Ising Model on Locally Tree-like Graphs: Uniqueness of Solutions to Cavity Equations

Qian Yu* and Yury Polyanskiy†
*Department of Electrical and Computer Engineering, Princeton University
†Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology

Abstract

In the study of Ising models on large locally tree-like graphs, in both rigorous and non-rigorous methods one is often lead to understanding the so-called belief propagation distributional recursions and its fixed point (also known as Bethe fixed point, cavity equation etc). In this work we prove there is at most one non-trivial fixed point for Ising models with zero or random (but "unbised") external fields. Previously this was only known for sufficiently “low-temperature” models. Our proof consists of constructing a metric under which the BP operator is a contraction (albeit non-multiplicative). This is achieved by introducing a concept of degradation index and proving a strengthening of the stringy tree lemma from [9].

This simultaneously closes the following 6 conjectures in the literature:
1) uselessness of global information for a labeled 2-community stochastic block model, or 2-SBM (Kanade-Mossel-Schramm’2014);
2) optimality of local algorithms for 2-SBM under noisy side information (Mossel-Xu’2015);
3) uniqueness of BP fixed point in broadcasting on trees with large degree limit (ibid);
4) independence of robust reconstruction accuracy to leaf noise in broadcasting on trees (Mossel-Neeman-Sly’2016);
5) boundary irrelevance in BOT (Abbe-Cornacchia-Gu-P.’2021);
6) characterization of entropy of community labels given the graph in 2-SBM (ibid).

I. INTRODUCTION

Consider an inference problem for the Ising model on infinite trees. We have an infinite rooted tree that is generated recursively, such that each vertex $v$ has an i.i.d. number of children sampled from a given degree distribution $P_d$. Each vertex $v$ is associated with a binary random variable $X_v$, and their joint distribution is defined based on a similar recursion. The variable on the root, denoted by $X_0$, is $Ber(\frac{1}{2})$. For any other vertex $v$, $X_v$ is identical to the variable on its parent with probability $1-\delta$, conditioned on all other variables that are not their descendant, given some fixed parameter $\delta \in [0,1]$.

We are interested in the process of estimating $X_0$ given the tree structure and the collection of variables on all vertices with a level of $h$, in the limit of $h \to \infty$. Let $O_h$ denote the collection of all observed information, which is random. We can consider the distribution of the log-likelihood ratio (LLR) for estimating $X_0$ conditioned on $X_0 = 0$, which is defined on $(-\infty, +\infty]$. Formally, we denote the LLR distribution for a finite tree of level $h$ by $\mu_{(h)}$, which is defined as

$$\mu_{(h)}(S) = P\left[\ln \frac{P(O_h|X_0 = 0)}{P(O_h|X_0 = 1)} \in S | X_0 = 0\right]$$

for any measurable $S \subseteq (-\infty, +\infty]$. The LLR distribution plays an important role in the estimation process, as it contains full information about fundamental concepts such as minimum error probability and mutual information.

In this basic setting, each $\mu_{(h)}$ can be determined through Belief Propagation (BP), and they converge to a distribution described by the following properties.

Definition 1 (Symmetric fixed point of BP). For any given degree distribution $P_d$ and parameter $\delta$, we say a distribution $\mu$ defined on domain $(-\infty, +\infty]$ is a symmetric BP fixed point if

$$d\mu(r) = e^r \cdot d\mu(-r),$$

and for $d \sim P_d$, $\tilde{R}_u \sim \mu$, $Y_u \sim (-1)^{Ber(\delta)}$ (all jointly independent), the following random variable also has law $\mu$,

$$R \triangleq \sum_{u=1}^{d} Y_u F_\delta(\tilde{R}_u),$$

where

$$F_\delta(x) \triangleq \ln \frac{(1-\delta)e^x + \delta}{\delta e^x + 1 - \delta} = 2 \tanh^{-1}\left(\frac{(1 - 2\delta) \tanh\left(\frac{x}{2}\right)}{2}\right).$$

Generally, we call a distribution $\mu$ symmetric if equation (2) holds, and we call it non-trivial if $\mu\{0\} < 1$.

For simplicity, let $Q$ denote the operator that represents the recursion process, i.e., $Q$ maps any symmetric distribution to a new distribution following the rule of equation (3).
**Proposition 2.** A distribution \(\mu\) is a BP fixed point if and only if \(Q\mu = \mu\).

The operator \(Q\) is also known as density evolution [1, Section 2.2], Bethe recursion [2, Definition 1.6]. It arises from a so-called cavity method [3], which (non-rigorously, but often correctly) allows one to infer important qualities and quantities of statistical physical systems based on the knowledge of the fixed points of the \(Q\). Correspondingly, the distributional identity \(Q\mu = \mu\) is known as the 1RSB cavity equation, with Parisi parameter set to \(x = 1\) [4, Section 14.6, (19.72)]. Evidently, uniqueness of a fixed point has implications for all of these.

This formulation has been widely studied in statistical physics [5], [6], evolutionary biology [7], [8], and information theory [9], [10]. The condition of existence for non-trivial symmetric BP fixed points has been determined exactly, which can be described using the branching number [9]. However, identifying the uniqueness of non-trivial symmetric BP fixed points is a long-unsolved open problem. In this work, we completely resolve this problem by proving the following theorem.\(^1\)

**Theorem 3** (Uniqueness of Symmetric BP fixed points). For any fixed degree distribution \(P_d\) and parameter \(\delta \in [0, 1]\), there is at most one unique non-trivial symmetric BP fixed point.

**Remark 1.** The BP recursion and the cavity equation can be defined for asymmetric distributions. In particular, for any binary hypothesis testing problem and any LLR distribution \(\mu\) that is defined in the form of equation (1), the following equation specifies a unique distribution \(\mu^-\) on \([-\infty, +\infty)\),

\[
d\mu^-(r) = e^r d\mu^-(r),
\]

which can be viewed as the law of the same LLR variable generated by \(X_0 = 1\). For any \(\mu\) satisfying this condition, \(Q\mu\) is defined to be the law of

\[
R \overset{d}{=} \sum_{a=1}^d F_\delta(\hat{R}_a),
\]

where \(\hat{R}_a \overset{iid}{=} (1 - \delta)\mu + \delta \mu^-\), and we call \(\mu\) a BP fixed point if \(Q\mu = \mu\). It can be seen that for symmetric \(\mu\) this definition coincides with the one we gave above.

Our main result implies the uniqueness of BP fixed point for general distributions as well. We present a proof for the following generalized result, as well as its implications, in Appendix B.

**Corollary 4** (Uniqueness of BP fixed points). For any fixed degree distribution \(P_d\) and parameter \(\delta \in [0, 1]\), there is at most one unique non-trivial BP fixed point, and it is symmetric.

There has been several extensions to this basic formulation. The robust reconstruction setting was considered in [11], where the leaf variables observed in the basic setting are instead received through noisy channels. It has been conjectured in [12] that the performance of the maximum likelihood estimator is independent of the noise condition as long as the observation contains non-zero information. We prove this conjecture by showing a stronger statement that the limiting distributions of LLR for robust reconstruction are well-defined and identical to a unique BP fixed point (see Section III-A).

Another variant is to include additional observations to the model, called the survey, that consists of noisy symmetric-channel measurements on vertices above level \(h\) [13]. Similarly, one can compute the LLR distributions through Belief Propagation and define the associated fixed point equations. Identifying the uniqueness of BP fixed points was also an open problem when survey channels are present, and we prove that it holds for all regimes (see Section I-B).

Our results imply complete solutions to several other open problems. Two of them are directly based on the broadcast on trees formulation. First, a concept called boundary irrelevance (BI) was studied in the literature [13], [14], which defines a scenario where the effect of leaf observations become negligible compared to the survey information. BI was conjectured to be true in [14] for binary erasure survey channels, and in [13] for general symmetric survey channels with non-zero capacity. We provide a positive proof of these conjectures, and show that BI holds for any degree distribution (see Section III-B). Second, the uniqueness of BP fixed point has been investigated under a simplified formulation, where the recursion is approximated using central limit theorem when the degrees are large with high probability. For example, see Conjecture 2.6 in [15], where the BP fixed point is represented by a single parameter related to its variance. Our proof techniques directly extend to those regimes, and we show that all earlier stated properties still hold (see Section III-C).

The rest of the resolved conjectures are related to a framework called 2-community stochastic block model, or 2-SBM, where the goal is to estimate a set of hidden labels by observing an associated random graph. The labels are defined on \(n\) vertices, each being i.i.d. Ber(\(\frac{1}{2}\)). The graph is constructed by independently connecting any pair of vertices, with probability \(\alpha\) for pairs with the same labels and with probability \(\beta\) for the rest. In one line of works, the authors of [16], [17], [12], [13] aim to characterize the optimal clustering accuracy and the SBM entropy for a practically important case called the sparse regime. Our leaf-independence results on robust reconstruction and BI lead to complete solutions to these problems, strengthening the accuracy characterization in [12] and resolving the conjecture on entropy characterization in [13] (see Section III-D).

\(^1\)For ease of discussion, in the basic setting we will ignore the trivial cases where \(P[d \leq 1] = 1\).
In another direction, the authors of [14], [15] considered a setting where a noisy version of the labels is provided as side information. It was conjectured that it is optimal to ignore global information and use local belief propagation for estimation, stated in [14], [15] separately for different noise conditions. The proof for such conjectures direct follows from BI. More generally, we present a result that holds for any symmetric noise conditions (see Section III-E).

In the rest of this section, we present the main technical proof steps for uniqueness theorems in the broadcast on tree settings. Then we provide detailed proofs in Section II.

A. Proof Techniques for the Main Theorem

The proof of Theorem 3 builds upon ideas of channel comparison, which were previously in [9] to show certain negative results, and more recently by us [10] for the positive side. Here we extend this methodology in two ways: a) strengthening the stringy tree lemma from [9]; and b) introducing of the concept of degradation index. The latter allows us to define a potential function over symmetric distributions (and LLR distributions in general) that is only stabilized at a unique solution. Further, it also proves BP convergence to BP fixed points for robust reconstruction (see Theorem 35). For clarity, we illustrate the main concepts over symmetric distributions, which enables simplifications compared to their general forms. We first state the definition of degradation.

Definition 5. For any two symmetric distributions \( \mu_Y, \mu_Z \) defined on \( (-\infty, +\infty) \), we say \( \mu_Y \) is a degraded version of \( \mu_Z \), denoted by \( \mu_Y \preceq \mu_Z \), if one can define a joint distribution \( \mu_{Y,Z} \) with \( \mu_Y, \mu_Z \) as marginal distributions, such that \( \mu_{Y|Z} \) is invariant under \( (Y,Z) \rightarrow (Y,-Z) \).

Intuitively, for any \( \mu_Y \preceq \mu_Z, \mu_Z \) can be viewed as the LLR distribution of a symmetric binary hypothesis testing problem and \( \mu_Y \) can be viewed as a noisy version of \( \mu_Z \) where the observation is corrupted by a symmetric noise channel \( \mu_{Y|Z} \). A more detailed discussion on degradation can be found in Appendix A.

Our construction of degradation index is based on the following operator.

Definition 6 (Box Convolution for Symmetric Distributions). For any \( \phi \in [0,1] \), let \( B_\phi \) denote the symmetric distribution defined on \( \{-\ln \frac{1-\phi}{\phi}, \ln \frac{1-\phi}{\phi}\} \). We define box convolution \( \boxplus \) to be the bilinear operator over the space of symmetric distributions satisfying the following condition

\[
B_{\phi_1} \boxplus B_{\phi_2} \preceq B_{\phi_1 + \phi_2 - 2\phi_1 \phi_2} \quad \text{for all } \phi_1, \phi_2 \in [0,1].
\]

It is clear that box convolution is commutative. One can show the following alternative definition, which proves box convolution is associative.

Proposition 7. Let \( X \sim \mu, Y \sim \nu \) be independent random variables with symmetric distributions, then \( \mu \boxplus \nu \) is identical to the law of

\[
Z \triangleq 2 \tanh^{-1} \left( \frac{\tanh \frac{X}{2} \tanh \frac{Y}{2}}{2} \right).
\]

Remark 2. As a physical interpretation, \( B_\phi \boxplus \mu \) can be viewed as the LLR distribution for an experiment where the observation channel is the concatenation of a binary symmetric channel (BSC) with crossover probability \( \phi \) and a symmetric channel with conditional distribution given by \( \mu \). The box convolution between general distributions can be interpreted as the LLR distribution for the same estimation problem, except that the crossover probability for the BSC is known and random.

The gist of our proof is to express the BP operator using elementary operations, so one can investigate their commutation relation with box convolution by characterizing the commutation rules between elementary operators.

Proposition 8. For any symmetric \( \mu \), we have \( Q\mu = \mathbb{E}_{P_{\delta}}[(B_\delta \boxplus \mu)^{*d}] \), where \( (\cdot)^{*d} \) denotes self convolution by \( d \) times.

In particular, we use the commutativity of box convolution, which is equivalently \( F_\delta \circ F_\delta = F_\delta \circ F_\delta \). Then we develop an exchange rule between convolution and box convolution. A key intermediate result is summarized in Theorem 10, which relies on the following definition.\(^2\)

Definition 9 (Strict Degradation). For any two distributions \( \nu \) and \( \mu \), we define \( \nu \prec \mu \) if \( \exists \phi \in (0, \frac{1}{2}) \) such that \( \nu \preceq B_\phi \boxplus \mu \).

