Revisiting Algebra and Complexity of Inference in Graphical Models

Siamak Ravanbakhsh, University of Alberta
Russell Greiner, University of Alberta

Graphical Models use the intuitive and well-studied methods of graph theory to implicitly represent dependencies between variables in large systems and model the global behaviour of a complex system by specifying only local factors. This paper studies the form and complexity of inference in graphical models using the abstraction offered by algebraic structures. In particular, we broadly formalize inference problems in graphical models by viewing them as a sequence of commutative semigroups and study their computational complexity. This is in part achieved by organizing the problems in an inference hierarchy, such that problems in the same class of the hierarchy also belong to the same complexity class. When the underlying structure of an inference problem is a commutative semiring – i.e., a combination of two commutative semigroups with the distributive law – a message passing procedure can leverage this additional structure to perform inference. The resulting procedure, known as belief propagation allows exact inference for tree structures and efficient approximations for general graphs. We prove that for general graphs, exact inference in any commutative semiring is \( \text{NP} \)-hard under randomized reduction. Moreover, we show that direct application of distributive law for exact inference on tree structures is limited to inference problems with two commutative semigroups. We then study a more involved form of message passing, called survey propagation, from an algebraic point of view and express its updates using a combination of two commutative semirings.

1. INTRODUCTION

Many complicated global system corresponds to a graphical structure of interacting local functions. Many fields have (almost independently) discovered this: graphical models have been used in bioinformatics (protein folding, pedagogy trees, regulatory networks), neuroscience (formation of associative memory and neuroplasticity), communication theory (low density parity check codes), statistical physics (physics of dense matter and spin-glass theory), image processing (inpainting, stereo/texture reconstruction, denoising and super-resolution), compressed sensing, robotics (particle filters), sensor networks, social networks, natural language processing, speech recognition, combinatorial optimization and artificial intelligence (artificial neural networks, Bayesian networks). Two general perspectives have emerged from these varied approaches to local computation, namely variational [Wainwright and Jordan 2007] versus algebraic [Aji and McEliece 2000] perspectives on graphical models. These two perspectives are to some extent “residuals” from the different origins of research in AI and statistical physics.

In the statistical study of physical systems, Boltzmann distribution relates the probability of each state of a physical system to its energy, which is often decomposed due to local interactions [Mézard et al. 1987; Mézard and Montanari 2009]. These studies have been often interested in modeling systems at thermodynamic limit of infinite variables and the average behaviour through the study of random ensembles. Inference techniques with this origin (e.g., mean-field and cavity methods) are asymptotically exact under these assumptions. Most importantly these studies have reduced inference to optimization through the notion of free energy –a.k.a. variational approach.

In contrast, graphical models in the AI community have emerged in the study of knowledge representation and reasoning under uncertainty [Pearl 1988]. These advances are characterized by their attention to the theory of computation and logic [Bacchus 1991], where interest in computational (as opposed to analytical) solutions has motivated the study of approximability, computational complexity [Cooper 1990; Roth 1993] and invention of inference techniques such as belief propagation that are efficient and exact on tree structures. But most relevant to the topic of this paper is the
fact that, these studies have lead to algebraic abstractions in modeling systems that allow local computation [Shenoy and Shafer 1990; Lauritsen and Jensen 1997].

The common foundation underlying these two approaches is information theory, where derivation of probabilistic principles from logical axioms [Jaynes 2003] leads to notions such as entropy and divergences that are closely linked to their physical counter-parts i.e., entropy and free energies in physical systems. At a less abstract level, it was shown that inference techniques in AI and communication are attempting to minimize (approximations to) free energy [Yedidia and Freeman 2001; Aji and McEliece 2001].

Another exchange of ideas between the two fields was in the study of critical phenomena in random constraint satisfaction problems by both computer scientists and physicists [Fu and Anderson 1986; Mitchell et al. 1992; Monasson et al. 1999]; Satisfiability is at the heart of theory of computation and an important topic to investigate reasoning in AI. On the other hand, study of critical phenomena and phase transitions is central in statistical physics of disordered systems. This was culminated when a variational analysis lead to discovery of survey propagation [Mezard et al. 2002] for constraint satisfaction, which significantly advanced state-of-the-art in solving random satisfiability problems.

