are merely non-negative functions, all proposed concepts directly apply to the detection and compact storage of approximately symmetric functions in more general settings.

An interesting direction for future work is to investigate approximate indistinguishability for FGs with continuous variables. Currently, the usage of a CRV restricts the present work to the discrete case and a continuous counterpart of a CRV is required to handle ε-commutativity in models with continuous variables.

Acknowledgments

This work was partially funded by the Ministry of Culture and Science of the German State of North Rhine-Westphalia. This preprint has not undergone peer review or any post-submission improvements or corrections. The Version of Record of this contribution is published in *Lecture Notes in Computer Science, Volume 17188*, and is available online at `https://doi.org/10.1007/978-3-032-43514-9_14`.

Declaration on Generative AI

During the preparation of this work, the authors used Claude Code (Anthropic) to discuss ideas and to polish text passages. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

1. Berger, J.O.: Robust Bayesian Analysis: Sensitivity to the Prior. Journal of Statistical Planning and Inference **25**, 303–328 (1990)
2. Chan, H., Darwiche, A.: A Distance Measure for Bounding Probabilistic Belief Change. International Journal of Approximate Reasoning **38**, 149–174 (2005)
3. Cooper, G.F.: The Computational Complexity of Probabilistic Inference using Bayesian Belief Networks. Artificial Intelligence **42**, 393–405 (1990)

4. Frey, B.J., Kschischang, F.R., Loeliger, H.A., Wiberg, N.: Factor Graphs and Algorithms. In: Proceedings of the Thirty-Fifth Annual Allerton Conference on Communication, Control, and Computing. pp. 666–680. Allerton House (1997)
5. Kschischang, F.R., Frey, B.J., Loeliger, H.A.: Factor Graphs and the Sum-Product Algorithm. IEEE Transactions on Information Theory **47**, 498–519 (2001)
6. Li, Z., Junna, S., Liao, W.: A Robust Factor Graph Framework for Navigation on PDR/Magnetic Field Integration. Measurement **245**, 116509 (2025)
7. Luttermann, M., Braun, T., Möller, R., Gehrke, M.: Colour Passing Revisited: Lifted Model Construction with Commutative Factors. In: Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-2024). pp. 20500–20507. AAAI Press (2024)
8. Luttermann, M., Machemer, J., Gehrke, M.: Efficient Detection of Commutative Factors in Factor Graphs. In: Proceedings of the Twelfth International Conference on Probabilistic Graphical Models (PGM-2024). pp. 38–56. PMLR (2024)
9. Luttermann, M., Möller, R., Gehrke, M.: On the Detection of Commutative Factors in Factor Graphs: Necessary and Sufficient Conditions. arXiv preprint, https://arxiv.org/abs/2605.26908 (2026)
10. Luttermann, M., Speller, J., Gehrke, M., Braun, T., Möller, R., Hartwig, M.: Approximate Lifted Model Construction. In: Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-2025). pp. 9077–9085. IJCAI Organization (2025)
11. Milch, B., Zettlemoyer, L.S., Kersting, K., Haimes, M., Kaelbling, L.P.: Lifted Probabilistic Inference with Counting Formulas. In: Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI-2008). pp. 1062–1068. AAAI Press (2008)
12. Miranda, E., Montes, I., Destercke, S.: A Unifying Frame for Neighbourhood and Distortion Models. In: Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications. pp. 304–313. PMLR (2019)
13. Niepert, M., Van den Broeck, G.: Tractability through Exchangeability: A New Perspective on Efficient Probabilistic Inference. In: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI-2014). pp. 2467–2475. AAAI Press (2014)
14. Pfeifer, T., Weissig, P., Lange, S., Protzel, P.: Robust Factor Graph Optimization – A Comparison for Sensor Fusion Applications. In: Proceedings of the Twenty-First IEEE International Conference on Emerging Technologies and Factory Automation. pp. 1–4. IEEE (2016)
15. Poole, D.: First-Order Probabilistic Inference. In: Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-2003). pp. 985–991. Morgan Kaufmann Publishers Inc. (2003)
16. Speller, J., Luttermann, M., Gehrke, M., Braun, T.: Compression versus Accuracy: A Hierarchy of Lifted Models. In: Proceedings of the Twenty-Eighth European Conference on Artificial Intelligence (ECAI-2025). pp. 5051–5058. IOS Press (2025)
17. Speller, J., Luttermann, M., Gehrke, M., Braun, T.: Towards Explainability of Approximate Lifted Model Construction: A Geometric Perspective. In: Proceedings of the Eleventh Workshop on Formal and Cognitive Reasoning (FCR-2025). pp. 41–56. CEUR (2025)
18. Taghipour, N.: Lifted Probabilistic Inference by Variable Elimination. Ph.D. thesis, KU Leuven (2013)

A Detailed Proofs

Lemma 1. *Let ϕ denote a factor that is ε-commutative with respect to a subset of its arguments $\boldsymbol{C}_\phi$. Then, for all permutations $\pi \in \Pi_{\boldsymbol{C}_\phi}$, it holds that*

$$\phi(r_1,\ldots,r_n) \in [\phi(r_{\pi(1)},\ldots,r_{\pi(n)}) \cdot \frac{1}{1+\varepsilon}, \phi(r_{\pi(1)},\ldots,r_{\pi(n)}) \cdot (1+\varepsilon)] \text{ and}$$
$$\phi(r_{\pi(1)},\ldots,r_{\pi(n)}) \in [\phi(r_1,\ldots,r_n) \cdot \frac{1}{1+\varepsilon}, \phi(r_1,\ldots,r_n) \cdot (1+\varepsilon)].$$

Proof. The proof follows the same structural argument as [10, Lemma 6]. For any permutation $\pi \in \Pi_{\boldsymbol{C}_\phi}$ of a given assignment $\boldsymbol{r} = (r_1,\ldots,r_n)$, by definition, the symmetric properties $\phi(\boldsymbol{r}) \leq \phi(r_{\pi(1)},\ldots,r_{\pi(n)}) \cdot (1+\varepsilon)$ and $\phi(r_{\pi(1)},\ldots,r_{\pi(n)}) \leq \phi(\boldsymbol{r}) \cdot (1+\varepsilon)$ hold, and therefore also the rearranged inequalities $\phi(\boldsymbol{r}) \cdot \frac{1}{1+\varepsilon} \leq \phi(r_{\pi(1)},\ldots,r_{\pi(n)})$ and $\phi(r_{\pi(1)},\ldots,r_{\pi(n)}) \cdot \frac{1}{1+\varepsilon} \leq \phi(\boldsymbol{r})$. □