Theorem 10. For any \( \delta \in [0,1] \) and degree distribution \( P_d \), let \( Q \) be the associated BP operator. Then for any \( \phi \in (0, \frac{1}{2}) \) and symmetric \( \mu \),

\[
B_\delta \boxplus Q^2 \mu \prec Q^2(B_\delta \boxplus \mu) \quad \text{if } P[d > 2] = 0, \tag{6}
\]

\[
B_\delta \boxplus Q\mu \prec Q(B_\delta \boxplus \mu) \quad \text{otherwise}. \tag{7}
\]

\(^2\)The proof of Theorem 10 can be found in Section II.
Remark 3. Note that the fixed point equation can be written as \( Q \mu = \mu \). Theorem 10 implies that \( \mu \) and \( B_\phi \boxplus \mu \) can not be both non-trivial symmetric fixed points. As we will show later in this work, Theorem 10 essentially states that either \( Q^2 \) or \( Q \) will always reduce a distance between distinct non-trivial symmetric distributions that is measured based on degradation index.

Remark 4. The stringy tree construction in [9] implies that \( B_\phi \boxplus Q \mu \leq Q(B_\phi \boxplus \mu) \) for any degree distribution. Theorem 10 provides a strict version of this result for \( P[d > 2] > 0 \). Note that this condition can not be further relaxed, and one can show that inequality (7) is not satisfied when \( P[d > 2] = 0 \).

We also have the following fact as degradation is transitive and box-convolution-preserving (see Appendix A).

Proposition 11. If \( \nu \preceq \tau \prec \mu \) or \( \nu \prec \tau \preceq \mu \), then \( \nu \prec \mu \).

Given these results, we are ready to prove the uniqueness of BP fixed points.

Proof of Theorem 3. Consider any two non-trivial symmetric fixed points \( \mu, \nu \), we first prove that \( \mu \preceq \nu \).

Definition 12 (Degradation Index). For any two symmetric distributions \( \mu \) and \( \nu \), we defined the degradation index from \( \nu \) to \( \mu \) to be

\[
\phi^*(\mu, \nu) \triangleq \inf \{ \phi \mid B_\phi \boxplus \mu \preceq \nu \}. \tag{8}
\]

Proposition 13. Degradation index has the following properties.

1. We always have \( \phi^*(\mu, \nu) < \frac{1}{2} \) for \( \nu \) non-trivial.
2. For any symmetric \( \mu \) and \( \nu \), we have \( B_{\phi^*(\mu, \nu)} \boxplus \mu \preceq \nu \).
3. For any \( \phi \in (0, \frac{1}{2}) \) and any symmetric \( \mu \) and \( \nu \) satisfying \( B_\phi \boxplus \mu \preceq \nu \), we have \( \phi^*(\mu, \nu) < \phi \).
4. For any symmetric \( \mu, \nu, \tau \), we have \( 1 - 2\phi^*(\mu, \tau) \geq (1 - 2\phi^*(\mu, \nu))(1 - 2\phi^*(\nu, \tau)) \).

The first property is proved in Appendix D by a probability-of-error argument. The second property indicates that the set defined in equation (8) has a minimum. It follows from the fact that degradation commutes with weak convergence, see Proposition 45. The third property shows that strict degradation provides strict upper bound on degradation index. It can be proved using Proposition 46. The fourth property states a triangle inequality for \( |\ln(1 - 2\phi^*(\mu, \nu))| \), which is due to the transitivity of degradation and associativity of box convolution.

We consider the fixed point condition and any \( \phi \in (0, \frac{1}{2}) \) satisfying \( B_\phi \boxplus \mu \preceq \nu \). From the first property in Proposition 13, such \( \phi \) exists. Then from the second property, we can choose \( \phi = \phi^*(\mu, \nu) \) unless \( \phi^*(\mu, \nu) = 0 \). Theorem 10 states that there is an integer \( k \) for any degree distribution such that

\[
B_\phi \boxplus \mu \preceq Q^k(B_\phi \boxplus \mu). \tag{9}
\]

Because \( Q \) describes BP, it preserves degradation (see Proposition 48). Hence

\[
Q^k(B_\phi \boxplus \mu) \preceq Q^k \nu = \nu.
\]

Recall the transitivity property stated in Proposition 11. This implies \( B_\phi \boxplus \mu \prec \nu \).

However, this conclusion is mutually exclusive with \( \phi = \phi^*(\mu, \nu) \) according to the third property in Proposition 13. Hence, we must have \( \phi^*(\mu, \nu) = 0 \), then \( \mu \preceq \nu \) follows from the second property in Proposition 13.

By symmetry, we have \( \nu \preceq \mu \) as well. Then by the antisymmetry of degradation (see Appendix A), \( \mu \) and \( \nu \) are identical. Therefore, there can be at most one unique non-trivial symmetric BP fixed points.

\[ \square \]

Remark 5. As shown in our proof steps, the identity condition of two non-trivial symmetric distributions is exactly given by \( \phi^*(\mu, \nu) = \phi^*(\nu, \mu) = 0 \). Recall the triangle inequality implied by Proposition 13. Degradation index defines a metric on the space of non-trivial symmetric distributions given by the following equation:

\[
d(\mu, \nu) = |\ln(1 - 2\phi^*(\mu, \nu)) + \ln(1 - 2\phi^*(\nu, \mu))|. \tag{9}
\]

The proof of uniqueness can exactly be interpreted as showing \( Q^k \) is a contractive map [18] with respect to \( d(\mu, \nu) \). Generally, convergence in this degradation metric implies weak convergence and the \( L_\infty \) distance convergence in a concept called \( \beta \)-curves to be defined later in this paper (See Appendix C and Proposition 54).

Compared to the literature, contraction methods have seen the most success for linear models, where metrics are constructed based on certain expectation classes of the distributions while being sublinear under mixtures [19]. Under such models, contraction properties can be proved by analyzing the operator norms [20]. The same approach does not directly apply to non-linear models, in particular, for \( F_\beta \) being non-linear. It was shown that metrics defined based on certain expectation functions are not BP-contractive under all regimes [13]. Degradation index is constructed with a different approach that violates sublinearity, but provides the needed contractive properties.

Remark 6. Recall that our results for the basic setting focuses on non-trivial cases where \( d \geq 2 \) with non-zero probability. Theorem 3 is not always true within the trivial cases. Particularly, in the deterministic case where \( d = 1 \) and \( \delta \in \{0, 1\} \),

\[ \Box \]
the uniqueness property does not hold and any distribution is a BP fixed point. This example is the only exception for the uniqueness of non-trivial BP fixed points.

### B. Extension to Broadcast with Survey

Now we consider the inference problem with survey observations [13]. In this setting, each vertex above level \( h \) is observed through identical symmetric channels, called survey channels. Note that any symmetric channel can be characterized by its LLR distribution, which is defined similar to equation (1). We use \( \mu_s \) to denote the LLR distribution for the survey channels. Clearly, \( \mu_s \) is symmetric.

The goal is to characterize the estimation process of \( X_0 \) based on all survey information and all observations in the basic formulation. The LLR distribution for the estimation problem converges as \( h \to \infty \) to a distribution satisfying the following condition.

**Definition 14** (Fixed point of BP for Broadcast with Survey). For any given degree distribution \( P_d \), parameter \( \delta \), and survey distribution \( \mu_s \), we say a distribution \( \mu \) defined on domain \((-\infty, +\infty]\) is a BP fixed point if

\[
\mu = (Q\mu) \ast \mu_s,
\]

where \( Q \) is BP operator in the basic formulation.

Similar to the basic setting, we let \( Q_s \) denote the BP operator when survey channels are present.

**Proposition 15.** For any distribution \( \mu \), let \( Q_s \mu \equiv (Q\mu) \ast \mu_s \). Then \( \mu \) is a BP fixed point if and only if \( Q_s \mu = \mu \).

For ease of discussion, we focus on survey channels with non-zero capacity, i.e., non-trivial \( \mu_s \), as the results for trivial \( \mu_s \) are already covered in the basic setting. We prove the uniqueness of BP fixed points, stated in the following Theorem.

**Theorem 16** (Uniqueness of BP fixed points for Broadcast with Survey). In the broadcast with survey setting, there is exactly one unique symmetric BP fixed point for any fixed degree distribution \( P_d \), parameter \( \delta \in [0,1] \), and non-trivial symmetric survey distribution \( \mu_s \).

The proof of uniqueness relies on the following intermediate step, which is proved in Section II-B.

**Theorem 17.** For any \( \delta \in [0,1] \), non-trivial symmetric \( \mu_s \), and degree distribution \( P_d \) with \( \mathbb{P}[d = 0] < 1 \), let \( Q_s \) be the associated BP operator. Then for any \( \phi \in (0, \frac{1}{2}) \) and symmetric \( \mu \), we have

\[
B_\phi \boxplus Q_s^k \mu \prec Q_s^k (B_\phi \boxplus \mu)
\]

if \( \mathbb{P}[d > 1] = 0 \),

\[
B_\phi \boxplus Q_s \mu \prec Q_s (B_\phi \boxplus \mu)
\]

otherwise,

for some \( k \in \mathbb{N}_+ \).

Assuming the correctness of Theorem 17, we present the proof for the uniqueness of BP fixed points.

**Proof of Theorem 16.** Note that \( \mu_s \) is non-trivial. Any BP fixed point with respect to \( Q_s \) must also be non-trivial. Then following the same steps in the proof of Theorem 3, one can show that Theorem 17 implies the uniqueness of non-trivial symmetric BP fixed points. To summarize, there can be at most one symmetric BP fixed point for any fixed \( P_d, \delta \), and symmetric \( \mu_s \).

On the other hand, the LLR distributions for the broadcast with survey problem always converges to a symmetric BP fixed point. This proves the existence of a symmetric BP fixed point. Thus, there is exactly one unique symmetric BP fixed point for any fixed \( P_d, \delta \), and symmetric \( \mu_s \).

**Remark 7.** The existence of survey channels significantly affects the properties of BP fixed points. As we have shown, when the survey channels have a non-zero capacity, there is always one unique BP fixed point, and it is non-trivial. However, when survey channels are absent, either the trivial and non-trivial fixed points coexist, or only the trivial solution remains. This difference is also reflected in the statements of the uniqueness theorems.

### II. PROOF OF MAIN RESULTS

In this section, we prove the key intermediate steps, i.e., Theorem 10 and Theorem 17. For convenience, we make the following definition.

**Definition 18.** For any symmetric distribution \( \mu \), we define its \( \beta \)-curve as a function on domain \( t \in \mathbb{R} \) given by the following equation.

\[
\beta(t; \mu) \triangleq \mathbb{E}_{R \sim \mu} \left[ \max \left\{ \tanh \frac{|R|}{2}, |t| \right\} \right].
\]
We also define

\[ t_{\max}(\mu) \triangleq \inf\{t \in [0, 1] \mid \beta(t) = t\}. \]

**Proposition 19 (Linearity of \( \beta \)-curves).** If \( \mu = E_\phi[\mu_0], \) then \( \beta(t; \mu) = E_\phi[\beta(t; \mu_0)]\).

The \( \beta \)-curve provides exact conditions for strict degradation, stated in the following Proposition and proved in Appendix E.

**Proposition 20.** For any non-trivial symmetric distributions \( \mu \) and \( \nu \), the following statements are equivalent.

1) \( \nu \prec \mu \).
2) \( \beta(t; \nu) < \beta(t; \mu) \) for all \( t \in [0, t_{\max}(\nu)] \).
3) \( \beta(t; \nu) < \beta(t; \mu) \) for all \( t \in [0, t_{\max}(\mu)] \) and \( t_{\max}(\nu) < t_{\max}(\mu) \).

**Remark 8.** By symmetry condition, one can show that \( \frac{1}{2}(1 - \beta(t; \mu)) \) for \( t \in [0, 1] \) equals the minimum error probability for a hypothesis testing problem with conditional distribution given by \( \mu \) and prior given by \( E_{\frac{1}{2}}[\cdot] \). Therefore, Proposition 20 proves that strict degradation between two distributions can be completely determined by comparing their associated inference problems in the Bayesian sense.

The proof of Proposition 20 relies on several elementary steps. In the following Theorem, we state a convenient tool for proving degradation relation, which can be derived from the celebrated Blackwell–Sherman–Stein (BSS) theorem. An equivalent form is also stated in [21], Theorem 4.7.6. For discrete distributions with finite supports, Theorem 21 can be proved by constructing joint distributions with induction. The statement can then be generalized to general distributions using a convergence argument.

**Theorem 21.** [Blackwell–Sherman–Stein] For any pair of symmetric distributions \( \nu \) and \( \mu \), we have \( \nu \leq \mu \) if and only if \( \beta(t; \nu) \leq \beta(t; \mu) \) for all \( t \in [0, 1] \).

Proposition 20 provides a strict version by connecting strict degradation with gap conditions between \( \beta \)-curves. This is achieved using the following fact, which states that box convolution with any \( B_\phi \) can be viewed as a homothetic transformation. Then equation (15) can be proved from the second definition of box convolution.