Despite this convergence, variational and algebraic perspectives are to some extent complementary – e.g., variational approach does not extend beyond (log) probabilities, while the algebraic approach cannot justify application of message passing to graphs with loops. This paper is concerned with the algebraic approach. We organize/generalize the previous work on the algebra of graphical models and give some new results on the complexity and limit of inference in this framework. To this end, section 2 broadly formalizes the problem of inference using factor-graphs and commutative semigroups. Section 3 organizes a subset of inference problems into an inference hierarchy with increasing levels computational complexity under PSPACE. Section 5 reviews the distributive property that make efficient inference possible and derive a negative result regarding the application of this properties beyond commutative semirings. Section 6 introduces an algebraic interpretation of survey propagation that generalizes its application to new settings.

2. THE PROBLEM OF INFERENCE

We use commutative semigroups to both define what a graphical model represents and also to define inference over this graphical model. The idea of using structures such as semigroups, monoids and semirings in expressing inference has a long history [Lauritzen and Jensen 1997; Schiex et al. 1995; Bistarelli et al. 1999]. Our approach based on factor-graphs [Kschischang and Frey 2001] and commutative semigroups, generalizes a variety of previous frameworks, including Markov networks [Clifford 1990], Bayesian networks [Pearl 1985], Forney graphs [Forney Jr 2001], hybrid models [Dechter and Larkin 2001], influence diagrams [Howard and Matheson 2005] and valuation networks [Shenoy 1992].

In particular the combination of factor-graphs and semigroups that we consider here generalizes the plausibility, feasibility and utility framework of Pralet et al. [2007], which is explicitly reduced to all the graphical models above and many more. The main difference in our approach is in keeping the framework free of semantics (e.g., decision and chance variables, utilities, constraints), that are often associated with variables, factors and operations without changing the expressive power. This notions can later be associated with individual inference problems to help with interpretation.

Definition 2.1. A **commutative semigroup** is a tuple \( G = (\mathcal{Y}^*, \otimes) \), where \( \mathcal{Y}^* \) is a set and \( \otimes : \mathcal{Y}^* \times \mathcal{Y}^* \rightarrow \mathcal{Y}^* \) is a binary operation that is (I) associative: \( a \otimes (b \otimes c) = \)
(a ⊗ b) ⊗ c and (II) commutative: a ⊗ b = b ⊗ a for all a, b, c ∈ Y*. If every element a ∈ Y* has an inverse a⁻¹ (often written 1/a), and an identity 1 such that a ⊗ a⁻¹ = 1, and a ⊗ 1 = a the commutative semigroup is an abelian group.

Here, the associativity and commutativity properties of a commutative semigroup make the operations invariant to the order of elements. In general, these properties are not “vital” and one may define inference starting from a magma.¹

Example 2.2. • the set of strings with concatenation operation form a semigroup with empty string as identity element. However this semigroup is not commutative.
• the set of natural numbers N with summation define a commutative semigroup.
• integers modulo n with addition defines an abelian group.
• the power-set 2^S of any set S, with intersection operation defines a commutative semigroup with S as its identity element.
• the set of natural numbers with greatest common divisor defines a commutative semigroup.
• in fact any semilattice is a commutative semigroup [Davey and Priestley 2002].
• given two commutative semigroups on two sets Y* and Z*, their Cartesian product is also a commutative semigroup.

Let z = (x₁, ..., x_N) be a tuple of N discrete variables xᵢ ∈ Xᵢ, where Xᵢ is the domain of xᵢ and z ∈ Z = X₁ × ... × Xₙ. Let I ⊆ N = {1, 2, ..., N} denote a subset of variable indices and z_I = {xᵢ | i ∈ I} ∈ X_I be the tuple of variables in z indexed by the subset I. A factor fᵢ : X_I → Yᵢ is a function over a subset of variables and Yᵢ = {fᵢ(z_I) | z_I ∈ X_I} is the range of this factor.

Definition 2.3. A factor-graph is a pair (F, G) such that
— F = {fᵢ} is a collection of factors with collective range Y = ⋃ᵢ Yᵢ.
— |F| = Poly(N) and |X_I| = Poly(N) ∀fᵢ ∈ F
— G = (Y*, ⊗) is a commutative semigroup, where Y* is the closure of Y w.r.t. ⊗.