Lemma 4. *Let ϕ be an ε-commutative factor with respect to $\boldsymbol{C}_\phi$, let $\mathcal{X}_{\mathcal{R}_\phi} = \dot{\bigcup}_{j=1}^{l} \{\boldsymbol{r}^j\}$ be a set of disjoint representatives $\boldsymbol{r}^j$ for all disjoint equivalence classes $\dot{\bigcup}_{j=1}^{l} [\boldsymbol{r}^j]_\phi = \dot{\bigcup}_{j=1}^{l} \mathfrak{S}_\phi(\boldsymbol{r}^j) = \mathcal{X}_{\boldsymbol{R}_\phi}$ ordered by size such that $m_1 := |[\boldsymbol{r}^1]_\phi| \geq \ldots \geq m_l := |[\boldsymbol{r}^l]_\phi| \geq 1$, and let ϕ^* be the symmetrised representative factor via the mean as in Eq. (2). Then,*

$$1 \;\leq\; \frac{\max_{\boldsymbol{r}} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})}}{\min_{\boldsymbol{r}} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})}} \;\leq\; \frac{\left(1 + \frac{m_2-1}{m_2}\varepsilon\right)(1+\varepsilon)}{1 + \frac{1}{m_1}\varepsilon}.$$

Proof. A direct worst-case analysis requires some care. Although the extrema $\max_{\boldsymbol{r}} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})}$ and $\min_{\boldsymbol{r}} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})}$ are taken independently over the entire assignment space, the mean value $\phi^*(\boldsymbol{r})$ is uniquely determined within each equivalence class $[\boldsymbol{r}^j]_\phi$. Consequently, the extremal configurations cannot always be chosen independently, and we therefore distinguish several cases.

First observe that for every equivalence class $[\boldsymbol{r}^j]_\phi$ and all $\boldsymbol{r}', \boldsymbol{r}'' \in [\boldsymbol{r}^j]_\phi$, Lemma 3 guarantees pairwise ε-equivalence and Lemma 1 implies $\phi(\boldsymbol{r}') \in [\phi(\boldsymbol{r}'') \cdot \frac{1}{1+\varepsilon}, \phi(\boldsymbol{r}'') \cdot (1+\varepsilon)]$ and $\phi(\boldsymbol{r}'') \in [\phi(\boldsymbol{r}') \cdot \frac{1}{1+\varepsilon}, \phi(\boldsymbol{r}') \cdot (1+\varepsilon)]$. Hence, the potentials within one equivalence class can be ordered as

$$0 < \frac{1}{1+\varepsilon} \cdot \phi(\boldsymbol{r}^{j,\max}) \leq \phi(\boldsymbol{r}^{j,\min}) \leq \phi(\boldsymbol{r}'') \leq \phi(\boldsymbol{r}^{j,\max}) \leq \phi(\boldsymbol{r}^{j,\min}) \cdot (1+\varepsilon),$$

where $\phi(\boldsymbol{r}^{j,\max}) := \max_{\boldsymbol{r}' \in [\boldsymbol{r}^j]_\phi} \phi(\boldsymbol{r}')$ and $\phi(\boldsymbol{r}^{j,\min}) := \min_{\boldsymbol{r}' \in [\boldsymbol{r}^j]_\phi} \phi(\boldsymbol{r}')$. This directly affects the class-wise mean

$$\phi^*(\boldsymbol{r}'') = \overline{\phi}(\boldsymbol{r}'') = \frac{1}{|[\boldsymbol{r}^j]_\phi|} \sum_{\boldsymbol{r}' \in [\boldsymbol{r}^j]_\phi} \phi(\boldsymbol{r}'),$$

which is constrained to vary only within a restricted range of values. Consider first the case where the assignments attaining the maximum and minimum originate from two different equivalence classes $[\boldsymbol{r}^i]_\phi$ and $[\boldsymbol{r}^j]_\phi$ with $i \neq j$. In this situation, the worst-case deviations can be chosen independently.

For the maximum, the denominator should be as small as possible. Hence, choose $\phi(\boldsymbol{r}) = \phi(\boldsymbol{r}^{i,\min})$ for the denominator, while all remaining potentials in the same equivalence class attain their maximal admissible value $\phi(\boldsymbol{r}') := \phi(\boldsymbol{r}^{i,\max}) = (1+\varepsilon)\phi(\boldsymbol{r}^{i,\min})$ for all $\boldsymbol{r}' \in [\boldsymbol{r}^i]_\phi \setminus \{\boldsymbol{r}\}$ to maximise the numerator. This yields

$$\begin{aligned}\max_{\boldsymbol{r}\in[\boldsymbol{r}^i]_\phi} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})} &= \frac{(\phi(\boldsymbol{r}^{i,\min}) + (|[\boldsymbol{r}^i]_\phi| - 1)\cdot(1+\varepsilon)\cdot\phi(\boldsymbol{r}^{i,\min})) \,/\, |[\boldsymbol{r}^i]_\phi|}{\phi(\boldsymbol{r}^{i,\min})} \\ &= \frac{1 + (|[\boldsymbol{r}^i]_\phi| - 1)(1+\varepsilon)}{|[\boldsymbol{r}^i]_\phi|} = 1 + \frac{m_i - 1}{m_i}\,\varepsilon.\end{aligned}$$