**Proposition 22.** For any \( \phi \in [0, \frac{1}{2}] \), the \( \beta \)-curve of \( B_\phi \) is given by

\[ \beta(t; B_\phi) = \max\{|t|, 1 - 2\phi\}. \] (14)

More generally, for any symmetric \( \mu \),

\[ \beta(t; B_\phi \boxplus \mu) = (1 - 2\phi)\beta\left(\frac{t}{1 - 2\phi}; \mu\right). \] (15)

In order to prove the needed \( \beta \)-curve gaps for Theorem 10 and Theorem 17, we define the following notations for some useful statements.

**Definition 23.** For any symmetric distributions \( \mu, \nu, \) and parameter \( s \geq 0 \), we define \( \nu \prec_s \mu \) if \( \beta(t; \nu) < \beta(t; \mu) \) for all \( t \) satisfying

\[ t \in \left( \tanh\left(\frac{s}{2}\right), t_{\max}(\nu) \right] \]

and \( \beta(t; \nu) \leq \beta(t; \mu) \) for all other \( t \). We also define

\[ r_{\max}(\mu) \triangleq \sup\{r \in [0, +\infty] \mid \mu([r, +\infty]) > 0\} = 2\tanh^{-1} t_{\max}(\mu). \]

**Proposition 24.** Let \( \mu, \nu, \tau \) be symmetric distributions.

1) We have \( \nu \prec_s \mu \) if \( \beta(t; \nu) < \beta(t; \mu) \) for all \( t \in (\tanh\left(\frac{s}{2}\right), t_{\max}(\mu)) \), \( \beta(t; \nu) \leq \beta(t; \mu) \) for all other \( t \), and \( t_{\max}(\nu) < t_{\max}(\mu) \); when \( s < r_{\max}(\mu) \), the converse also holds true.
2) If \( \nu \leq \tau \prec_s \mu \) or \( \nu \prec_s \tau \leq \mu \), then \( \nu \prec_s \mu \).
3) \( \nu \prec \mu \) implies \( \nu \prec_s \mu \) for any \( s \), the latter implies \( \nu \leq \mu \).

**Proposition 25.** For any \( \phi \in (0, 1) \) and \( \delta_1, \delta_2 \in [0, 1] \), let \( s_{\min} = |F_{\phi}(r_{\max}(B_{\delta_1})) - F_{\phi}(r_{\max}(B_{\delta_2}))| \). Then we have

\[ B_\phi \boxplus (B_{\delta_1} \boxplus B_{\delta_2}) \prec_s (B_\phi \boxplus B_{\delta_1}) \boxplus (B_\phi \boxplus B_{\delta_2}). \]

The proof of Proposition 25 is in Appendix F. We show that the \( \beta \)-curves for the distributions on both sides are piecewise linear. Moreover, we provide the exact condition for the inequality between \( \beta \)-curves to be strict. By integrating the \( \beta \)-curves for the above special case and applying Proposition 24, one can prove the following corollary.

**Corollary 26.** For any \( \phi \in (0, 1) \) and any two symmetric distributions \( \mu_1, \mu_2 \), we have

\[ B_\phi \boxplus (\mu_1 \boxplus \mu_2) \prec_s (B_\phi \boxplus \mu_1) \boxplus (B_\phi \boxplus \mu_2) \] (16)
Proof. Inequality (18) directly follows from Corollary 26. Inequality (19) clearly holds when $\prec$ operations, providing alternative proofs for Proposition 47. Similar properties can be extended to For any symmetric $\mu$, conditions. Then we apply law of total probability to obtain the stated equation. By integrating the $\prec$-curves in Proposition 27, we have the following result, which is proved in Appendix H.

**Definition 28.** For any symmetric $\mu$, define $\text{supp}(\mu) \triangleq \{ v \in \mathbb{R} \mid \mu([v - \epsilon, v + \epsilon]) > 0 \forall \epsilon > 0 \}$.

**Proposition 29** (Rule of Convolution for Symmetric Distributions). For any symmetric $\mu$, $\nu$, $\tau$, and $\ell \in \text{supp}(\tau)$, if $\nu \prec_{s} \mu$, then

$$\beta(t; B_{\phi} \ast \mu) = \left( \frac{1 + t - 2t\phi}{1 + t - 2t\phi} \right) \beta(t_0; \mu) + \left( \frac{1 - t + 2t\phi}{1 - t + 2t\phi} \right) \beta(t_1; \mu),$$

where

$$t_0 \triangleq \frac{1 + t - 2\phi}{1 + t - 2t\phi} = \tanh \left( \frac{r_{\text{max}}(B_{\phi})}{2} + \tanh^{-1}(t) \right),$$

$$t_1 \triangleq \frac{-1 - t + 2\phi}{1 - t + 2t\phi} = \tanh \left( \frac{r_{\text{max}}(B_{\phi})}{2} - \tanh^{-1}(t) \right).$$

Proposition 27 is proved in Appendix G. In particular, we view $\prec$-curves as linear functions of error probabilities for certain estimation problems, and $B_{\phi}$-convolution corresponds to side information measured though a BSC. Conditioned on the side information, the experiment reduces to inference problems defined by the unconvolved distribution with different prior conditions. Then we apply law of total probability to obtain the stated equation. By integrating the $\prec$-curves in Proposition 27, we have the following result, which is proved in Appendix H.

**Proposition 30.** For any symmetric $\mu$, $\nu$, and $\tau$, we have the following facts.

1) If $\nu \prec \mu$, then $\tau \boxplus \nu \prec \tau \boxplus \mu$.
2) If $\nu \prec_{s} \mu$, then $\tau \boxplus \nu \prec_{s} \tau \boxplus \mu$, where $s_{\tau} \triangleq \ln \frac{e^{\frac{\tau}{\nu}\text{supp}(\nu)} + 1}{e^{\frac{\tau}{\mu}\text{supp}(\mu)} + 1}$.
3) If $\nu \prec \mu$ and $r_{\text{max}}(\tau) < r_{\text{max}}(\mu)$, then $\tau \ast \nu \prec \tau \ast \mu$.

The above results can be used to prove the following statements.

**Proposition 31.** For any $\phi \in (0, 1)$ and any symmetric $\nu$, we have

$$B_{\phi} \boxplus (\nu^{(2)}) \prec_{s} (B_{\phi} \boxplus \nu)^{(2)},$$

$$B_{\phi} \boxplus (\nu^{(d)}) \prec (B_{\phi} \boxplus \nu)^{(d)} \quad \forall \ d > 2.$$  \hspace{1cm} (18) \hspace{1cm} (19)

**Proof.** Inequality (18) directly follows from Corollary 26. Inequality (19) clearly holds when $\phi = \frac{1}{2}$ or $\nu$ is trivial. We prove inequality (19) by induction for both $B_{\phi}$ and $\nu$ are non-trivial.

(a) Consider the base case where $d = 3$. Let $r_{\nu} = r_{\text{max}}(\nu)$. Because convolution preserves degradation, by stringy tree lemma, or inequality (18) and Proposition 24, we have

$$\left( B_{\phi} \boxplus \nu \right) \ast (B_{\phi} \boxplus (\nu^{(2)})) \leq (B_{\phi} \boxplus \nu)^{(3)}.$$  \hspace{1cm} (20)

Next, from Corollary 26, we have

$$B_{\phi} \boxplus (\nu^{(3)}) \prec_{s} (B_{\phi} \boxplus \nu) \ast (B_{\phi} \boxplus (\nu^{(2)})),$$

where

$$s = F_{\phi}(r_{\text{max}}(\nu^{(2)})) - F_{\phi}(r_{\text{max}}(\nu)) = F_{\phi}(2r_{\nu}) - F_{\phi}(r_{\nu}).$$

The above two steps form a chain of degradation. By transitivity stated in Proposition 24, this implies

$$B_{\phi} \boxplus (\nu^{(3)}) \prec_{s} (B_{\phi} \boxplus \nu)^{(3)}.$$
To prove the needed statement, it suffices to show that any of these two steps has strict inequality in $\beta$-curves for $t \in [0, s]$. We apply Rule of Convolution to inequality (20) and let $\ell = F_\phi(r_\nu) = r_{\text{max}}(B_\phi \Box \nu) \in \text{supp}(B_\phi \Box \nu)$, the strict condition holds for

$$t \in \left[0, \tanh \left( \frac{r_{\text{max}}(B_\phi \Box (\nu^{(2)})) - \ell}{2} \right) \right] = \left[0, \tanh \left( \frac{F_\phi(r_\nu)}{2} \right) \right].$$

Because $F_\phi(r_\nu) > s$ for both $B_\phi$ and $\nu$ non-trivial, the strict degradation statement is implied by Proposition 20.

(b) Assume inequality (19) is proved for some $d = d_0 \geq 3$. By induction assumption, we have the following chain similar to the base case.

$$B_\phi \Box (\nu^{(d_0+1)}) \not{\preceq} (B_\phi \Box \nu) \ast (B_\phi \Box (\nu^{(d_0)}) \prec (B_\phi \Box \nu)^{*(d_0+1)}.$$  

In particular, we apply the third property in Proposition 30 to show strict degradation in the second step. Therefore, the induction step follows from Proposition 11.

(c) To conclude, inequality (19) is proved for any $d \geq 3$.  

Let $\nu = B_* \Box \mu$, Proposition 31 states degradation relationships between $B_\phi \Box (Q\mu)$ and $Q(B_\phi \Box \mu)$ when $d$ is deterministic.

**Corollary 32.** If $\mathbb{P}[d = d_0] = 1$ for some fixed $d_0$, then

$$B_\phi \Box (Q\mu) \prec_0 Q(B_\phi \Box \mu) \quad \text{if } d_0 = 2, \quad (22)$$

$$B_\phi \Box (Q\mu) \prec Q(B_\phi \Box \mu) \quad \text{if } d_0 > 2. \quad (23)$$

**A. Proof of Theorem 10**

**Proof.** For brevity, we focus on non-trivial cases where $\mu$ is non-trivial and $\delta \neq \frac{1}{2}$. We first consider the deterministic $d$ case and fill in the gap for $d = 2$. Note that $Q$ preserves degradation (see Proposition 48). By Corollary 32 and Proposition 24, we have the following chain of degradation.

$$B_\phi \Box Q^2 \mu \prec_0 Q(B_\phi \Box Q^2 \mu) \prec Q^2(B_\phi \Box \mu).$$

In particular, the first step is obtained by replacing $\mu$ with $Q\mu$ in Corollary 32. The above chain implies strict inequality in $\beta$-curves for $t \in (0, t_{\text{max}}(B_\phi \Box Q^2 \mu)]$. By non-trivial condition, from Proposition 20, it remains to prove strict inequality of $\beta$-curves at $t = 0$.

To that end, we zoom in on the second step and apply Rule of Convolution to the first inequality of the following chain.

$$Q(B_\phi \Box Q\mu) = (B_\phi \Box B_\delta \Box Q\mu)^{(2)} \quad (24)$$

$$\preceq (B_\delta \Box B_* \Box Q\mu) \ast (B_\delta \Box Q(B_\phi \Box \mu)) \quad (25)$$

$$\prec (B_\delta \Box Q(B_\phi \Box \mu))^{*(2)} = Q^2(B_\phi \Box \mu). \quad (26)$$

Note that Corollary 32 and Proposition 30 implies

$$B_\phi \Box B_\delta \Box Q\mu \prec_0 B_\delta \Box Q(B_\phi \Box \mu).$$

By Proposition 48, the non-trivial condition implies that the $r_{\text{max}}$ functions for both sides are different. Therefore, we can choose $\ell = r_{\text{max}}(B_\phi \Box B_\delta \Box Q\mu)$ for the Rule of Convolution and apply it to inequality (25), which leads to the needed strict condition at $t = 0$.

Now we consider general degree distributions. First for $\mathbb{P}[d \geq 2] = 0$, recall that our formulation assumes non-trivial cases where $\mathbb{P}[d \leq 1] < 1$. We have $d = 2$ with non-zero probability. Then the $r_{\text{max}}$ function of $B_\phi \Box (Q^2 \mu)$ is identical to that of its $d = 2$ component. Thus, by linearity, our earlier proof for the deterministic $d = 2$ case implies strict inequality of $\beta$-curves for the full range $t \in [0, t_{\text{max}}(B_\phi \Box (Q^2 \mu))]$, and the needed statement is implied.

On the other hand, we have $\mathbb{P}[d > 2] > 0$. If $d$ is upper bounded by some fixed integer almost surely, we can let $d_0$ be the largest possible $d$ for such degree distribution and apply the same linearity argument to inequality (23) to prove the statement. Otherwise, $d$ is unbounded, and we have strict inequality on $\beta$-curves for any $t < t_{\text{max}}(Q(B_\phi \Box \mu)) = 1$. Note that $t_{\text{max}}(B_\phi \Box (Q^2 \mu)) = 1 - 2\phi < 1$. The statement follows from Proposition 24.  

**B. Proof of Theorem 17**

We start by formulating two useful results.