It compactly represents the expanded (joint) form

q(z) = ⋂ᵢ fᵢ(zᵢ) \hspace{1cm} (1)

Note that the connection between the set of factors F and the commutative semigroup is through the “range” of factors. In practice, fᵢ(z_I) is often explicitly expressed for each z_I ∈ X_I as a ||I|-dimensional array. The requirements in the definition above ensure that 1) evaluation of the expanded form, q(z), polynomially tractable 2) F has a polynomial representation.

F can be conveniently represented as a bipartite graph, that includes two sets of nodes: variable nodes xᵢ, and factor nodes I. A variable node i (note that we will often identify a variable xᵢ with its index “i”) is connected to a factor node I if and only if i ∈ I – i.e., I is a set that is also an index. We will use ∂ᵢ to denote the neighbours of a variable or factor node in the factor graph – that is ∂ᵢ = {i | i ∈ I} (which is the set I) and ∂I = {I | i ∈ I}.

Example 2.4. figure 1 shows a factor-graph with 12 variables and 12 factors. Here z = (x₁, x₂, x₃, x₄, x₅, x₀, x₆, x₇, x₈, x₉, x₁₀, x₁₁) and ∂₁ = {1, j, k}, z_k = z_{(k,w,v)} and ∂ₗ = {l, v, w}. Assuming G₁ = (R, min), the expanded form represents q(z) = min{f₁(z₁), f₁(z₂), ..., f₁(z₁₁)}.

¹Magma [Pinter 2012] generalizes semigroup, as it does not require associativity property nor an identity element. Here, to use magma (in Definition 2.3), the elements of Y* and/or X should be ordered so as to avoid ambiguity in the order of pairwise operations over the set.
Now, assume that all variables are binary \( i.e., \mathcal{X} = \{0,1\}^{12} \) and \( q(x) \) is 12-dimensional hypercube, with one assignment at each corner. Also assume all the factors count the number of non-zero \( e.g., \) for \( z_W = (1,0,1,0,1,0,1,0,1,0,1,1) \in \mathcal{X} \), we have \( f_W(z_W) = 2 \).

Then, for the complete assignment \( z = (0,1,0,1,0,1,0,1,0,1,0,1) \in \mathcal{X} \), it is easy to check that the expanded form is \( q(z) = \min\{2,0,1,\ldots,1\} = 0 \).

A marginalization operation shrinks the expanded form \( q(z) \) using another commutative semigroup with binary operation \( \oplus \). Inference is a combination of an expansion and one or more marginalization operations, which can be difficult due to the exponential size of the expanded form.

**Definition 2.5.** Given a function \( q : \mathcal{X}_J \to \mathcal{Y} \), and a commutative semigroup \( \mathcal{G} = (\mathcal{Y}^*, \oplus) \), where \( \mathcal{Y}^* \) is the closure of \( \mathcal{Y} \) w.r.t. \( \oplus \), the marginal of \( q \) for \( I \subset J \) is

\[
q(\mathcal{X}_J \setminus I) \overset{\text{def}}{=} \bigoplus_{\mathcal{X}_I} q(\mathcal{X}_J)
\]

where \( \bigoplus_{\mathcal{X}_I} q(\mathcal{X}_J) \) is short for \( \bigoplus_{\mathcal{x}_I \in \mathcal{X}_I} q(\mathcal{X}_J) \), and it means to get \( q(\mathcal{X}_J \setminus I) \) for each \( \mathcal{X}_I \), the operation \( \oplus \) is performed over the set of all the values that the tuple \( \mathcal{X}_I \) can take.

We can think of \( q(\mathcal{X}_J) \) as a \(|J|\)-dimensional tensor and marginalization as performing \( \oplus \) operation over the axes in the set \( I \). The result is another \(|J \setminus I|\)-dimensional tensor (or function) that we call the **marginal**. Here if the marginalization is over all the dimensions in \( J \), we denote the marginal by \( q(\emptyset) \) instead of \( q(\mathcal{X}_J) \) and call it the **integral** of \( q \).

Now we define an inference problem as a sequence of marginalizations over the expanded form of a factor-graph.