For the minimum, the denominator should instead be as large as possible. Thus, choose $\phi(\boldsymbol{r}) = \phi(\boldsymbol{r}^{j,\max})$, while all remaining values in the same equivalence class attain their minimal admissible value $\phi(\boldsymbol{r}') := \phi(\boldsymbol{r}^{j,\min}) = \frac{1}{1+\varepsilon}\phi(\boldsymbol{r}^{j,\max})$ for all $\boldsymbol{r}' \in [\boldsymbol{r}^j]_\phi \setminus \{\boldsymbol{r}\}$. Consequently,

$$\begin{aligned}\min_{\boldsymbol{r}\in[\boldsymbol{r}^j]_\phi} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})} &= \frac{((|[\boldsymbol{r}^j]_\phi| - 1)\cdot\frac{1}{1+\varepsilon}\cdot\phi(\boldsymbol{r}^{j,\max}) + \phi(\boldsymbol{r}^{j,\max})) \,/\, |[\boldsymbol{r}^j]_\phi|}{\phi(\boldsymbol{r}^{j,\max})} \\ &= \frac{(|[\boldsymbol{r}^j]_\phi| - 1)\cdot\frac{1}{1+\varepsilon} + 1}{|[\boldsymbol{r}^j]_\phi|} = \frac{|[\boldsymbol{r}^j]_\phi| + \varepsilon}{(1+\varepsilon)\cdot|[\boldsymbol{r}^j]_\phi|} = \frac{1 + \frac{1}{m_j}\varepsilon}{1+\varepsilon}.\end{aligned}$$

Observe, that $1 + \frac{m_i - 1}{m_i}\,\varepsilon$ is monotonically increasing in m_i, whereas $\frac{1+\frac{1}{m_j}\varepsilon}{1+\varepsilon}$ is monotonically decreasing in m_j. Hence, the largest equivalence classes by group size produce the extremal deviations. To determine which class size appear in the outer quotient as the numerator and denominator, respectively, consider $g(x) := \left(1 + \frac{x-1}{x}\,\varepsilon\right)\left(1 + \frac{1}{x}\varepsilon\right)$. Its derivative equals $g'(x) = \frac{\varepsilon^2(2-x)}{x^3}$, which is non-positive for all $x \geq 2$. Therefore, for $x_1, x_2 \in \mathbb{N}_{\geq 2}$ it follows

$$\begin{aligned}&\left(1 + \frac{x_1 - 1}{x_1}\,\varepsilon\right)\left(1 + \frac{1}{x_1}\varepsilon\right) = g(x_1) \geq g(x_2) = \left(1 + \frac{x_2 - 1}{x_2}\,\varepsilon\right)\left(1 + \frac{1}{x_2}\varepsilon\right) \\ \Leftrightarrow\quad &\frac{1 + \frac{x_1 - 1}{x_1}\,\varepsilon}{1 + \frac{1}{x_2}\varepsilon} \geq \frac{1 + \frac{x_2 - 1}{x_2}\,\varepsilon}{1 + \frac{1}{x_1}\varepsilon}.\end{aligned}$$

Consequently, the sharpest bound is obtained by using m_2 for the maximum term and m_1 for the minimum term. If several equivalence classes with maximal group size exist, then simply $m_1 = m_2$.

It remains to analyse the case where both extrema originate from the same equivalence class. Since $\phi^*(\boldsymbol{r}) = \phi^*(\boldsymbol{r}')$ for all $\boldsymbol{r}, \boldsymbol{r}' \in [\boldsymbol{r}^j]_\phi$, we obtain

$$\frac{\max_{\boldsymbol{r}\in[\boldsymbol{r}^j]_\phi} \frac{\phi^*(\boldsymbol{r})}{\phi(\boldsymbol{r})}}{\min_{\boldsymbol{r}'\in[\boldsymbol{r}^j]_\phi} \frac{\phi^*(\boldsymbol{r}')}{\phi(\boldsymbol{r}')}} = \frac{\max_{\boldsymbol{r}\in[\boldsymbol{r}^j]_\phi} \frac{1}{\phi(\boldsymbol{r})}}{\min_{\boldsymbol{r}'\in[\boldsymbol{r}^j]_\phi} \frac{1}{\phi(\boldsymbol{r}')}}.$$

If the equivalence class contains only a single assignment, the quotient equals 1. Otherwise, $m_1 \geq m_2 \geq 2$ holds, and the quotient is maximised by choosing

$\phi(\boldsymbol{r}) = \phi(\boldsymbol{r}^{j,\min}) = \frac{1}{1+\varepsilon}\phi(\boldsymbol{r}^{j,\max})$ and $\phi(\boldsymbol{r}') = \phi(\boldsymbol{r}^{j,\max}) = (1+\varepsilon)\phi(\boldsymbol{r}^{j,\min})$. Hence, for $m_2 \leq m_1$ we get

$$\frac{\max_{\boldsymbol{r}\in[\boldsymbol{r}^j]_\phi} \frac{1}{\phi(\boldsymbol{r})}}{\min_{\boldsymbol{r}'\in[\boldsymbol{r}^j]_\phi} \frac{1}{\phi(\boldsymbol{r}')}} \leq \frac{\frac{1}{\phi(\boldsymbol{r}^{j,\min})}}{\frac{1}{(1+\varepsilon)\phi(\boldsymbol{r}^{j,\min})}} = 1+\varepsilon \leq \frac{\left(1+\frac{m_2-1}{m_2}\varepsilon\right)(1+\varepsilon)}{1+\frac{1}{m_1}\varepsilon},$$

which establishes the bound in all cases. □

Theorem 5. *The bound given in Thm. 4 is optimal.*

For the proof, consider the following example.