**Proposition 33.** For any $\phi \in (0, 1)$ and any symmetric distributions $\mu, \nu$, let $s_{\text{max}} \triangleq F_\phi(r_{\text{max}}(\mu)) - r_{\text{max}}(\nu)$. We have

$$B_\phi \Box (\mu \ast \nu) \prec (B_\phi \Box \mu) \ast \nu \quad (27)$$

if $s_{\min} < 0$, and
\[ B_\phi \boxplus (\mu \ast \nu) \prec_{s_{\min}} (B_\phi \boxplus \mu) \ast \nu \]
otherwise.

Proof. Similar to the basic setting, our technique is to prove strict inequalities for beta-curves by forming a chain of degradation. First observe that $B_\phi \boxplus \nu \prec \nu$ for any $\phi \in (0, 1)$. By convolving $B_\phi \boxplus \mu$ on both sides and apply Proposition 47, we have
\[ (B_\phi \boxplus \mu) \ast (B_\phi \boxplus \nu) \preceq (B_\phi \boxplus \mu) \ast \nu. \tag{28} \]
On the other hand, Corollary 26 implies following step, which completes the chain.
\[ B_\phi \boxplus (\mu \ast \nu) \preceq (B_\phi \boxplus \mu) \ast (B_\phi \boxplus \nu) \tag{29} \]
When $s_{\min} < 0$, we have $r_{\max}(B_\phi \boxplus \mu) = F_\phi(r_{\max}(\mu)) < r_{\max}(\nu)$. Hence, we can apply the third statement in Proposition 30 to obtain a strict version of inequality (28), i.e., $(B_\phi \boxplus \mu) \ast (B_\phi \boxplus \nu) \preceq (B_\phi \boxplus \mu) \ast \nu$. Then, inequality (27) follows from the transitivity statement in Proposition 11.

For $s_{\min} \geq 0$, we need to verify that the beta-curves have an overall non-zero gap for all $t = \tanh \frac{s}{2}$ with
\[ s \in (s_{\min}, r_{\max}(B_\phi \boxplus (\mu \ast \nu))]. \]
Note that the guarantee provided by Corollary 26 for inequality (29) already covers the subset where $s > |r_{\max}(B_\phi \boxplus \mu) - r_{\max}(B_\phi \boxplus \nu)| = |F_\phi(r_{\max}(\mu)) - F_\phi(r_{\max}(\nu))|$. By the assumption of $s_{\min} \geq 0$, we have $F_\phi(r_{\max}(\mu)) \geq r_{\max}(\nu) \geq F_\phi(r_{\max}(\nu))$. Hence, inequality (29) contributes non-zero gaps all $t = \tanh \frac{s}{2}$ with
\[ s \in (F_\phi(r_{\max}(\mu)), r_{\max}(B_\phi \boxplus (\mu \ast \nu))]. \]
On the other hand, the gap condition for inequality (28) can be analyzed using the rule of convolution with $\tau = B_\phi \boxplus \mu$ and $\ell = r_{\max}(B_\phi \boxplus \mu) = F_\phi(r_{\max}(\mu))$. This implies strict inequality for $t = \tanh \frac{s}{2}$ with
\[ s_{\min} = F_\phi(r_{\max}(\mu)) - r_{\max}(\nu) < s \leq F_\phi(r_{\max}(\mu)), \]
which covers the entire rest of the interval. \hfill $\square$

Next, consider the case of $d = 1$, so that $Q_\alpha \ast \mu = (B_\phi \boxplus \mu) \ast \mu$. By induction, one can derive the following result (proof in Appendix I).

Proposition 34. For $\phi \in (0, 1)$ and $d = 1$, let $r_s \triangleq r_{\max}(\mu), s_0 \triangleq F_\phi(r_{\max}(\mu)), \text{then for any } k \in \mathbb{N}, \text{we have}$
\[ B_\phi \boxplus (Q_k^s \mu) \prec Q_k^s(B_\phi \boxplus \mu) \]
if $s_k \triangleq F_\phi(s_{k-1}) - r_s < 0$, and
\[ B_\phi \boxplus (Q_k^s \mu) \prec_{s_k} Q_k^s(B_\phi \boxplus \mu) \]
otherwise.

One can show that there is a finite $k$ for $s_k < 0$. For example, as a rough estimate, we have $s_0 \leq F_\phi(+\infty) < +\infty$ and $s_k \leq s_{k-1} - r_s$ for any positive $s_k$. Because $\mu$ is non-trivial, we also have $r_s > 0$. Hence, we can find $k \leq 1 + s_0/r_s$ for strict degradation to hold, which gives the needed statement.

Proof of Theorem 17. First, consider the case $P[d > 1] = 0$. Recall that theorem statement assumes $P[d = 0] < 1$, so that we have $d = 1$ with non-zero probability. Consequently, the $\beta$-curve analysis is dominated by the $d = 1$ component, and our earlier proof steps shows strict inequalities in the full range of $t$.

Formally, $Q_\alpha$ can be written as a linear combination of two operators, each corresponds to the BP operator for a deterministic $d \in \{0, 1\}$. Thus, $Q_\alpha^k$ can be expanded into a linear combination of $2^k$ chains, and each side of inequality (11) can be decomposed into $2^k$ terms. Each corresponding pair can be compared individually. Recall that $Q_\alpha \ast \mu = (Q_\mu) \ast \mu$. By Corollary 32 and Proposition 33, we have that $B_\phi \boxplus (Q_\alpha \ast \mu) \preceq Q_\alpha(B_\phi \boxplus \mu)$. By applying this inequality recursively, each corresponding terms in the decomposition satisfies degradation.

Among all terms, the ones with all BP operators corresponds to the $d = 1$ case achieves strict degradation due to Proposition 34. Note that these terms have the largest $t_{\max}$ values on both sides. The interval $t \in [0, t_{\max}(B_\phi \boxplus (Q_\alpha \ast \mu))]$ must be contained within the range where the $\beta$-curve inequality between these two terms is strict. The $P[d = 0] < 1$ condition ensures that this non-zero gap has non-zero weights in the overall $\beta$ functions. Then inequality (11) is proved by definition.

It remains to consider general degree distributions with $P[d > 1] > 0$. Note that in the basic setting, we have essentially proved that if $P[d \leq 1] < 1$, then
\[ B_\phi \boxplus Q_\mu \prec_0 Q(B_\phi \boxplus \mu). \tag{30} \]
Specifically, the above inequality is directly implied by Theorem 10 if \( \mathbb{P}[d > 2] > 0 \). In the other case, we have that the \( r_{max} \) function of \( B_\phi \boxplus Q \mu \) is dominated by its \( d = 2 \) component. Thus, inequality (22) implies non-zero gaps in \( \beta \)-curves for all \( t \in (0, t_{max}(B_\phi \boxplus Q \mu)) \), which proves inequality (30).

Let \( r_s = r_{max}(\mu_s), r_Q = r_{max}(Q \mu), r_\tilde{Q} = r_{max}(Q(B_\phi \boxplus \mu)) \). We first apply Proposition 33 and then the Rule of Convolution to obtain

\[
B_\phi \boxplus Q_s \mu \preceq (B_\phi \boxplus Q \mu) * \mu_s \preceq Q_s(B_\phi \boxplus \mu).
\]

Consider the first step of inequality (31), the statement of Proposition 33 implies that the gap between the \( \beta \)-curves on both sides is strict for any \( t = \tanh \frac{|s|}{2} \) with

\[
F_\phi(r_Q) - r_s < s < F_\phi(r_Q) + r_s.
\]

Then by the rule of convolution, the \( \beta \)-curve inequality for the second step is strict for \( t = \tanh \frac{|s|}{2} \) with

\[
-r_s < s < r_\tilde{Q} - r_s.
\]

Note that inequality (30) implies that \( r_\tilde{Q} = r_{max}(B_\phi \boxplus Q \mu) = F_\phi(r_Q) \). Thus, (32) and (33) cover all \( 0 \leq s < F_\phi(r_Q) + r_s \), concluding the proof of

\[
B_\phi \boxplus Q_s \mu \preceq Q_s(B_\phi \boxplus \mu).
\]

\[\square\]

### III. Implications

#### A. Extension to Robust Reconstruction

Consider a variant to the basic setting, called robust reconstruction [11], where all leaf observations are obtained through some identical noisy channels. The estimation problem is to infer the root variable given the tree structure and the noisy observations. In particular, let \( \mu_s \) denote the LLR distribution of the observation channels. Then the LLR distribution for estimating the root variable for trees of level \( h \) is exactly given by \( Q^h \mu_s \), where \( Q \) is the BP operator defined in Section I.

Using concepts developed for the uniqueness theorems, we prove that the LLR distributions converge as \( h \to \infty \) for any initialization.\(^3\) Moreover, we prove the following theorem.

**Theorem 35.** For any fixed \( P_d \) and \( \delta \), the distributions in the following classes all exist and are identical.

(a) The limiting LLR distribution for the basic setting.

(b) The limiting LLR distribution for robust construction with any non-trivial initialization.

(c) The dominant BP fixed point.

**Remark 9.** In [12], it was conjectured that the error probability for the maximum likelihood estimator is independent of the observation channels when \( h \to \infty \), as long as the channel capacity is non-zero. Note that this error probability can be written as an expectation over the LLR distribution (see Remark 8). The unique convergence stated in Theorem 35 provides a positive proof to this conjecture. More generally, the same guarantee holds for any quantity that can be written as the expectation of a bounded continuous function on \( (-\infty, +\infty) \), such as mutual information and Bayesian estimation errors under different prior distributions. Theorem 35 also provides an independent proof of Proposition 1 in [22] for robust reconstruction.

**Remark 10.** The existence of a dominant BP fixed point is guaranteed by Theorem 3 (and by Corollary 4 to include asymmetric distributions). Whenever non-trivial BP fixed point exists, the dominant BP fixed point is identical to the unique non-trivial symmetric BP fixed point. Otherwise, it is the unique trivial fixed point. Therefore, Theorem 35 shows that when non-trivial and trivial fixed points coexist, BP recursion converges to the non-trivial fixed point for any non-trivial initialization.

**Proof.** To prove Theorem 35, we essentially aim to show that BP recursion stabilizes the LLR distributions under different initializations. We first consider the basic setting as a reference, which can be viewed as a special case of robust reconstruction with noiseless observation, i.e., by having \( \mu_s = B_0 \). For clarity, we denote the LLR distributions in robust reconstruction by \( \hat{\mu}_h \), and the LLR distributions in the basic setting by \( \mu_h \). The convergence of \( \mu_h \) for \( h \to \infty \) is well known\(^4\), and we denote their limiting distribution by \( \mu^* \).

The LLR distributions for robust reconstruction are always degraded versions compared to those for the basic setting. Therefore, with any initialization, we always have \( B_\phi \preceq \hat{\mu}_h \preceq \mu_h \). This implies the needed convergence when \( \mu^* \) is trivial, by applying Proposition 40 with \( h \) and \( \mu = \mu_h \).

When \( \mu^* \) is non-trivial, we have to develop new lower bound constructions, and prove they converges to \( \mu^* \). In particular, there are no guaranteed degradation relation between \( \hat{\mu}_h \) and \( \hat{\mu}_{(h+1)} \), so the weak convergence of LLR distributions does not follow directly from the filtration argument. However, if we let \( \mu_{(0)} = B_\phi^{\mu^*} \), then \( \phi^*(\mu^*, \hat{\mu}_0) \) defined in

\(^3\) For ease of discussion, in this section we will also ignore trivial cases where \( \mathbb{P}[d \leq 1] = 1 \).

\(^4\) The proof follows from a filtration argument, which relies on the fact that the LLR distributions \( \{\mu_h\}_{h \in \mathbb{N}} \) form a degrading sequence.
equation (8), and let \( \mu_h \triangleq Q^h \mu_0 \) for any \( h \geq 1 \), then \( \mu_0 \) can be viewed as a non-trivial initialization that satisfies the filtration requirement for any non-trivial \( \mu_{(0)} \).

More specifically, note that degradation is preserved under mixture of distributions, Corollary 32 (or the stringy tree lemma) implies \( \mu_0 \preceq \mu_1 \) for any degree distribution. Then by Proposition 48 and induction, we have \( \mu_h = Q^h \mu_0 \preceq Q^h \mu_1 = \mu_{h+1} \).

Hence, from the filtration argument, the limiting distribution of \( \{ \mu_h \}_{h \in \mathbb{N}} \) is well defined, and we denote it by \( \mu^* \).

By the first property in Proposition 13, we have \( \mu_0 \) is non-trivial. Then by Proposition 45 and transitivity of degradation, \( \mu^* \succeq \mu_0 \) is also non-trivial. On the other hand, because the BP operator commutes with weak convergence, \( \mu^* \) is a BP fixed point. Therefore, from Theorem 3, \( \mu^* \) and \( \mu^* \) must be identical. Consequently, we meet the conditions of Proposition 49 and the convergence of \( \mu_{(h)} \) to \( \mu^* \) is proved.

Note that any BP fixed point can be viewed as the limiting distribution in a robust reconstruction setting with itself as initialization, it has to be a degraded version of \( \mu^* \). Thus, by antisymmetry, \( \mu^* \) is identical to any dominant BP fixed point. Therefore, we have proved that all distributions in Theorem 35 are well defined and identical to \( \mu^* \).