**Definition 2.6.** An **inference problem** seeks

\[
q(\mathcal{X}_J_0) = \bigoplus_{\mathcal{X}_{J_M}} \bigoplus_{\mathcal{X}_{J_{M-1}}} \cdots \bigoplus_{\mathcal{X}_{J_1}} \bigotimes_{I} f_I(\mathcal{X}_I)
\]

where

\( - \mathcal{Y}^* \) is the closure of \( \mathcal{Y} \) (the collective range of factors), w.r.t. \( \oplus, \ldots, \oplus \) and \( \otimes \).
— $G_m = (Y^m, \bigoplus)$ for $1 \leq m \leq M$ and $G_e = (Y^e, \otimes)$ are all commutative semigroups.
— $J_0, \ldots, J_L$ partition the set of variable indices $N$.
— $q(\mathbf{x}_{J_0})$ has a polynomial representation in $N$.

Note that $1, \ldots, M$ are referring to potentially different operations as each belong to a different semigroup. When $J_0 = \emptyset$, we call the inference problem integration (denoting the inquiry by $q(\emptyset)$) and otherwise we call it marginalization. Here, having a constant sized $J_0$ is not always enough to ensure that $q(\mathbf{x}_{J_0})$ has a polynomial representation in $N$. This is because the size of $q(\mathbf{x}_{J_0})$ for any individual $\mathbf{x}_{J_0} \in \mathcal{X}_{J_0}$ may grow exponentially with $N$ (e.g., see claim 3.2).

**Example 2.7.** Going back to example 2.7, the shaded region in figure 1 show a partitioning of the variables that we use to defines the following inference problem:

$$q(\mathbf{x}_{J_0}) = \max_{\mathbf{x}_{J_0}} \prod_{I} \min \min_{I} f_{I}(\mathbf{x}_{I})$$

We can associate this problem with the following semantics: we may think of each factor as an agent, where $f_{I}(\mathbf{x}_{I})$ is the payoff for agent $I$, only depends on a subset of variables $\mathbf{x}_{I}$. We have adversarial variables $\mathbf{x}_{A}$, environmental or chance variables $\mathbf{x}_{E}$, controlled variables $\mathbf{x}_{C}$, and query variables $\mathbf{x}_{Q}$. The inference problem above for each query $\mathbf{x}_{Q}$, seeks to maximize the expected minimum payoff of all agents, without observing the adversarial or chance variables, and assuming the adversary makes its decision after observing control and chance variables.

**Example 2.8.** A “probabilistic” graphical model is defined using an expansion semigroup $G_e = (\mathbb{R}^\geq, \times)$ and often a marginalization semigroup $G_m = (\mathbb{R}^\geq, +)$ and the expanded form represents the unnormalized joint probability $q(\mathbf{z}) = \prod_{I} f_{I}(\mathbf{x}_{I})$, whose marginal probabilities are simply called marginals. Replacing the summation with marginalization semigroup $G_m = (\mathbb{R}^\geq, \max)$, seeks the maximum probability state and the resulting integration problem $q(\emptyset) = \max_{\mathbf{z}} \prod_{I} f_{I}(\mathbf{x}_{I})$ is known as maximum a posteriori (MAP) inference. Alternatively by adding a second marginalization operation to the summation, we get the marginal MAP inference

$$q(\mathbf{x}_{J_0}) = \max_{\mathbf{x}_{J_0}} \prod_{I} f_{I}(\mathbf{x}_{I}).$$

where here $\otimes = \prod$, $\bigoplus = \sum$ and $\bigoplus = \max$.

If the object of interest is the negative log-probability (a.k.a. energy), the product expansion semigroup is replaced by $G_e = (\mathbb{R}, \text{sum})$. Instead of sum marginalization semigroup, we can use the log-sum-exp semigroup, $G_m = (\mathbb{R}, +)$ where $a \oplus b \overset{\text{def}}{=} \log(e^{-a} + e^{-b})$. The integral in this case is the log-partition function. If we change the marginalization semigroup to $G_m = (\mathbb{R}, \text{min})$, the integral is the minimum energy (corresponding to MAP).

The operation $\min$ (also $\max$) is a choice function, which means $\min_{a \in A} a \in A$.