Example 3. Let M be the FG with factors $\phi_1, \ldots, \phi_m$, of which $\phi_1, \ldots, \phi_k$ with $0 \leq k \leq m$ are ε-commutative with respect to disjoint argument sets $\boldsymbol{C}_{\phi_1}, \ldots, \boldsymbol{C}_{\phi_k}$ with finite domains, respectively, each satisfying $|\boldsymbol{C}_{\phi_i}| \geq 2$. Let $\mathcal{X}_{\mathcal{R}_{\phi_i}} = \dot{\cup}_{j=1}^{l_i}\{\boldsymbol{r}_i^j\}$ be the set of disjoint representatives $\boldsymbol{r}_i^j$ for all disjoint equivalence classes $[\boldsymbol{r}_i^j]_{\phi_i}$ united to $\dot{\cup}_{j=1}^{l_i}[\boldsymbol{r}_i^j]_{\phi_i} = \mathcal{X}_{\boldsymbol{R}_{\phi_i}}$ ordered by group size with $m_{1_i} := |[\boldsymbol{r}_i^1]_{\phi_i}| \geq \ldots \geq m_{l_i} := |[\boldsymbol{r}_i^l]_{\phi_i}| \geq 1$ for every ε-commutative factor ϕ_i with $i = 1, \ldots, k$. Let $\mathcal{X}_{R_j} = \text{range}(R_j)$ have at least two deviating elements and within a group $\boldsymbol{C}_{\phi_i}$ the same ranges. For $i = 1, \ldots, k$ define

$$\phi_i(\boldsymbol{r}_i) := \begin{cases} (1+\varepsilon)\cdot 1 & \text{if } \boldsymbol{r}_i = \boldsymbol{r}_i^1, \\ 1 & \text{if } \boldsymbol{r}_i \in [\boldsymbol{r}_i^1]_{\phi_i} \setminus \{\boldsymbol{r}_i^1\}, \\ 2 & \text{if } \boldsymbol{r}_i = \boldsymbol{r}_i^2, \\ (1+\varepsilon)\cdot 2 & \text{if } \boldsymbol{r}_i \in [\boldsymbol{r}_i^2]_{\phi_i} \setminus \{\boldsymbol{r}_i^2\}, \\ j & \text{for } \boldsymbol{r}_i \in [\boldsymbol{r}_i^j]_{\phi_i} \text{ and } 2 < j \leq l_i. \end{cases}$$

Thus, the first equivalence class contains exactly one maximal value and otherwise minimal values, whereas the second class contains exactly one minimal value and otherwise maximal ones, following the same construction as in the proof of Lemma 4. All remaining parts are constant and therefore contribute no approximation error.

Proof of Thm. 5. We show that the construction from Ex. 3 attains the bound. Let M' denote the model obtained from M by replacing every ε-commutative factor ϕ_i by the arithmetic mean as in Eq. (2). Notice that $\phi_i^*(\boldsymbol{r}_i) = \phi_i(\boldsymbol{r}_i)$ for $i = k+1, \ldots, m$ and all $\boldsymbol{r}_i \in \mathcal{X}_{\boldsymbol{R}_{\phi_i}}$, but also $\phi_i^*(\boldsymbol{r}_i) = \phi_i(\boldsymbol{r}_i)$ for $i = 1, \ldots, k$ and $\boldsymbol{r}_i \in [\boldsymbol{r}_i^j]_{\phi_i}$ and $2 < j \leq l_i$ implying $\frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)} = 1$. Consequently, all extremal deviation are attained exclusively within the first two equivalence classes. For $\boldsymbol{r}_i \in [\boldsymbol{r}_i^1]_{\phi_i}$, the definition results in

$$\phi_i^*(\boldsymbol{r}_i) = \frac{(1+\varepsilon) + (m_{1_i} - 1)}{m_{1_i}} = 1 + \frac{1}{m_{1_i}}\varepsilon,$$

and for $\boldsymbol{r}_i' \in [\boldsymbol{r}_i^2]_{\phi_i}$ in

$$\phi_i^*(\boldsymbol{r}_i') = \frac{(1+\varepsilon)\cdot 2\cdot(m_{2_i} - 1) + 2}{m_{2_i}} = 2\cdot\left(1 + \frac{(m_{2_i} - 1)}{m_{2_i}}\varepsilon\right).$$

For $\boldsymbol{r}_i \in [\boldsymbol{r}_i^1]_{\phi_i}$ the maximum becomes

$$\max_{\boldsymbol{r}_i \in [\boldsymbol{r}_i^1]_{\phi_i}} \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)} = \frac{1 + \frac{1}{m_{1_i}}\varepsilon}{1} = 1 + \frac{1}{m_{1_i}}\varepsilon > 1,$$

and for $\boldsymbol{r}_i' \in [\boldsymbol{r}_i^2]_{\phi_i}$ it becomes

$$\max_{\boldsymbol{r}_i' \in [\boldsymbol{r}_i^2]_{\phi_i}} \frac{\phi_i^*(\boldsymbol{r}_i')}{\phi_i(\boldsymbol{r}_i')} = \frac{2 \cdot \left(1 + \frac{(m_{2_i}-1)}{m_{2_i}}\varepsilon\right)}{2} = 1 + \frac{(m_{2_i}-1)}{m_{2_i}}\varepsilon > 1,$$

which is larger than the first one due to $m_{2_i} \geq 2$.

Analogously, for $\boldsymbol{r}_i \in [\boldsymbol{r}_i^1]_{\phi_i}$ the minimal quotient inside this class reaches

$$\min_{\boldsymbol{r}_i \in [\boldsymbol{r}_i^1]_{\phi_i}} \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)} = \frac{1 + \frac{1}{m_{1_i}}\varepsilon}{1 + \varepsilon} < 1,$$

which is as small as the following one for $\boldsymbol{r}_i' \in [\boldsymbol{r}_i^2]_{\phi_i}$ due to $m_{2_i} \geq 2$

$$\min_{\boldsymbol{r}_i' \in [\boldsymbol{r}_i^2]_{\phi_i}} \frac{\phi_i^*(\boldsymbol{r}_i')}{\phi_i(\boldsymbol{r}_i')} = \frac{2 \cdot \left(1 + \frac{(m_{2_i}-1)}{m_{2_i}}\varepsilon\right)}{2 \cdot (1+\varepsilon)} = \frac{1 + \frac{(m_{2_i}-1)}{m_{2_i}}\varepsilon}{1 + \varepsilon} < 1.$$