Remark 11. A less straightforward, but more intuitive proof for the convergence result in robust reconstruction is to consider the sequence of degradation indices \( \phi_h \triangleq \phi^*(\mu^*, \tilde{\mu}_{(h)}) \), which can be viewed as a potential function that captures the stabilization effect of BP recursion. Corollary 32 shows its monotonicity, and Theorem 10 shows that any positivity of \( \phi_h \) leads to a strict decrease within two steps of recursion. The first statement implies that \( B_{\phi_h} \boxtimes \mu^* \) converges. Then the second statement combined with the following continuity proposition shows that the converged distribution has to be \( B_0 \boxtimes \mu^* = \mu^* \).

**Proposition 36.** If a sequence of LLR distributions \( \{ \mu_n \}_{n \in \mathbb{N}} \) converges to a non-trivial \( \mu \), then the following equation holds for any symmetric \( \nu \).

\[
\lim_{n \to \infty} \phi^*(\nu, \mu_n) = \phi^*(\nu, \mu).
\]

Proposition 36 is proved in Appendix J. Note that the same conclusion does not hold if \( \nu \) is \( n \)-dependent.

Remark 12. Recall that we assumed the nontrivial cases where \( \mathbb{P}[d \leq 1] = 1 \). Similar to the basic setting, Theorem 35 does not always hold in trivial cases. The only exception is when \( \mathbb{P}[d = 1] = 1 \) and \( \delta \in \{0, 1\} \), where the limiting distribution for any robust reconstruction problem is identical to its initialization, and the dominant BP fixed point is identical to \( \mu^* \) and \( B_0 \).

**B. Boundary Irrelevance for Broadcast with Survey**

We first present a definition of boundary irrelevance in terms of LLR distributions.

**Definition 37.** For any degree distribution \( P_d, \delta \in [0, 1] \), and symmetric non-trivial survey distribution \( \mu_s \), let \( Q_s \) be the associated BP operator. We say boundary irrelevance (BI) is satisfied if both \( \mu_{(h)} \triangleq Q^h B_0 \) and \( \mu_{(h)} \triangleq Q^h B_2 \) weakly converges to the same distribution on domain \( (-\infty, +\infty] \).

Note that \( \mu_{(h)} \) represents the LLR distribution for estimation with full leaf information, and \( \mu_{(h)} \) represents the corresponding LLR distribution with no leaf information. The above definition essentially states that ignoring leaf information will not affect estimation as \( h \to \infty \), which is consistent with the notation of BI defined in the literature. In particular, one can show that our definition is equivalent to the version in [13], and is stronger than the error-probability guarantee in [14]. Therefore, we present the following Theorem, which simultaneously proves Conjecture 1 in [13] and Conjecture 1 in [14].

**Theorem 38.** BI holds for any combination of \( P_d, \delta \in [0, 1] \), and symmetric non-trivial \( \mu_s \).

**Proof.** By directly following the definition, we have \( \mu_{(h)} \preceq \mu_{(h+1)} \) and \( \mu_{(h)} \succeq \mu_{(h+1)} \). Hence, by the filtration argument, both sequences converge and the converged distributions are symmetric BP fixed points. Then BI directly follows from Theorem 16, which states that all symmetric BP fixed points are identical.

Boundary irrelevance can be viewed as a stronger claim compared to the independence of initialization and estimation for robust reconstruction. Particularly, because the LLR distribution that corresponds to any leaf-observation channel is always sandwiched between \( \mu_{(h)} \) and \( \mu_{(h)} \), we can directly apply Proposition 49 and obtain the following statement.

**Corollary 39.** Consider a robust reconstruction problem where survey information is also provided, characterized by parameters \( P_d, \delta \), non-trivial symmetric \( \mu_s \), and any choice of leaf-observation channels. The resulting LLR distributions always converge to the unique symmetric BP fixed point.

**C. Uniqueness and Convergence in the Large \( d \) Limit**

We consider a recursion process characterized by the following operator.
Definition 40. For any symmetric $\mu_s$ and any distribution $P_d$ on domain $[0, +\infty]$, let $Q_d$ denote the following operator.

$$Q_d \mu = E_{P_d} \left[ N(\overline{d} \cdot V_{\mu}) \right] \ast \mu_s,$$

where $N(s) \triangleq N\left(\frac{s}{2}, s\right)$ for any $s \in [0, +\infty]$ and $V_{\mu} \triangleq E \left[ 4\tanh \left( \frac{R}{2} \right) \right]$ for $R \sim \mu$.

This operator was considered in [15] as the limit of $Q_d$ (or $Q$ when $\mu_s$ is trivial) as $\delta \rightarrow \frac{1}{2}$, where the degree distribution $P_d$ is parameterized by $\delta$, and $\overline{d} \triangleq d(1-\delta)^2$ converges in distribution to $P_d$ on domain $[0, +\infty]$. Similar to earlier sections, one can define the fixed point equation to be $\mu = Q_d \mu$ and define BP recursion as $\mu_{(h+1)} = Q_d \mu_{(h)}$ for both basic and robust reconstruction.

To extend our earlier results to $Q_d$, we prove its contractivity in terms of the degradation index. In particular, note that the contraction implied by the BP operator is non-multiplicative, a careful investigation is needed to show that strict inequalities in $\beta$-curves are maintained in the limit of large $d$. We present the results in following theorem, and provide a proof in Appendix K.

Theorem 41. Consider the large $d$ regime defined by any $P_d$ and any symmetric $\mu_s$.

1) There is at most one unique non-trivial symmetric BP fixed point. (Uniqueness of non-trivial BP fixed point)
2) Non-trivial symmetric BP fixed point exists if and only if $E[\overline{d}] \in (1, +\infty)$ or $\mu_s$ is non-trivial. (Existence of non-trivial BP fixed point)
3) BP recursion with any initialization converges to the dominant BP fixed point. (Independence of Convergence and Initialization)
4) When $\mu_s$ is non-trivial, the above convergence statement also applies for trivial initialization. (Boundary Irrelevance)

Remark 13. The uniqueness statement in the above theorem proves Conjecture 2.6 in [15], by applying the special case where $P_d$ is a delta distribution and $\mu_s = B_\alpha$. Generally, our formulation does not assume $d$ scales with $(1-2\delta)^{-2}$ in high probability. The sublinear and superlinear components of $P_d$ are naturally captured by non-zero mass points in $P_d$ at $0$ and $+\infty$.

D. Full Characterizations of Accuracy and Entropy in Stochastic Block Model

Consider a 2-SBM problem with a set of $n$ vertices $V = \{v_1, v_2, \ldots, v_n\}$. Let $X_v$ denote the label on vertex $v$, $X$ denotes the collection of all labels, and $G$ denotes the random graph generated based on the SBM. The goal in this setting is to design algorithms that use the random graph to produce an estimate of $X$.

There are two main quantities of interests. For any estimator $\hat{X}$, its estimation accuracy, denoted by $acc_n(\hat{X})$, is defined as follows [12].

$$acc_n(\hat{X}) \triangleq \frac{1}{2} + \frac{1}{n} \sum_{v \in V} \left| X_v - \hat{X}_v \right| - \frac{1}{2}. \quad (35)$$

In particular, note that the conditioned graph distribution is invariant under a global bit flip of hidden labels. No algorithm can achieve a non-trivial estimation in the expected number of correctly estimated labels. The accuracy defined in equation (35) captures the correlation between the partitions induced by the labels, which removes the global bit-flip effect.

Note that $acc_n(\hat{X})$ is random. The quantity $p_G(a, b)$ was introduced in [12] to measure the performance of estimators, defined as the maximum accuracy that can be achieved by any estimator for large $n$ with non-zero probability. The problems of interests are to characterize $p_G(a, b)$ and to prove whether it can be achieved by any algorithm with high-probability. Both were only resolved when $a$ and $b$ satisfy certain conditions. However, with the leaf-independence result proved in Section III-A, the proofs in [16], [12], [17] can be extended to all regimes.

The other quantity was called the SBM entropy [13], denoted by $H(a, b)$, which is defined to be the limit of the normalized conditional entropy of all labels $X$ given the graph $G$, as $n \rightarrow \infty$.

The SBM entropy also characterizes the normalized mutual information between the labels and the graph, of which the limit for large $n$ was denoted by $I(a, b)$. Similar to the accuracy metric, it was an open problem to characterize $H(a, b)$ and $I(a, b)$ for all parameter values using BP fixed points. It was pointed out in [13] that a complete characterization can be obtained once the BI result stated in Section III-B is proved.

To summarize, we have the following theorem, which strengthens Theorem 2.9 in [12] and Theorem 1 in [13].

Theorem 42. For any $a$ and $b$,

1) we have $p_G(a, b) = p(\mu_{a,b}^\ast)$, where $\mu_{a,b}^\ast$ is the dominant BP fixed point for the broadcast on tree problem with $P_d = \text{Pois}(a + b)$ and $\delta = \frac{b}{a + b}$, and $p(\mu_{a,b}^\ast) \triangleq \mu_{a,b}^\ast(0, +\infty) + \frac{1}{2} \mu_{a,b}^\ast(\{0\})$;
2) there is an algorithm that achieves $p_G(a, b)$ with high probability, i.e., with $acc_n(\hat{X})$ converges in probability to $p_G(a, b)$ as $n \rightarrow \infty$;
3) we have $H(a, b) = \log 2 - I(a, b) = \int_0^1 E_{R \sim \mu_{a,b}} \left[ \log(2 \cosh \left( \frac{R}{2} \right)) - \frac{R}{2} \tanh \left( \frac{R}{2} \right) \right] dR$, where $\mu_{a,b}$ is the unique BP fixed point for the broadcast with survey setting with $P_d = \text{Pois}(a + b)$, $\delta = \frac{b}{a + b}$, and BEC survey $\mu_s = \epsilon B_0 + (1 - \epsilon) B_\perp$, where $H(a, b) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} H(X|G)$. 

$^5$Formally, $H(a, b) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} H(X|G)$. 

E. Stochastic Block Model with Side Information

Consider a variant of the 2-SBM formulation, where the estimator has additional access to a noisy version of all hidden labels. Similar to the broadcast with survey setting, each label is observed through an independent symmetric channel, and we denote their LLR distribution by $\mu_s$.

In the presence of this side information, a different notion of accuracy was considered in the literature. In [14], [15], the authors considered estimators that asymptotically maximizes the expected fraction of correctly estimated labels. Formally, we denote this function by $p_n(\hat{X})$, which can be defined by the following equation.

$$p_n(\hat{X}) \triangleq \mathbb{E} \left[ \frac{1}{n} \sum_{v \in V} |X_v - \hat{X}_v| \right] = \frac{1}{n} \sum_{v \in V} \mathbb{P} \left[ X_v = \hat{X}_v \right].$$

(36)

The estimation accuracy defined in equation (36) can be maximized by applying the ML estimator individually for each $X_v$. However, the ML estimator becomes computationally intractable when the graph is large as it relies on global information. Therefore, local algorithms has been studied, and they have been conjectured to be optimal in the sparse regime [14], [15]. In particular, for any fixed parameter $t \in \mathbb{N}_+$, an algorithm is called $t$-local if it estimates each $X_v$ only using the information within the subgraph induced by vertices with a distance from $v$ less than $t$. Such local information resembles the distribution of local observation in the broadcast with survey setting as $n \to \infty$ for fixed $a$ and $b$, up to a graph isomorphism. Hence, one can estimate each $X_v$ using the same belief propagation whenever the graph is locally tree-like.

Local BP is asymptotically optimal among local algorithms. We present the following theorem, which states that there are no gaps between the estimation accuracies of local and global algorithms.6

**Theorem 43.** For 2-SBM with any fixed $a$, $b$ and side information generated based on any non-trivial $\mu_s$, we have

$$\lim_{t \to \infty} \lim_{n \to \infty} p_n(\hat{X}^{(t)}_{\text{BP}}) = \lim_{n \to \infty} p^*_n,$$

(37)

where $\hat{X}^{(t)}_{\text{BP}}$ is any estimator that runs local BP with parameter $\delta = \frac{b}{a+b}$, and $p^*_n$ is the optimal estimation accuracy over all estimators.

**Remark 14.** Theorem 43 implies the correctness of two related conjectures in [14], [15], which can be obtained by specializing the observation model $\mu_s$ to the cases of binary erasure channels or binary symmetric channels. The proof of Theorem 43 can be established by first connecting SBM to the broadcast with survey setting, similar to the approach presented in [15, Section 2.4]. Formally, one can show that Lemma 3.7 and Lemma 3.9 in [15] holds for general symmetric $\mu_s$, which bound the accuracies on both sides of equation (37) using convergence distributions of BP recursion. As a consequence, the equality condition of equation (37) is implied if all convergence distributions are identical, which is essentially the BI condition stated in Section III-B. In this work, we proved that the BI condition holds for all regimes (Theorem 38), which completes the proof of optimality of local BP algorithms.

APPENDIX A

**Properties of Degradation**

The concept of degradation has been studied as early as in [23], [24], [25], [26], under a topic called comparison of experiments. It has also appeared later in [27], [28], [21], [29], [30], [31], [32] for communication channels. Degradation commonly defines a preorder in the literature. However, the relationship given by Definition 5 satisfies antisymmetry as well. We state the following proposition, which is proved in Appendix L. This property is due to the fact that LLR contains the minimum information as a sufficient statistic for estimation. Hence, any equivalence class of LLR distributions defined based on the degradation preorder has a unique element.