The implication is that if $\bigoplus$ in definition 2.6 is $\min$ (or $\max$), we can replace it with $\arg_{\mathbf{x}_{J_0}} \max$ without changing the inference problem. For example, in performing min-max inference we may ask for $\mathbf{x}^* = \arg_{\mathbf{x}} \min \max_{I} f_{I}(\mathbf{x}_{I})$
3. THE INFERENCE HIERARCHY

Often, the complexity class is concerned with the decision version of the inference problem in definition 2.6. The decision version of an inference problem asks a yes/no question about the integral: \( q(\emptyset) \geq q \) for a given \( q \).

Here, we produce a hierarchy of inference problems in analogy to polynomial [Stockmeyer 1976], the counting [Wagner 1986] and arithmetic [Rogers 1987] hierarchies.

To define the hierarchy, we assume the following in the definition 2.6:

- any two consecutive marginalization operations are distinct (\( l \neq l+1 \) \( \forall 1 < l < M \)).
- the marginalization index sets \( J_i \) \( \forall 1 \leq l \leq M \) are non-empty. Moreover if \( |J_l| = O(\log(N)) \) we call this marginalization operation a polynomial marginalization as here \( |X_J| = \text{Poly}(N) \).
- In defining the factor-graph we required each factor to be polynomially computable.

In building the hierarchy we require the operations over each semigroup to be polynomially computable as well. To this end we consider \( \mathbb{V} = \{0, \ldots, M \} \). Note that this automatically eliminates semigroups that involve operations such as exponentiation and logarithm, (because \( \mathbb{Q} \) is not closed under these operations) and only consider summation, product, minimization and maximization.

We can always re-express any inference problem to enforce the first two conditions and therefore they do not impose any restriction.

Define five inference families \( \Sigma, \Pi, \Phi, \Psi, \Delta \). The families are associated with that "last" marginalization operation – i.e., \( \oplus \) in definition 2.6. \( \Sigma \) is the family of inference problems where \( \oplus = \text{sum} \). Similarly, \( \Pi \) is associated with product, \( \Phi \) with minimization and \( \Psi \) with maximization. \( \Delta \) is the family of inference problems where the last marginalization is polynomial (i.e., \( |I_M| = O(\log(N)) \)) regardless of \( \oplus \).

Now we define inference classes in each family, such that all the problems in the same class have the same computational complexity. Moreover, the hierarchy is exhaustive – i.e., it includes all inference problems with four operations sum, min, max and product that are within \( \text{PSPACE} \) – and inference classes are exclusive. For this, each family is parameterized by a subscript \( M \) and two sets \( S \) and \( D \) (e.g., \( \Phi_M(S, D) \) is an inference "class" in family \( \Phi \)). As before, \( M \) is the number of marginalization operations, \( S \) is the set of indices of the (exponential) sum-marginalization and \( D \) is the set of indices of polynomial marginalizations.

**Example 3.1.** Sum-min-sum-product is a short notation for the decision problem

\[
\sum_{x_{J_3}} \min_{x_{J_2}} \sum_{x_{J_1}} \prod_{i} f_i(x_i) \geq q
\]

where \( J_1, J_2 \) and \( J_3 \) partition \( \mathcal{N} \). Let\( s \) assume \( J_1 = \{2, \ldots, N \} \), \( J_2 = \{N + 1, \ldots, N \} \) and \( J_3 = \{1\} \). Since we have three marginalization operations \( M = 3 \). Here the first and second marginalizations are exponential and the third one is polynomial (since \( |J_3| \) is constant). Therefore \( D = \{3\} \). Since the only exponential summation is \( \sum_{x_{J_1}} = \sum_{x_{J_3}} \), \( S = \{1\} \). In our inference hierarchy, this problem belongs to the class \( \Delta_3(\{1\}, \{3\}) \).