Using Eq. (4) from the definition of the Chan-Darwiche distance leads to

$$\begin{aligned}
D_{CD}(P_M, P_{M'}) &= \ln \max_{\boldsymbol{r}} \frac{\psi'(\boldsymbol{r})}{\psi(\boldsymbol{r})} - \ln \min_{\boldsymbol{r}} \frac{\psi'(\boldsymbol{r})}{\psi(\boldsymbol{r})} \\
&= \ln \max_{\boldsymbol{r}} \frac{\prod_{i=1}^k \phi_i^*(\boldsymbol{r}_i) \prod_{i=k+1}^m \phi_i(\boldsymbol{r}_i)}{\prod_{i=1}^m \phi_i(\boldsymbol{r}_i)} - \ln \min_{\boldsymbol{r}} \frac{\prod_{i=1}^k \phi_i^*(\boldsymbol{r}_i) \prod_{i=k+1}^m \phi_i(\boldsymbol{r}_i)}{\prod_{i=1}^m \phi_i(\boldsymbol{r}_i)} \\
&= \ln \max_{\boldsymbol{r}} \prod_{i=1}^k \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)} - \ln \min_{\boldsymbol{r}} \prod_{i=1}^k \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)}.
\end{aligned}$$

Since the argument sets are pairwise disjoint, the extrema can be attained independently for every factor. Hence,

$$\begin{aligned}
D_{CD}(P_M, P_{M'}) &= \ln \prod_{i=1}^k \max_{\boldsymbol{r}_i} \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)} - \ln \prod_{i=1}^k \min_{\boldsymbol{r}_i} \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)} \\
&= \ln \prod_{i=1}^k \max_{\boldsymbol{r}_i' \in [\boldsymbol{r}_i^2]_{\phi_i}} \frac{\phi_i^*(\boldsymbol{r}_i')}{\phi_i(\boldsymbol{r}_i')} - \ln \prod_{i=1}^k \min_{\boldsymbol{r}_i \in [\boldsymbol{r}_i^1]_{\phi_i}} \frac{\phi_i^*(\boldsymbol{r}_i)}{\phi_i(\boldsymbol{r}_i)} \\
&= \ln \prod_{i=1}^k \left(1 + \frac{(m_{2_i}-1)}{m_{2_i}}\varepsilon\right) - \ln \prod_{i=1}^k \frac{1 + \frac{1}{m_{1_i}}\varepsilon}{1+\varepsilon} \\
&= \ln \prod_{i=1}^k \frac{\left(1 + \frac{m_{2_i}-1}{m_{2_i}}\varepsilon\right)(1+\varepsilon)}{1 + \frac{1}{m_{1_i}}\varepsilon}.
\end{aligned}$$

Thus, the bound of Thm. 4 is attained exactly and is therefore optimal. □

A.1 Additional Result Symmetry Set

Proposition 2 (Cardinality of the Symmetry Set). *Let* $\boldsymbol{r} = (r_1, \ldots, r_n) \in \mathcal{X}_{\boldsymbol{R}_\phi}$ *and* $I_{\boldsymbol{C}_\phi} := \{\, i \mid R_i \in \boldsymbol{C}_\phi \,\}$ *denote the indices of the commutative arguments. Further, let* $\{a_1, \ldots, a_k\} = \{\, r_i \mid i \in I_{\boldsymbol{C}_\phi} \,\}$ *be the set of distinct values occurring among the commutative arguments, and let* $n_j := \left|\{\, i \in I_{\boldsymbol{C}_\phi} \mid r_i = a_j \,\}\right|$ *denote the multiplicity of value* a_j *for a fixed assignment* $\boldsymbol{r}$*. Then the cardinality of the symmetry set is*

$$|\mathfrak{S}_\phi(\boldsymbol{r})| = \frac{|\boldsymbol{C}_\phi|!}{\prod_{j=1}^{k} n_j!}$$

with $1 \leq |\mathfrak{S}_\phi(\boldsymbol{r})| \leq |\boldsymbol{C}_\phi|!$*, where the upper bound is attained if and only if all commutative arguments have pairwise non-*ε*-equivalent values.*

Proof. Since permutations in $\Pi_{\boldsymbol{C}_\phi}$ act only on the positions corresponding to ε-commutative arguments, it holds that $|\Pi_{\boldsymbol{C}_\phi}| = |\boldsymbol{C}_\phi|!$.

If all values among the commutative arguments are pairwise distinct, every permutation produces a distinct tuple, and thus $|\mathfrak{S}_\phi(\boldsymbol{r})| = |\boldsymbol{C}_\phi|!$.

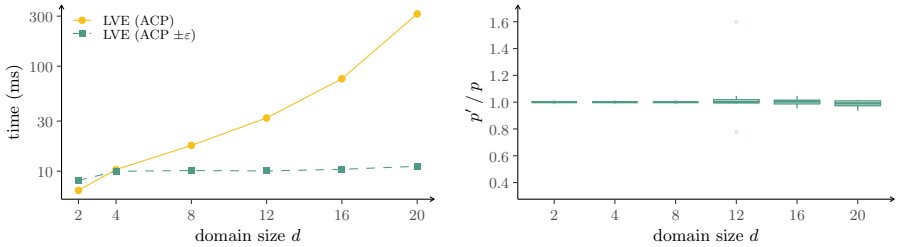
Now assume that some values coincide. For each distinct value a_j, there are exactly m_j positions among the commutative arguments whose assigned value equals a_j. Permuting these m_j positions among themselves does not change the resulting tuple. Hence, every distinct tuple in the symmetry set is generated exactly $\prod_{j=1}^{k} n_j!$ times by permutations in $\Pi_{\boldsymbol{C}_\phi}$. Since there are $|\boldsymbol{C}_\phi|!$ permutations in total, the number of distinct tuples is

$$|\mathfrak{S}_\phi(\boldsymbol{r})| = \frac{|\boldsymbol{C}_\phi|!}{\prod_{j=1}^{k} n_j!}. \qquad \square$$

B Additional Experimental Results

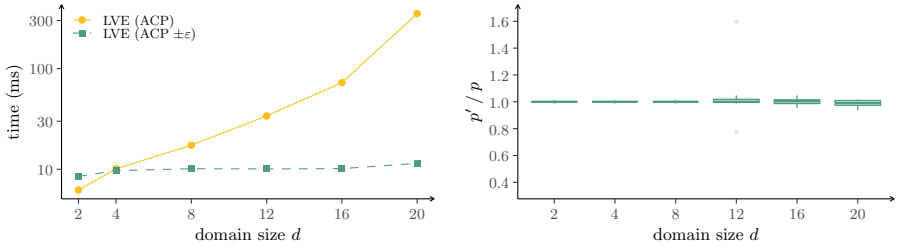
In addition to the experimental results provided in Sec. 4, we provide further experimental results for individual scenarios in this section.