**Proposition 44.** Degradation defines a partial order on the set of all symmetric distributions.

Degradation commutes with weak convergence, stated as follows, which can be proved by the sequential compactness of joint distributions. Note that here we consider the most general formulation where distributions can have non-zero mass at $+\infty$, similar to [33], [34], we adopt the natural definition of weak convergence over domain $(-\infty, +\infty]$.

**Proposition 45.** For any two sequences of distributions $\{\mu_n\}_{n \in \mathbb{N}}$ and $\{\nu_n\}_{n \in \mathbb{N}}$ that weakly converge to $\mu$ and $\nu$ respectively and satisfy $\mu_n \preceq \nu_n$ for any $n$, we have $\mu \preceq \nu$.

Degradation also commutes with several elementary operations, including convolution, box convolution, and their compositions.7 They can be proved by considering the binary-input-channel equivalence of the related distributions, and the needed

---

6In certain parts of [15], a generalized setting was considered, where $a$, $b$ are $n$-dependent. It is clear that the same generalization is not considered in their Conjecture 1, otherwise the stated limits may not converge. However, one can still prove a similar optimality result using the BI presented in Section III-B and III-C, stated in terms of the absolute difference between estimation accuracies. More generally, this asymptotical optimality can hold whenever the local tree-like condition is satisfied for large $n$.

7The definition of box convolution can be found in Section I-A.
Markov chain constructions naturally follow from the physical interpretations of these operations. In particular, box convolution can be viewed as channel concatenation (see Remark 2), and convolution can be viewed as parallel observation.

**Proposition 46.** For any \(\phi \in [0, \frac{1}{2}]\) and any distributions \(\mu\) and \(\nu\), the statements \(\mu \preceq \nu\) and \(B_\phi \boxplus \mu \preceq B_\phi \boxplus \nu\) are equivalent.

**Proposition 47.** For any distributions \(\mu\), \(\nu\), and \(\tau\) satisfying \(\mu \preceq \nu\), we have \(\tau \boxast \mu \preceq \tau \boxast \nu\) and \(\tau \boxplus \mu \preceq \tau \boxplus \nu\).

**Proposition 48.** For any \(\delta \in [0, 1]\) and degree distribution \(P_\delta\), let \(Q\) be the operator defined in Section I. Then for any two symmetric distributions \(\mu \preceq \nu\), we have
\[
Q\mu \preceq Q\nu.
\]

Finally, we present the following proposition, which can be used to prove the convergence of LLR distributions. Proposition 49 can be proved using the sequential compactness of joint distributions and antisymmetry of degradation.

**Proposition 49.** For any sequences of distributions \(\{\mu_n\}_{n \in \mathbb{N}}, \{\nu_n\}_{n \in \mathbb{N}}, \{\pi_n\}_{n \in \mathbb{N}}\) satisfying \(\mu_n \preceq \nu_n \preceq \pi_n\) for any \(n\) and \(\mu_n, \pi_n\) weakly converges to \(\mu^*\) for some \(\mu^*\), we have \(\mu_n\) weakly converges to \(\mu^*\).

**Remark 15.** Commonly, to provide a rigorous definition for distributions that are not discrete, degradation is defined based on the existence of a stochastic transformation function, which satisfies certain measure theoretic requirements (see [26]) and serves as the conditional distribution. Our proof can be made independent of such concepts by defining \(\mu \preceq \nu\) in terms of the existence of two sequences of distributions, with each pair of corresponding terms satisfying the same degradation order, and each weakly converges to \(\mu\) and \(\nu\), respectively.

**Remark 16.** Degradation can be defined for asymmetric distributions, and a formal definition is provided in Appendix B. All properties stated within this appendix hold under this extension, except for the statements on box convolution, which requires symmetry condition by definition. Therefore, we omit all symmetric distribution requirements whenever no ambiguity arises.

### Appendix B

**Uniqueness of BP Fixed Points for General Distributions and Its Implications**

We first present a proof of Corollary 4. For brevity, we define the following class of distributions.

**Definition 50.** For any distribution \(\mu\) defined on domain \((-\infty, +\infty]\), we call it an LLR distribution, if we can find a distribution \(\mu^-\) to satisfy equation (4). We call \(\mu^-\) the complement distribution of \(\mu\).

**Proposition 51.** The following statements hold true.

1. All symmetric distributions are LLR distributions.
2. Any LLR distribution has a unique complement.
3. Any distribution \(\mu\) is an LLR distribution if and only if \(\mathbb{E}_{R \sim \mu}[e^{-R}] \leq 1\).

The notion of degradation can be generalized to LLR distributions. One can verify that the following definition is consistent with Definition 5. Moreover, the antisymmetry property holds as explained in Remark 19.

**Definition 52.** For any two LLR distributions \(\mu_Y, \mu_Z\), we say \(\mu_Y\) is a degraded version of \(\mu_Z\), denoted by \(\mu_Y \preceq \mu_Z\), if one can define joint distributions \(\mu_{Y,Z}\) and \(\mu_{Y^-, Z^-}\), with \(\mu_Y, \mu_Z\), and their complements being the marginal distributions, such that \(\mu_{Y|Z}\) and \(\mu_{Y^-|Z^-}\) are identical.

**Proposition 53.** Degradation defines a partial order on the set of all LLR distributions.

**Proof of Corollary 4.** The corollary is proved by showing that any non-trivial BP fixed point is identical to the limiting LLR distribution for the broadcast on tree formulation. Specifically, we extend the proof steps of Theorem 35 based on the fact that all needed properties of degradation hold for general LLR distributions.

Consider any non-trivial BP fixed point \(\mu\). By setting \(\mu_{(0)} = \mu\), we have \(\mu \preceq \mu_{(h)}\), which is proved by the generalized version of Proposition 48. This essentially implies that \(\mu \preceq \mu^*\), which follows from the generalized version of Proposition 45.

On the other hand, because \(\mu^*\) is symmetric, we can directly extend the definition of degradation index \(\phi^*(\mu^*, \mu)\) according to equation (8). The first and the second property in Proposition 13 hold under this extension\(^8\), implying that \(\mu_{(0)}\) is non-trivial and \(\mu_{(0)} \preceq \mu\). Note that \(\mu_{(0)}\) is symmetric, we have already shown that \(\mu_{(h)}\) converges to \(\mu^*\). Therefore, by the generalized version of Proposition 48 and Proposition 49, we have \(\mu = \mu^*\).

**Remark 17.** The same extension can be proved for all other main results. In particular, Theorem 16, Theorem 35, Corollary 39, and Theorem 41 do not rely on the symmetry condition of BP fixed points or that of the BP initialization.

\(^8\)See Remark 18 for the proof of the first property.
APPENDIX C
CONVERGENCE IN DEGRADATION METRIC

Recall the definition of degradation metric \( d(\mu, \nu) \) and \( \beta \)-curves \( \beta(t; \mu) \) (see Remark 5 and Definition 18). We have the following fact.

**Proposition 54.** For any sequence of symmetric distributions \( \{\mu_n\}_{n \in \mathbb{N}} \) and any symmetric \( \mu \), the convergence of degradation metric \( \lim_{n \to \infty} d(\mu_n, \mu) = 0 \) implies that

1. \( \mu_n \) converges weakly to \( \mu \),
2. \( \lim_{n \to \infty} \sup_t |\beta(t; \mu_n) - \beta(t; \mu)| = 0 \).

**Proof.** We first prove the second property. The proof follows from the fact that both \( \phi^*(\mu_n, \mu) \) and \( \phi^*(\mu, \mu_n) \) converges to 0 as \( d(\mu_n, \mu) \to 0 \), and each degradation index function implies a uniform bound on \( \beta(t; \mu) - \beta(t; \mu_n) \).

Consider any pair of symmetric distributions \( \mu \) and \( \nu \), we derive a uniform bound on \( \beta(t; \mu) - \beta(t; \nu) \) using \( \phi^*(\mu, \nu) \). Recall the second property of Proposition 13 states that

\[
B_{\phi^*(\mu, \nu)} \supseteq \mu \preceq \nu.
\]

By applying Theorem 21 and equation (15), we have

\[
\beta(t; \nu) \geq (1 - 2\phi^*(\mu, \nu))\beta\left(\frac{t}{1 - 2\phi^*(\mu, \nu)}; \mu\right).
\]

(38)

Note that any \( \beta \)-curve is a non-negative, non-decreasing function of \(|t|\) that is upper bounded by 1 for \(|t| \leq 1\). If \( \phi^*(\mu, \nu) < \frac{1}{2} \), we have the following inequality.

\[
\beta(t; \nu) \geq (1 - 2\phi^*(\mu, \nu))\beta(t; \mu) \geq \beta(t; \mu) - 2\phi^*(\mu, \nu)
\]

for \(|t| \leq 1\). (39)

Because \( \beta(t; \nu) = \beta(t; \nu) = |t| \) for \(|t| > 1\), we have obtained an bound of \( \sup_t (\beta(t; \mu) - \beta(t; \nu)) \), which converges to 0 as \( \phi^*(\mu, \nu) \to 0 \).

By symmetry, we also have

\[
\beta(t; \mu) \geq \beta(t; \nu) - 2\phi^*(\nu, \mu)
\]

(40)

for \( \phi^*(\nu, \mu) < \frac{1}{2} \). Therefore, the convergence of degradation index function implies that

\[
\lim_{n \to \infty} \sup_t |\beta(t; \mu_n) - \beta(t; \mu)| \leq \lim_{n \to \infty} 2(\phi^*(\mu, \nu) + \phi^*(\nu, \mu)) = 0.
\]

To prove the first property, we need to show the following convergence for any function \( f \) that is bounded and continuous on \((\infty, \infty)\).

\[
\lim_{n \to \infty} \mathbb{E}_{\mu_n}[f(R)] = \mathbb{E}_{\mu}[f(R)].
\]

By symmetry condition, the expectation of function \( f \) over any symmetric distribution is identical to the expectation of the following even function

\[
g(R) = \frac{e^R f(R) + f(-R)}{e^R + 1}.
\]

Any such \( g \) is bounded and continuous on \([-\infty, \infty] \). Thus, they can each be uniformly approximated by even Lipschitz functions of tanh \( \frac{R}{2} \). By linear interpolation, uniform convergence of \( \beta \)-curves implies convergence in expectation of even Lipschitz functions. Therefore, the proof is complete.

**APPENDIX D**
PROOF OF PROPOSITION 13

**Proof.** Here we provide the proof for the first property. By the definition of degradation index, we only need to show the existence of a \( \phi \in [0, \frac{1}{2}] \) that satisfies \( B_{\phi} \supseteq \mu \preceq \nu \). We prove this fact by choosing \( \phi = \mathbb{E}_{R \sim \nu}[\frac{1 - \text{sgn}(R)}{2}] \). Recall that \( \nu \) is non-trivial and symmetric. We have \( \phi \in [0, \frac{1}{2}] \).

Following the commutativity of box convolution and the physical interpretation, we have \( B_{\phi} \supseteq \mu \preceq B_{\phi} \). On the other hand, \( B_{\phi} \) can be viewed as the LLR distribution of the 1-bit maximum likelihood estimator for the estimation problem characterized by \( \nu \). Hence, there is a natural joint distribution that implies \( B_{\phi} 
preceq \nu \). Then we have \( B_{\phi} \supseteq \mu \preceq \nu \) from the transitivity of degradation.

**Remark 18.** The above proof can be generalized for asymmetric \( \nu \), by choosing \( \phi \) as follows.

\[
\phi = \inf_{f \mid \nu} \max \{\mathbb{P}_{Y \sim \nu}[Z = 0], \mathbb{P}_{Y \sim \nu^{-}}[Z = 1]\},
\]

where \( \nu^{-} \) is the complement distribution of \( \nu \). From elementary statistics, it is known that \( B_{\phi} \) is the LLR distribution of a thresholding quantized version of the estimation problem characterized by \( \nu \), where the quantizer is chosen to achieve the minimax error probability. Therefore, we have \( B_{\phi} \preceq \nu \), and the rest of the proof follows the same steps.
Appendix E
Proof of Proposition 20

Proof. Proof of 1) ⇒ 3). From \( \nu < \mu \) and the non-trivial condition, we can choose \( \phi \in (0, \frac{1}{2}) \) to satisfy \( \nu \leq B_\phi \boxplus \mu \). By either applying Theorem 21 or Jensen’s inequality over any joint distribution consistent with Definition 5, we have \( \beta(t; \nu) \leq \beta(t; B_\phi \boxplus \mu) \) for any \( t \in \mathbb{R} \). Using equation (15), we have

\[
(1 - 2\phi)\beta(t; \mu) \geq \beta((1 - 2\phi)t; \nu).
\]

Therefore, following the definition of \( t_{\max} \) and non-trivial condition, we have \( t_{\max}(\nu) \leq (1 - 2\phi)t_{\max}(\mu) < t_{\max}(\mu) \). On the other hand, note that any \( \beta \)-curve is 1-Lipschitz. For \( t \in [0, t_{\max}(\mu)] \), we have

\[
\beta((1 - 2\phi)t; \nu) \geq \beta(t; \nu) - 2\phi t.
\]

\[
> \beta(t; \nu) - 2\phi \beta(t; \mu).
\]

Combining the above inequalities, we obtain the stated gap condition.