Alternatively, if we assume \( J_1, J_2 \) and \( J_3 \) all linearly grow with \( N \), the corresponding inference problem becomes a member of \( \Sigma(\{1, 3\}, \emptyset) \).
Define the base members of families as
\[\begin{align*}
\Sigma_0(\emptyset, \emptyset) & \overset{\text{def}}{=} \{\text{sum}\} & \Phi_0(\emptyset, \emptyset) & \overset{\text{def}}{=} \{\text{min}\} \\
\Psi_0(\emptyset, \emptyset) & \overset{\text{def}}{=} \{\text{max}\} & \Pi_0(\emptyset, \emptyset) & \overset{\text{def}}{=} \{\text{prod}\} \\
\Delta_1(\emptyset, \{1\}) & \overset{\text{def}}{=} \{\text{sum} \setminus \text{sum}, \text{min} \setminus \text{min}, \text{max} \setminus \text{max}\}
\end{align*}\]

where none of these initial members are actual inference problems and only provide the basis for recursive definition of different families. The exception is \(\Delta_1(\emptyset, \{1\})\), which contains three inference problems.\(^2\) Now define the inference family members recursively

\[\begin{align*}
\forall |J_M| = \omega(\log(N)), M > 0 \\
\Sigma_{M+1}(S \cup \{M + 1\}, D) & \overset{\text{def}}{=} \{\text{sum} - \xi \mid \xi \in \Pi_M(S, D) \cup \Phi_M(S, D) \cup \Psi_M(S, D) \cup \Delta_M(S, D)\} \\
\Phi_{M+1}(S, D) & \overset{\text{def}}{=} \{\text{min} - \xi \mid \xi \in \Sigma_M(S, D) \cup \Pi_M(S, D) \cup \Psi_M(S, D) \cup \Delta_M(S, D)\} \\
\Psi_{M+1}(S, D) & \overset{\text{def}}{=} \{\text{max} - \xi \mid \xi \in \Sigma_M(S, D) \cup \Pi_M(S, D) \cup \Phi_M(S, D) \cup \Delta_M(S, D)\} \\
\Pi_{M+1}(S, D) & \overset{\text{def}}{=} \emptyset
\end{align*}\]

\[\forall |J_M| = O(\log(N)), M > 1 \text{ and } \oplus \in \{\text{sum}, \text{min}, \text{max}\} \]
\[\Delta_{M+1}(S, D \cup \{M + 1\}) \overset{\text{def}}{=} \{\oplus - \xi \mid \xi \in \Sigma_M(S, D) \cup \Pi_M(S, D) \cup \Psi_M(S, D) \cup \Phi_M(S, D) \cup \Delta_M(S, D)\}\]

In definition above, we ignored the inference problems in which product appears in any of the marginalization semigroups (e.g., product-sum). The following claim, explains this choice.

**Claim 3.2.** \(^3\) For \(\oplus_M = \text{prod}\), the inference query \(q(\mathcal{L}_0)\) can have an exponential representation in \(N\) and inference is outside \(\text{PSPACE}\).

\(\text{PSPACE}\) is the class of problems that can be solved by a (non-deterministic) Turing machine in polynomial space.

### 3.1. Single marginalization

The inference classes in the hierarchy with one marginalization are

\[\begin{align*}
\Delta_1(\emptyset, \{1\}) & = \{\text{min} \setminus \text{min}, \text{max} \setminus \text{max}, \text{sum} \setminus \text{sum}\} \quad (5) \\
\Psi_1(\emptyset, \emptyset) & = \{\text{max} \setminus \text{min}, \text{max} \setminus \text{sum}, \text{max} \setminus \text{prod}\} \quad (6) \\
\Phi_1(\emptyset, \emptyset) & = \{\text{min} \setminus \text{min}, \text{min} \setminus \text{sum}, \text{min} \setminus \text{prod}\} \quad (7) \\
\Sigma_1(\{1\}, \emptyset) & = \{\text{sum} \setminus \text{prod}, \text{sum} \setminus \text{sum}, \text{min} \setminus \text{max}\} \quad (8)
\end{align*}\]

Now we review all the problems above and proved that \(\Delta_1, \Psi_1, \Phi_1\) and \(\Sigma_1\) are complete w.r.t. \(\mathbb{P}, \text{NP}, \text{coNP}\) and \(\text{PP}\) respectively. Starting from \(\Delta_1\):

**Proposition 3.3.** sum-sum, min-min and max-max inference are in \(\mathbb{P}\).

\(^2\)We treat \(M = 1\) for \(\Delta\) specially as in this case the marginalization operation can not be polynomial. This is because if \(|J_1| = O(\log(N))\), then \(|J_0| = \Omega(N)\) which violates the conditions in the definition of the inference problem.