The averaged results in Fig. 2 aggregate over the number of ε-commutative factors k and the tolerance ε. To disentangle the influence of these two parameters, Figs. 3 to 11 report the LVE run time and the per-query quotient $p' \,/\, p$ separately for each combination of $k \in \{1, 3, 7\}$ and $\varepsilon \in \{0.001, 0.01, 0.1\}$. The run time behaviour is qualitatively consistent across all configurations: The run time of LVE on the ε-relaxed model stays almost constant in the domain size, whereas exact ACP, which must leave the ε-commutative factors uncompressed, slows down steeply, so the speedup grows exponentially with the domain size d. The speedup is essentially independent of ε (given that the ε-deviation in the model is within the chosen ε-value used when running ACP, which is the case in our experiments), since the structure of the compressed model depends only on the commutativity pattern and not on the magnitude of ε, while it increases



(a) LVE run time (ms). (b) Query result quotient p' / p.

Fig. 3: (a) Run time of LVE on the compressed model returned by ACP and its ε-relaxed variant (ACP $\pm\varepsilon$), and (b) distribution of the per-query quotient p' / p between the approximate result p' and the exact result p for input FGs containing $k = 1$ ε-commutative factors with $\varepsilon = 0.001$.

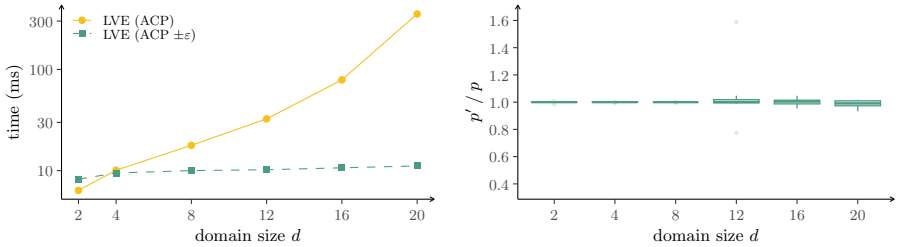


(a) LVE run time (ms). (b) Query result quotient p' / p.

Fig. 4: (a) Run time of LVE on the compressed model returned by ACP and its ε-relaxed variant (ACP $\pm\varepsilon$), and (b) distribution of the per-query quotient p' / p between the approximate result p' and the exact result p for input FGs containing $k = 1$ ε-commutative factors with $\varepsilon = 0.01$.

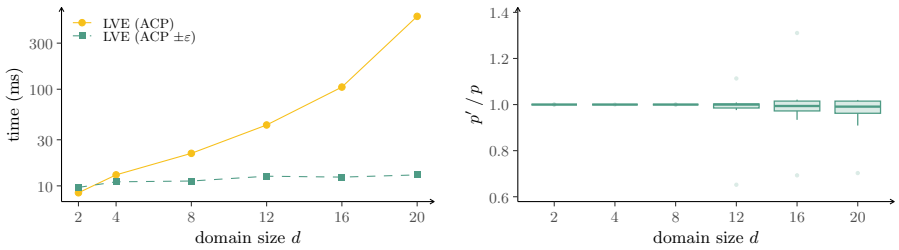
with the number of ε-commutative factors k, as each additional compressed factor widens the gap between the two models. At the same time, the per-query quotient remains tightly concentrated around the optimal value of one in every scenario. In line with the theoretical bound, increasing ε from 0.001 to 0.1 slightly widens the spread of the quotient, and increasing the number of ε-commutative factors k admits a few more outliers, since each compressed factor potentially contributes to the overall deviation. Even for the largest tested $k = 7$ and for the largest tested $\varepsilon = 0.1$, however, the quotient stays well within the theoretical guarantee, confirming that the bound is loose in practice.

A plot showcasing the resulting speedup of online inference (averaged over all choices of k and ε) is given in Fig. 12a. It becomes evident that the speedup grows exponentially with the domain size d and eventually reaches a factor of more than 30 for the choice of $d = 20$. Figure 12b further displays the offline run times of ACP and its ε-relaxed variant (averaged over all choices of k and ε). As expected, the offline run time of the ε-relaxed variant does not surpass the offline



(a) LVE run time (ms). (b) Query result quotient $p' \,/\, p$.

Fig. 5: (a) Run time of LVE on the compressed model returned by ACP and its ε-relaxed variant (ACP $\pm\varepsilon$), and (b) distribution of the per-query quotient $p' \,/\, p$ between the approximate result p' and the exact result p for input FGs containing $k = 1$ ε-commutative factors with $\varepsilon = 0.1$.

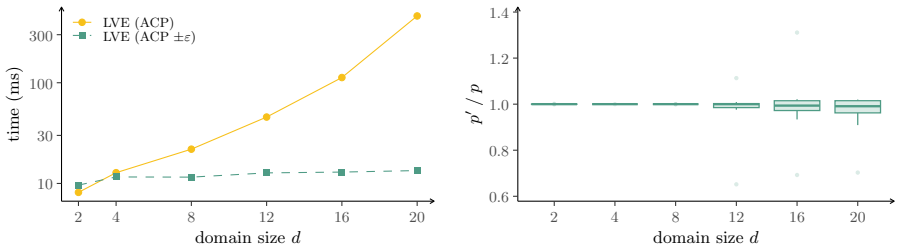


(a) LVE run time (ms). (b) Query result quotient $p' \,/\, p$.

Fig. 6: (a) Run time of LVE on the compressed model returned by ACP and its ε-relaxed variant (ACP $\pm\varepsilon$), and (b) distribution of the per-query quotient $p' \,/\, p$ between the approximate result p' and the exact result p for input FGs containing $k = 3$ ε-commutative factors with $\varepsilon = 0.001$.