Proof of 3) ⇒ 2). If \( t_{\max}(\nu) < t_{\max}(\mu) \), then \( [0, t_{\max}(\nu)] \) is a subset of \( [0, t_{\max}(\mu)] \). Thus, the same gap condition applies.

Proof of 2) ⇒ 1). Recall that both \( \mu \) and \( \nu \) are non-trivial. The \( \beta \)-curves for both distributions are positive and continuous over \( \mathbb{R} \). Therefore, the following quantity is well-defined

\[
\phi \triangleq \min_{t \in [0, t_{\max}(\nu)]} \frac{1}{2} \left( 1 - \frac{\beta(t; \nu)}{\beta(t; \mu)} \right).
\]

From the gap condition, we have \( \phi \in (0, \frac{1}{2}) \). It remains to show \( \nu \leq B_\phi \boxplus \mu \).

One can verify that the \( \beta \)-curve is non-decreasing for any measure. Hence, for any \( t \in [0, t_{\max}(\nu)] \), the above definition implies the following inequality.

\[
\beta(t; \nu) \leq (1 - 2\phi)\beta(t; \mu) \leq (1 - 2\phi)\beta \left( \frac{t}{1 - 2\phi}; \mu \right).
\]

Using equation (15), the above inequality can be written as

\[
\beta(t; \nu) \leq \beta(t; B_\phi \boxplus \mu).
\]

Because the \( \beta \)-curve is lower bounded by the identity function, recall the definition of \( t_{\max}(\nu) \), inequality (41) holds for \( t > t_{\max}(\nu) \) as well. Apply this conclusion to Theorem 21, we have proved that \( \nu \leq B_\phi \boxplus \mu \), which implies strict degradation.

\[\square\]

Appendix F
Proof of Proposition 25

Proof. For convenience, we define

\[
\tau_{(T)} = B_\phi \boxplus (B_{\delta_1} \ast B_{\delta_2}),
\]

\[
\tau_{(S)} = (B_\phi \boxplus B_{\delta_1}) \ast (B_\phi \boxplus B_{\delta_2}).
\]

To prove Proposition 25, we first derive closed-form expressions for the \( \beta \)-curves. Note that \( \tau_{(T)} \) is a symmetric discrete distribution that can be written as a linear combination of \( B_{\delta_{\alpha}} \) and \( B_{\delta_{\beta}} \) for some \( 0 \leq \delta_{(T)}^+ \leq \delta_{(T)}^- \leq \frac{1}{2} \). From Proposition 22, the \( \beta \)-curve of \( \tau_{(T)} \) must be a piecewise linear function with at most two corner points. Concretely, let\(^9\)

\[
t_{\tau_{(T)}}(\delta_{(T)}^-) \triangleq t_{\max} \left( B_{\delta_{(T)}^-} \right) = (1 - 2\phi) \tanh \frac{r_{\max}(B_{\delta_1}) - r_{\max}(B_{\delta_2})}{2},
\]

\[
t_{\tau_{(T)}}(\delta_{(T)}^+) \triangleq t_{\max} \left( B_{\delta_{(T)}^+} \right) = (1 - 2\phi) \tanh \frac{r_{\max}(B_{\delta_1}) + r_{\max}(B_{\delta_2})}{2},
\]

\[
\beta_{(T),0} \triangleq \beta \left( 0; \tau_{(T)} \right) = (1 - 2\phi) \max\{(1 - 2\delta_1), (1 - 2\delta_2)\}.
\]

The value-function pair \( \{t, \beta(t; \tau_{(S)})\} \) for \( t \in [0, 1] \) is on the lower convex envelope of the following finite set of points.

\[
\{ (0, \beta_{(T),0}), \left( t_{\tau_{(T)}}^-, \beta_{(T),0} \right), \left( t_{\tau_{(T)}}^+, \beta_{(T),0} \right), (1, 1) \}.
\]

\(^9\)When \( r_{\max}(B_{\delta_1}) = r_{\max}(B_{\delta_2}) = +\infty \), \( \tau_{(T)} \) is simply \( B_{\delta_{(T)}} \), and we can let \( t_{\tau_{(T)}}^+ \) take any value in \( [0, t_{\tau_{(T)}}^+] \).
Similarly, let
\[ t_{(S)}^- \triangleq t_{\max}(B_{\delta}) = \tanh \left( \frac{F_\phi(r_{\max}(B_{\delta})) - F_\phi(r_{\max}(B_{\delta^2}))}{2} \right), \]
\[ t_{(S)}^+ \triangleq t_{\max}(B_{\delta}) = \tanh \left( \frac{F_\phi(r_{\max}(B_{\delta})) + F_\phi(r_{\max}(B_{\delta^1}))}{2} \right), \]
\[ \beta_{(S),0} \triangleq \beta(0; \tau_{(S)}) = (1 - 2\phi) \max\{(1 - 2\delta_1), (1 - 2\delta_2)\}. \]

Then function \( \beta(t; \tau_{(S)}) \) for \( t \in [0, 1] \) is given by the lower convex envelope of the following finite set.
\[ \left\{ (0, \beta_{(S),0}), (t_{(S)}^-, \beta_{(S),0}), (t_{(S)}^+, \beta_{(S),0}), (1, 1) \right\}. \]

By some elementary calculus\(^{10}\), one can prove the following equation and inequalities, which shows that \( \beta(t; \tau_{(T)}) \leq \beta(t; \tau_{(S)}) \) for all \( t \in \mathbb{R} \).
\[ t_{(S)}^- \leq t_{(T)}^-, \]
\[ t_{(S)}^+ \geq t_{(T)}^+, \]
\[ \beta_{(S),0} = \beta_{(T),0}. \]

Moreover, we always have \( t_{(S)}^+ > t_{(T)}^+ \) except when any of \( \phi, \delta_1, \delta_2 \) is \( \frac{1}{2} \). In all cases, this leads to \( \beta(t; \tau_{(T)}) < \beta(t; \tau_{(S)}) \) for any \( t \in (t_{(S)}^-, t_{(T)}^+) \). Note that \( t_{(S)}^+ = \tanh \frac{\phi}{2} \) and \( t_{(T)}^+ = t_{\max}(\tau_{(T)}) \). We have \( \tau_{(T)} \preceq_{s_{\min}} \tau_{(S)}. \)

**APPENDIX G**

**PROOF OF PROPOSITION 27**

**Proof.** Recall that \( \beta \)-curves are characterized by the error probabilities of maximum a posteriori (MAP) estimators. We construct an estimation problem with prior \( X \sim \text{Ber}(\frac{t + 1}{2}) \), and independent observations \( Y, Z \) such that \( Y \) is measured through a symmetric channel characterized by \( \mu \) and \( Z \) is measured through a BSC with crossover probability \( \phi \). Let \( \hat{X} \) be the MAP estimator. Because the LLR distribution of this experiment is characterized by \( B_{\phi} \ast \mu \), we have
\[ \mathbb{P}[\hat{X} = X] = \frac{1 + \beta(t; B_{\phi} \ast \mu)}{2}. \]

Conditioned on \( Z = 0 \) or \( Z = 1 \), the inference problem reduces to estimating \( X \) given \( Y \) with different priors. Note that the MAP estimator remains the same. We have
\[ \mathbb{P}[\hat{X} = X| Z = 0] = \frac{1 + \beta(t_0; B_{\phi} \ast \mu)}{2}, \]
\[ \mathbb{P}[\hat{X} = X| Z = 1] = \frac{1 + \beta(t_1; B_{\phi} \ast \mu)}{2}. \]

Combining above results, we obtain the following equation, which is identical to the needed statement.
\[ \beta(t; B_{\phi} \ast \mu) = \mathbb{P}[Z = 0] \beta(t_0; B_{\phi} \ast \mu) + \mathbb{P}[Z = 1] \beta(t_1; B_{\phi} \ast \mu). \]

**APPENDIX H**

**PROOF OF PROPOSITION 29**

**Proof.** Consider any fixed \( r \in (s - \ell, r_{\max}(\mu) - \ell) \). Let \( t = \tanh \left( \frac{|r|}{2} \right) \), we have \( t \in [0, 1] \). Then we can apply Proposition 27 to evaluate \( \beta(t; \tau \ast \mu) - \beta(t; \tau \ast \nu) \) by writing \( \tau \) as an integration of \( B_{\phi} \) distributions. In particular, let \( Z \sim \tau \) and \( \phi = \frac{1}{e^{\pi r^2 + 1}} \).

In general, we have
\[ \beta(t; \tau \ast \mu) = \mathbb{E} \left[ \left( \frac{1 + t - 2t\phi}{2} \right) \beta(t_0; \mu) + \left( \frac{1 - t + 2t\phi}{2} \right) \beta(t_1; \mu) \right] \]
for any symmetric \( \mu \), where
\[ t_0 = \tanh \left( \frac{|Z| + |r|}{2} \right), \]
\[ t_1 = \tanh \left( \frac{|Z| - |r|}{2} \right). \]

\(^{10}\)In particular, the concavity of \( F_\phi \) on \( \mathbb{R}_{\geq 0} \).
When \( \nu \prec_s \mu \), we always have \( \beta(t_0; \nu) \leq \beta(t_0; \mu) \) and \( \beta(t_1; \nu) \leq \beta(t_1; \mu) \). To prove \( \beta(t; \tau \ast \nu) < \beta(t; \tau \ast \mu) \) using equation (42), it remains to show that \( \beta(t_0; \nu) < \beta(t_0; \mu) \) or \( \beta(t_1; \nu) < \beta(t_1; \mu) \) with non-zero probability. Let \( tZ = \tanh \left( \frac{t + Z}{2} \right) \).

Because \( r + \ell \in (s, r_{\max}(\mu)) \) and \( \ell \in \text{supp}(\tau) \), we have \( \beta(tZ; \mu) > \beta(tZ; \nu) \) for \( Z \) in a neighbourhood of \( \ell \), which holds with non-zero probability. Note that \( tZ \in \{t_0, t_1\} \). The needed inequality is proved.

For the second statement, we can assume \( \mu \) is non-trivial, otherwise the stated interval is an empty set and nothing needs to be proved. Note that in this case, \( \nu \prec \mu \) can be viewed as an extension of \( \nu \prec_s \mu \) by allowing \( s \) to choose any value within \( (-r_{\max}(\mu), 0) \). Therefore, following exactly the same steps, we have the needed strict inequality holds for \( r \in (s - \ell, r_{\max}(\mu) - \ell) \) for any such \( s \). Hence, inequality (17) holds if \( r \) belongs to \( (-r_{\max}(\mu) - \ell, r_{\max}(\mu) - \ell) \), which contains \[ -\ell, r_{\max}(\mu) - \ell \] as a subset.

**APPENDIX I**

**PROOF OF PROPOSITION 34**

**Proof.** We apply induction over \( k \in \mathbb{N} \). The base case \( k = 1 \) follows directly from Proposition 33. Indeed, when \( s_1 < 0 \), it implies

\[
B_\phi \boxplus Q_s \mu = B_\phi \boxplus (B_\delta \boxplus \mu_\ell) \prec (B_\phi \boxplus B_\delta \boxplus \mu_\ell) \prec (B_\phi \boxplus B_\delta \boxplus \mu) = Q_s(B_\phi \boxplus \mu)
\]

as \( s_{\min} = F_\phi(r_{\max}(B_\delta \boxplus \mu)) - r_s = F_\phi(F_\delta(r_{\max}(B_\delta \boxplus \mu))) - r_s = s_1 < 0 \). Here we used the fact that \( r_{\max}(B_\tau \boxplus \nu) = F_\tau(r_{\max}(\nu)) \) and

\[
F_\phi \circ F_\delta = F_\phi \circ F_\delta,
\]

due to commutativity of \( B_\phi \boxplus B_\delta \). Similarly, when \( s_1 \geq 0 \), the statement of Proposition 33 implies \( B_\phi \boxplus Q_s \mu \prec s_1 Q_s(B_\phi \boxplus \mu) \).

Assume Proposition 34 holds for some \( k \in \mathbb{N} \). We prove that it holds for \( k + 1 \). As everywhere before, our method is to start from a non-strict degradation chain given by

\[
B_\phi \boxplus (Q_s^{k+1} \mu) \leq Q_s(B_\phi \boxplus Q_s^k \mu) \leq Q_s(Q_s^k(B_\phi \boxplus \mu))
\]

and keep track of areas where the comparison of \( \beta \)-curves is strict. Denote for convenience \( \mu_k = Q_s^k \mu, \nu_k = Q_s^k(B_\phi \boxplus \mu) \) and \( r_k = r_{\max}(\mu_k) \). We will repeatedly use the fact that \( r_{\max}(B_\tau \boxplus \nu) = F_\tau(r_{\max}(\nu)) \), so that for example \( r_{k+1} = F_\delta(r_k) + r_s \).