\(^3\)All proofs appear in the appendix.
Max-sum and max-prod are widely studied and it is known that their decision version is \( \text{NP-complete} \) [Shimony 1994]. By reduction from satisfiability we can show that max-min inference [Ravanbakhsh et al. 2014] is also \( \text{NP-hard} \).

**Proposition 3.4.** The decision version of max-min inference that asks 
\[
\max_x \min_q f_1(x) \geq q \text{ is } \text{NP-complete}.
\]

This means all the problems in \( \Psi_1(\emptyset, 0) \) are in \( \text{NP} \) (and in fact are complete w.r.t. their complexity class). In contrast, problems in \( \Phi_1(\emptyset, 0) \) are in \( \text{coNP} \). \( \text{coNP} \) is the class of decision problems in which the “NO instances” have a polynomial time verifiable witness or proof.

Note that by changing the decision problem from \( q(\emptyset) \geq q \) to \( q(\emptyset) \leq q \), the complexity classes of problems in \( \Psi \) and \( \Phi \) family are reversed (i.e., problems in \( \Phi_1(\emptyset, 0) \) become \( \text{NP-complete} \) and the problems in \( \Psi_1(\emptyset, 0) \) become \( \text{coNP-complete} \)).

To our knowledge the only complexity result available for \( \max \)-marginalization over \( \mathcal{X} \) [Park and Darwiche 2004]. Now suppose that the max-marginalization over \( \mathcal{X}_J \) is polynomial (e.g., \( |J| \) is constant). Then marginal-MAP belongs to \( \Delta_2(\{1\}, \{2\}) \) with complexity \( \text{P}^{\text{PP}} \). This is because a Turing machine can enumerate all \( \mathcal{X}_J \in \mathcal{X}_J \) in polynomial time and

\[\text{Example 3.6.} \text{ Consider the marginal-MAP inference of eq (4). The decision version of this problem, } q(\emptyset) \geq q, \text{ is a member of } \Psi^2(\{1\}, \emptyset) \text{ which also includes } \max - \text{sum} - \text{min} \text{ and } \max - \text{sum} - \text{max}. \text{ The complexity of this class according to eq (10) is } \text{P}^{\text{PP}}(\emptyset(\emptyset, 0)) = \text{NP}^{\text{PP}}. \text{ However, marginal-MAP is also known to be “complete” w.r.t. } \text{NP}^{\text{PP}} \text{ [Park and Darwiche 2004]. Now suppose that the max-marginalization over } \mathcal{X}_J \text{ is polynomial (e.g., } |J| \text{ is constant). Then marginal-MAP belongs to } \Delta_2((1), \{2\}) \text{ with complexity } \text{P}^{\text{PP}}. \text{ This is because a Turing machine can enumerate all } \mathcal{X}_J \in \mathcal{X}_J \text{ in polynomial time and}\]

\[\text{We do not prove the completeness w.r.t. complexity classes beyond the first level of the hierarchy and only assert the membership.}\]
call its PP oracle to see if 

\[ q(x_{J_0} \mid z_{J_2}) \geq q \]

where 

\[ q(x_{J_0} \mid z_{J_2}) = \sum_{x_{J_2}} \prod_{I} f_I(x_{I \setminus J_2}, z_{I \cap J_2}) \]

and accept if any of its calls to oracle accepts and rejects otherwise. Here, \( f_I(x_{I \setminus J_2}, z_{I \cap J_2}) \) is the reduced factor, in which all the variables in \( x_{J_2} \) are fixed to \( z_{J_2} \cap I \).

The example above also hints at the rationale behind the recursive definition of complexity class for each inference class in the hierarchy. Consider the inference family \( \Phi \): Adding an exponential-sized \( \min \)-marginalization to an inference problem with known complexity \( A \), requires a Turing machine to non-deterministically enumerate \( z_{J_M} \in X_{J_M} \) possibilities, then call the \( A \) oracle with the “reduced factor-graph” – in which \( x_{J_M} \) is clamped to \( z_{J_M} \) – and reject iff any of the calls to oracle rejects. In contrast, for inference family \( \Psi \), the Turing machine accepts iff any of the calls to the oracle accepts.