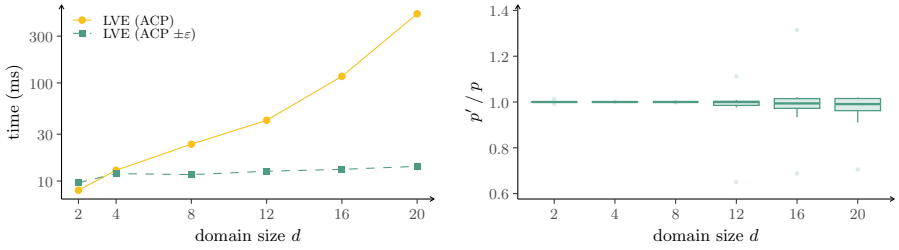
run time of exact ACP, demonstrating that the relaxation does not introduce any overhead at all. In fact, ε-relaxed ACP even slightly reduces the run time compared to exact ACP, since ε-relaxed ACP is able to find ε-commutative factors and hence requires less iterations when searching for subsets of ε-commutative arguments (ACP needs more iterations as it checks all subsets of arguments until it is able to conclude that no subset of strictly commutative arguments exists).

Finally, we examine the structural reduction achieved by the compression stemming from exact ACP and its ε-relaxed variant. Figures 13 to 15 show the compression ratio $|M'| \,/\, |M|$ between the compressed model M' and the original model M for $k \in \{1, 3, 7\}$ at the smallest and largest tolerance $\varepsilon \in \{0.001, 0.1\}$ (that is, a compression ratio of one means that no compression is achieved at all, whereas a smaller ratio indicates more compression). The ε-relaxed variant attains a substantially smaller compression ratio than exact ACP, because it is able to compress the ε-commutative factors that exact ACP cannot compress.



(a) LVE run time (ms). (b) Query result quotient p' / p.

Fig. 7: (a) Run time of LVE on the compressed model returned by ACP and its ε-relaxed variant (ACP $\pm\varepsilon$), and (b) distribution of the per-query quotient p' / p between the approximate result p' and the exact result p for input FGs containing $k = 3$ ε-commutative factors with $\varepsilon = 0.01$.

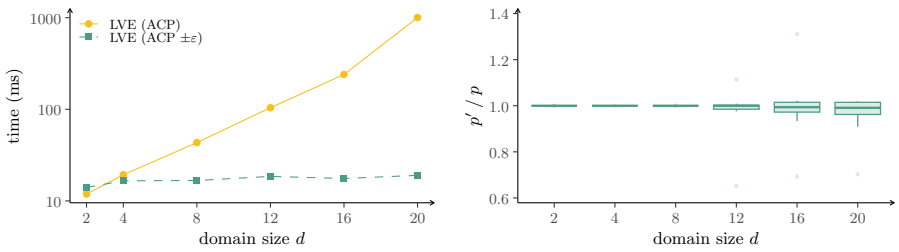


(a) LVE run time (ms). (b) Query result quotient p' / p.

Fig. 8: (a) Run time of LVE on the compressed model returned by ACP and its ε-relaxed variant (ACP $\pm\varepsilon$), and (b) distribution of the per-query quotient p' / p between the approximate result p' and the exact result p for input FGs containing $k = 3$ ε-commutative factors with $\varepsilon = 0.1$.

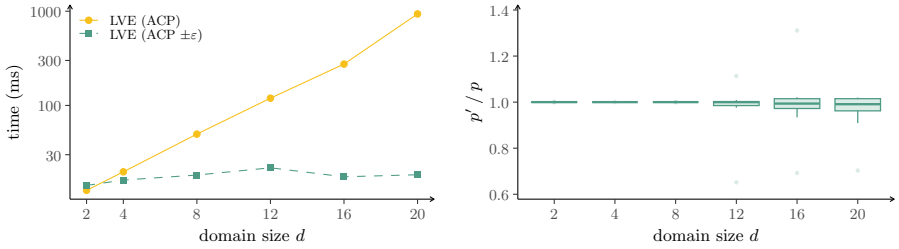
For both variants, the compression ratio decreases as the domain size grows, illustrating that larger models offer more indistinguishability to be exploited.

Interestingly, as shown in Fig. 16, ACP is not able to obtain any compression at all for some of the input FGs: The input instances are generated from two different classes of FGs, *employee* and *epidemic*, which have different structural properties, and exact ACP is only able to achieve compression for the *employee* instance class, while it fails to achieve any compression for the *epidemic* instance class. This particular behaviour highlights that even if there is just a single ε-commutative factor in a given FG, exact ACP may fail to find any compression at all (although there are other parts in the input FG that could be compressed and that do not involve commutativity), which is due to the propagation of information throughout ACP's colour passing procedure.



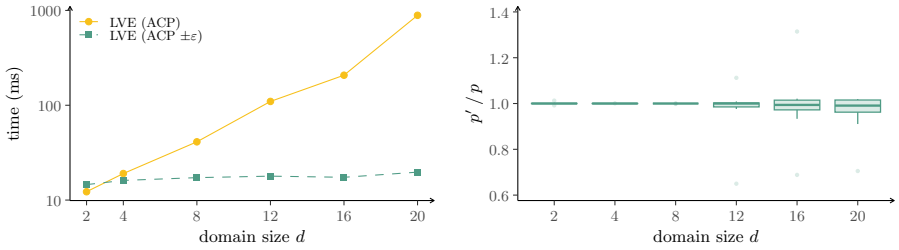
(a) LVE run time (ms). (b) Query result quotient p' / p.

Fig. 9: (a) Run time of LVE on the compressed model returned by ACP and its ε-relaxed variant (ACP $\pm\varepsilon$), and (b) distribution of the per-query quotient p' / p between the approximate result p' and the exact result p for input FGs containing $k = 7$ ε-commutative factors with $\varepsilon = 0.001$.



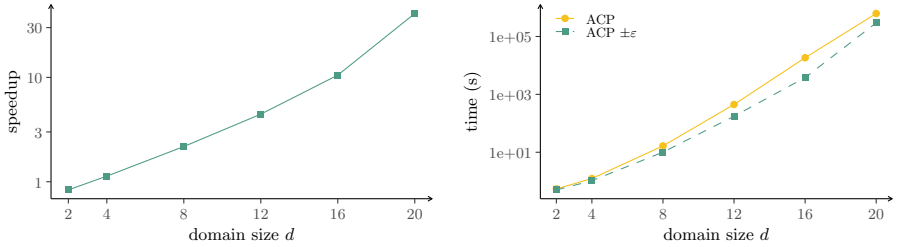
(a) LVE run time (ms). (b) Query result quotient p' / p.