Observe the first step of inequality (44), which essentially states that

\[
B_\phi \boxplus ((B_\delta \boxplus \mu_\ell) \boxplus \mu_k) \prec (B_\delta \boxplus B_\phi \boxplus \mu) \boxplus \mu_k.
\]

This is simply an instance of applying Proposition 33. Thus, we have that the comparison of \( \beta \)-curves is strict for

\[
F_\phi(F_\delta(r_k)) - r_s < s < F_\phi(F_\delta(r_k)) + r_s.
\]

Therefore, we can assume \( r_s \) is finite for the rest of the proof because otherwise we have already established strict comparison for the entire \( s \in \mathbb{R} \), which includes \( [0, r_{\max}(B_\phi \boxplus \mu_{k+1})] \) as a subset.

Now we analyze the second step in (44). From the induction hypothesis we know that

\[
B_\phi \boxplus \mu_k \prec_{s_k} \nu_k
\]

for \( s_k \geq 0 \), and \( B_\phi \boxplus \mu_k \prec \nu_k \) otherwise. Applying box convolution with \( B_\delta \) to both sides of these inequalities and by equation (15), we obtain

\[
B_\delta \boxplus B_\phi \boxplus \mu_k \prec_{s_k'} B_\delta \boxplus \nu_k,
\]

for \( s_k' = F_\delta(s_k) \geq 0 \), and \( B_\delta \boxplus B_\phi \boxplus \mu_k \prec B_\delta \boxplus \nu_k \) otherwise. Next, by convolving with \( \mu_s \) on both sides, and then Proposition 29 with \( \ell = r_s \), we get that

\[
(B_\delta \boxplus B_\phi \boxplus \mu_k) \boxplus \mu_s \leq (B_\delta \boxplus \nu_k) \boxplus \mu_s = \nu_{k+1}
\]

with inequality for \( \beta \)-curves strict for all \( t = \tanh \left( \frac{t + Z}{2} \right) \) with

\[
s_{k+1} = s_k - r_s < s < F_\delta(r_{\max}(\nu_k)) - r_s.
\]

From induction hypothesis (46) we have \( r_{\max}(\nu_k) > r_{\max}(B_\phi \boxplus \mu_k) = F_\phi(r_k) \). Therefore, because of the strictness of the inequality we have that (47) and (45) together imply comparison of first and last \( \beta \)-curves in (44) for all \( t = \tanh(|s|/2) \) with

\[
s_{k+1} < s < F_\phi(F_\delta(r_k)) + r_s.
\]

Finally, notice that \( F_\phi(x + y) < F_\phi(x) + y \) for \( x \geq 0, y > 0 \), and \( \phi \in (0, 1) \). Thus the right-hand side of (48) is strictly bigger than \( r_{\max}(B_\phi \boxplus \mu_{k+1}) = F_\phi(r_{k+1}) = F_\phi(F_\delta(r_k) + r_s) \). In all, we have established strict comparison of \( \beta \)-curves for

\[
s_{k+1} < s \leq r_{\max}(B_\phi \boxplus \mu_{k+1})
\]

\( \square \).
We first prove equation (49) for the special case of $\nu$. Hence, it remains to show that

$$\limsup_{n \to \infty} \phi^*(\nu, \mu_n) \leq \phi^*(\nu, \mu).$$

We first prove equation (49) for the special case of $\nu = \mu$. In this case, it is clear that the RHS of equation (49) equals 0.

By weak convergence, we can find a sequence $\{\epsilon_n\}_{n \in \mathbb{N}}$ converging to 0 such that for any $n$ and any $r \in [0, +\infty]$, $\mu_n([-F_n(r), F_n(r)]) \leq (1 - 2\epsilon_n)\mu([-r, r]) + 2\epsilon_n$.

Hence, for any $\epsilon_n < \frac{1}{2}$, let $\tau_n$ be the symmetric distribution satisfying

$$\tau_n([-F_n(r), F_n(r)]) = (1 - 2\epsilon_n)\mu([-r, r]) + 2\epsilon_n,$$

by Skorokhod representation, we have $\tau_n \leq \mu_n$. More explicitly, consider independent variables $R \sim \mu$, $S \sim \text{Ber}(1 - 2\epsilon_n)$, $X \sim \text{Ber}(\epsilon_n)$, the distribution of $V = (-1)^X SF_n(R)$ gives an exact construction of $\tau_n$.

We utilize this construction to show $B_{2\epsilon_n(1 - \epsilon_n)} \boxplus \mu \leq \tau_n$. Let $U$ be a random variable that satisfies $U = \pm F_{2\epsilon_n(1 - \epsilon_n)}(R)$ each w.p. $\frac{1}{2}$ for $V = 0$, and $U = F_{2\epsilon_n}(V)$ otherwise. One can verify that $U \sim B_{2\epsilon_n(1 - \epsilon_n)} \boxplus \mu$, and the joint distribution of $U, V$ satisfies the needed degradation requirements.

By transitivity, we can conclude that $B_{2\epsilon_n(1 - \epsilon_n)} \boxplus \mu \leq \mu_n$. Hence,

$$\limsup_{n \to \infty} \phi^*(\mu, \mu_n) \leq \limsup_{n \to \infty} 2\epsilon_n(1 - \epsilon_n) = 0.$$  \hspace{1cm} (50)

Now for general $\nu$, we have $B_{\phi^*(\nu, \mu)} \boxplus \nu \leq \mu$. Recall the triangle inequality for degradation index.

$$\phi^*(\nu, \mu_n) \leq \phi^*(\mu, \mu_n) + \phi^*(\nu, \mu) - 2\phi^*(\mu, \mu_n)\phi^*(\nu, \mu).$$  \hspace{1cm} (51)

Applying equation (50), we have

$$\limsup_{n \to \infty} \phi^*(\nu, \mu_n) \leq \limsup_{n \to \infty} \phi^*(\mu, \mu_n) + \phi^*(\nu, \mu) = \phi^*(\nu, \mu),$$  \hspace{1cm} (52)

which proves the needed inequality. \qed

**APPENDIX K**

**PROOF OF THEOREM 41**

Note that the definition of degradation index remain unchanged. Theorem 41 can be proved using the same steps upon the following proposition.\(^{11}\)

**Proposition 55.** For any symmetric $\mu$ and $\nu$,

1) if $\nu \preceq \mu$, then $\mathcal{Q}_L \nu \preceq \mathcal{Q}_L \mu$;
2) we have $B_\phi \boxplus \mathcal{Q}_L \mu \prec \mathcal{Q}_L (B_\phi \boxplus \mu)$ for any $\phi \in (0, \frac{1}{2})$;
3) if $\nu$ is nontrivial, then $\phi^*(\mathcal{Q}_L \mu, \mathcal{Q}_L \nu) < \phi^*(\mu, \nu)$ or $\mu \preceq \nu$.

The first statement in Proposition 55 is proved by viewing $\mathcal{Q}_L$ as a limit of $\mathcal{Q}$ or $\mathcal{Q}_S$. The third statement can be proved using the second statement and Proposition 13. Therefore, we focus on the proof for the second statement, which is obtained by deriving and comparing related $\beta$-curve functions.

**Proof.** For brevity, we shall ignore cases where $\mu$ is trivial or $P_\tau$ is a delta distribution at 0, where the statement is obviously true. Hence, the LHS of the needed inequality is non-trivial, and it suffices to examine the $\beta$-curves on both sides due to Proposition 20. We first focus on the case where $\mu_t$ is trivial, which implies that any $\mathcal{Q}_L \mu$ is a mixture of Gaussians, and their $\beta$-curves can be written using integrals of elementary functions.

**Proposition 56.** For any $t \in [0, 1]$ and $s \in [0, +\infty]$, we have

$$\beta(t; \mathcal{N}(s)) = \max_{r \in [0, +\infty]} \frac{1}{2} \left( (1 - t) \text{erf} \left( \frac{\sqrt{s}}{2\sqrt{2}} - r \right) + (1 + t) \text{erf} \left( \frac{\sqrt{s}}{2\sqrt{2}} + r \right) \right).$$

\(^{11}\)Except for the existence of non-trivial fixed points, which can be proved by analyzing the evolution of $V_{\mu_t}$ with noiseless initialization.
where \( \text{erf}(z) \triangleq \frac{2}{\sqrt{\pi}} \int_0^z e^{-s^2} \, ds \), and a maximizer is given by \( r = r^*(s, t) \triangleq \sqrt{\frac{2}{s}} \arctanh(t) \). Moreover, for any fixed \( t \in [0, 1] \),

\[
\frac{d}{d\sqrt{s}} \beta(t; N(s)) = \sqrt{\frac{1 - t^2}{2\pi}} e^{-r^*(s, t)^2} \frac{1}{\sqrt{s}}.
\]

Due to convexity of \( \arctanh \), we have \( r^*(\theta^2 s, \theta t) \leq r^*(s, t) \) for any \( \theta \in (0, 1) \) and \( t \in [0, 1] \). Therefore, the following proposition is implied by the Lagrange mean value theorem.

**Proposition 57.** The following inequality holds for any \( \theta \in (0, 1) \), \( t \in [0, \theta] \), and \( s \in (0, +\infty) \):

\[
\beta(t; N(\theta^2 s)) > \theta \beta(t/\theta; N(s)).
\]

From linearity of \( \beta \)-curves, for \( \mu_s \) being the trivial distribution, we have

\[
\beta(t; Q_{t \mu}) = \mathbb{E}_{\mu_s} \left[ \beta(t; N \left( \bar{d} \cdot V_{t \mu} \right)) \right].
\]

For brevity, let \( \theta \triangleq (1 - 2\phi)^2 \). Note that \( V_{B_\phi \otimes \mu} = \theta^2 V_{\mu} \) and recall Proposition 22. For any \( t \in [0, \theta] \) we have

\[
\beta(t; B_\phi \otimes Q_{t \mu}) = \frac{\theta^2}{\theta} \mathbb{E}_{\mu_s} \left[ \beta(t/\theta; N \left( \bar{d} \cdot V_{t \mu} \right)) \right],
\]

\[
\beta(t; Q_{t \mu}) = \mathbb{E}_{\mu_s} \left[ \beta(t; N \left( \bar{d} \cdot \theta^2 V_{t \mu} \right)) \right].
\]

Recall that we only need to provide a prove for \( \mathbb{P}[\bar{d} > 0] > 0 \) and non-trivial \( \mu \). We have \( V_{\mu} > 0 \) and \( t_{\max}(B_\phi \otimes Q_{t \mu}) = \theta \). By an integration argument, the above results implies that

\[
\beta(t; Q_{t \mu}) > \beta(t; B_\phi \otimes Q_{t \mu})
\]

for any \( t \in [0, t_{\max}(B_\phi \otimes Q_{t \mu})] \), which proves the needed statement. \( \square \)

**Appendix L**

**Proof of Proposition 44**

**Proof.** While reflexivity and transitivity are straightforward, we prove antisymmetry by constructing potential functions that are monotone w.r.t. this order, and showing that the equality condition requires identity in distribution. We first prove that for any variables \( Y, Z \) with distributions satisfying \( \mu_Y \preceq \mu_Z \) and for any \( \alpha \in (0, 1) \), we have \( \mathbb{E}[e^{-\alpha Y}] \geq \mathbb{E}[e^{-\alpha Z}] \).

Consider any joint distribution \( \mu_{Y,Z} \) that satisfies the conditions in Definition 5. By the symmetry condition of \( Z \), for any measurable \( A \subseteq (-\infty, +\infty) \) we have

\[
\mathbb{P}[-Y \in A] = \mathbb{E}[e^{-Z} \cdot \mathbb{I}(Y \in A)] + \mathbb{P}[-Y \in A \text{ and } Z = +\infty].
\]

Then by the symmetry condition of \( Y \),

\[
\mathbb{P}[-Y \in A] = \mathbb{P}[e^{-Y} \cdot \mathbb{I}(Y \in A)].
\]

Hence, we have the following inequality w.p.1,\(^{12}\)

\[
e^{-Y} \geq \mathbb{E}[e^{-Z} | Y].
\]

Note that Jensen’s inequality implies

\[
\mathbb{E}[e^{-Z} | Y]^\alpha \geq \mathbb{E}[e^{-\alpha Z} | Y].
\]

Combining the above results, we have proven that

\[
\mathbb{E}[e^{-\alpha Y}] \geq \mathbb{E}[\mathbb{E}[e^{-Z} | Y]^\alpha] \geq \mathbb{E}[e^{-\alpha Z}],
\]

If \( \mu_Y \preceq \mu_Z \) and \( \mu_Z \preceq \mu_Y \), we have \( \mathbb{E}[e^{-\alpha Y}] = \mathbb{E}[e^{-\alpha Z}] \). Then, the equality conditions for the above steps need to be satisfied. As a consequence, \( |e^{-\alpha Y} - e^{-\alpha Z}| \equiv 0 \), which implies \( \mu_Y = \mu_Z \). \( \square \)

**Remark 19.** The above proof can be generalized to cover all LLR distributions. In particular, a proof of Proposition 53 can be obtained by replacing any \( -Y \) in the proof with a variable \( Y^- \) that follows the complement distribution of \( \mu_Y \), and any \( Z \) jointly distributed with \( Y^- \) with a variables \( -Z^- \) such that the joint distribution \( \mu_{Y^-, Z^-} \) satisfies the conditions stated in Definition 52.

\(^{12}\)Rigorously, one can take the supremum when defining conditional expectations to avoid differentiability issues.
REFERENCES