Fig. 10: (a) Run time of LVE on the compressed model returned by ACP and its ε-relaxed variant (ACP $\pm\varepsilon$), and (b) distribution of the per-query quotient p' / p between the approximate result p' and the exact result p for input FGs containing $k = 7$ ε-commutative factors with $\varepsilon = 0.01$.



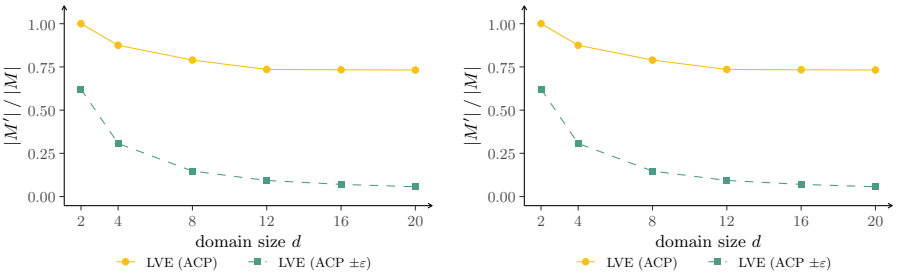
(a) LVE run time (ms). (b) Query result quotient p' / p.

Fig. 11: (a) Run time of LVE on the compressed model returned by ACP and its ε-relaxed variant (ACP $\pm\varepsilon$), and (b) distribution of the per-query quotient p' / p between the approximate result p' and the exact result p for input FGs containing $k = 7$ ε-commutative factors with $\varepsilon = 0.1$.



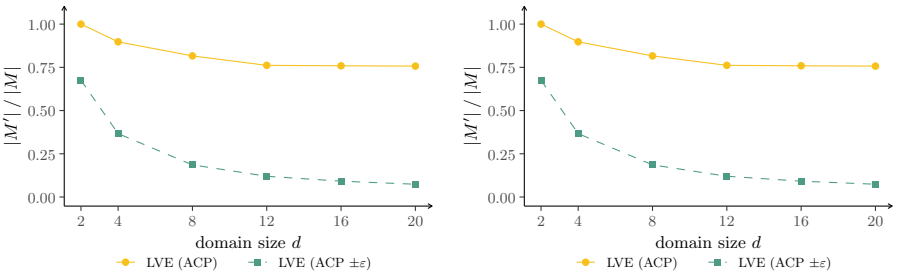
(a) Speedup factor for LVE. (b) Color passing run times.

Fig. 12: (a) Speedup for online inference run times (i.e., ratio of the run time of LVE on the compressed model returned by ACP and the run time of LVE on the compressed model returned by the ε-relaxed variant of ACP). (b) Offline run times of ACP and its ε-relaxed variant (ACP $\pm\varepsilon$).



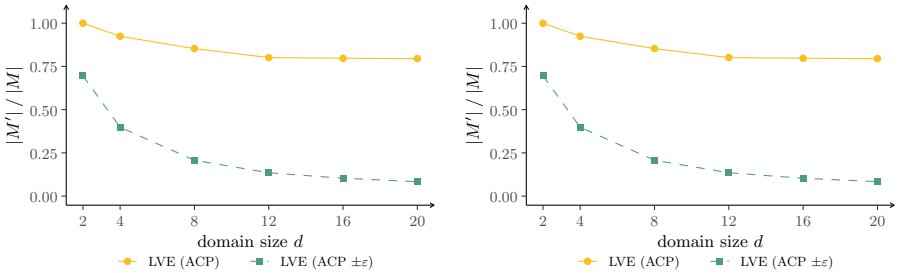
(a) Compression ratio for $\varepsilon = 0.001$. (b) Compression ratio for $\varepsilon = 0.1$.

Fig. 13: Compression ratio $|M'|\,/\,|M|$ between the compressed model M' and the original model M for input FGs containing $k = 1$ ε-commutative factors with (a) $\varepsilon = 0.001$ and (b) $\varepsilon = 0.1$.



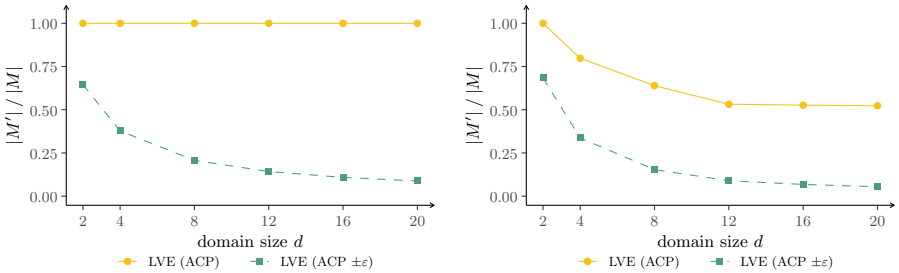
(a) Compression ratio for $\varepsilon = 0.001$. (b) Compression ratio for $\varepsilon = 0.1$.

Fig. 14: Compression ratio $|M'|\,/\,|M|$ between the compressed model M' and the original model M for input FGs containing $k = 3$ ε-commutative factors with (a) $\varepsilon = 0.001$ and (b) $\varepsilon = 0.1$.



(a) Compression ratio for $\varepsilon = 0.001$. (b) Compression ratio for $\varepsilon = 0.1$.

Fig. 15: Compression ratio $|M'|\,/\,|M|$ between the compressed model M' and the original model M for input FGs containing $k = 7$ ε-commutative factors with (a) $\varepsilon = 0.001$ and (b) $\varepsilon = 0.1$.



(a) *employee* instance class. (b) *epidemic* instance class.

Fig. 16: Compression ratio $|M'|\,/\,|M|$ between the compressed model M' and the original model M broken down per instance class, averaged over the number of commutative factors k and the tolerance ε.