A similar argument explains the recursive definition of complexity for the family \( \Sigma \) and \( \Delta \). Here, Toda’s theorem [Toda 1991] has an interesting implication w.r.t. the hierarchy. This theorem states that \( \text{PP} \) is as hard as the polynomial hierarchy, which means \( \min - \max - \min - \ldots - \max \) inference for an arbitrary, but constant, number of \( \min \) and \( \max \) operations appears below the sum-product inference in the inference hierarchy.

3.3. Complexity of the hierarchy

By restricting the domain \( Y^\ast \) to \( \{0, 1\} \), \( \min \) and \( \max \) become isomorphic to logical AND (\( \land \)) and OR (\( \lor \)) respectively, where \( 1 \equiv \text{TRUE}, 0 \equiv \text{FALSE} \). By considering the restriction of the inference hierarchy to these two operations we can express quantified satisfiability (QSAT) as inference in a graphical model, where \( \land \equiv \forall \) and \( \lor \equiv \exists \). Let each factor \( f_i(x_i) \) be a disjunction – e.g., \( f(x_{i,j,k}) = x_i \lor \neg x_j \lor \neg x_k \). Then we have

\[
\forall x_{J_M} \exists x_{J_{M-1}} \ldots \exists x_{J_2} \exists x_{J_1} \prod_i f_i(x_i) \equiv \min_{x_{J_M}} \max_{x_{J_{M-1}}} \ldots \max_{x_{J_2}} \min_{x_{J_1}} f_i(x_i)
\]

By adding the summation operation, we can express the stochastic satisfiability [Littman et al. 2001] and by generalizing the constraints from disjunctions we can represent any quantified constraint problem (QCP) [Bordeaux and Monfroy 2002]. QSAT, stochastic SAT and QCPs are all \( \text{PSPACE} \)-complete. Therefore if we can show that inference in the inference hierarchy is in \( \text{PSPACE} \), it follows that inference hierarchy is \( \text{PSPACE} \)-complete as well.

**Theorem 3.7.** Inference hierarchy is \( \text{PSPACE} \)-complete.

4. EFFICIENT INFERENCE

Our definition of inference was based on an expansion operation and one or more marginalization operations. If we assume only a single marginalization semigroup, polynomial time inference is still not generally possible. However, if we further assume that the expansion operation is distributive over marginalization and the factor-graph has no loops, exact polynomial time inference is possible.

**Definition 4.1.** A commutative semiring \( \mathcal{S} = (Y^\ast, \oplus, \otimes) \) is the combination of two commutative semigroups \( \mathcal{S}_c = (\otimes, Y^\ast) \) and \( \mathcal{S}_m = (\oplus, Y^\ast) \) with two additional properties
identity elements 1 and 1 such that 1 ⊕ a = a and 1 ⊗ a = a. Moreover 1 is an annihilator for (S, +): a ⊗ 1 = 1 ∀a ∈ S.

— distributive property:

\[ a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes b) \forall a, b, c \in S\]

The mechanism of efficient inference using distributive law can be seen in a simple example: instead of calculating \( \min(a + b, a + c) \), using the fact that summation distributes over minimization, we may instead obtain the same result using \( a + \min(b, c) \), which requires fewer operations.

**Example 4.2.** The following are some examples of commutative semirings: • sum-product \((\mathbb{R}^{\geq 0}, +, \times)\) • max-product \((\mathbb{R}^{\geq 0} \cup \{-\infty\}, \max, \times)\) and \((\{0, 1\}, \max, \times)\), • min-max \((\mathbb{S}, \min, \max)\) on any ordered set \(\mathbb{S}\), • min-sum \((\mathbb{R} \cup \{\infty\}, \min, +)\) and \((\{0, 1\}, \min, +)\), • or-and \((\{\text{TRUE, FALSE}\}, \lor, \land)\), • union-intersection \((2^{\mathbb{S}}, \cup, \cap)\) for any powerset \(2^{\mathbb{S}}\), • the semiring of natural numbers with greatest common divisor and least common multiple \((\mathbb{N}, \text{lcm}, \text{gcd})\), • symmetric difference-intersection semiring for any powerset \((2^{\mathbb{S}}, \Delta, \cap)\).

Many of the semirings above are isomorphic – e.g., \(y' = -\log(y)\) defines an isomorphism between min-sum and max-product. It is also easy to show that or-and semiring is isomorphic to min-sum/max-product semiring on \(Y^* = \{0, 1\}\).

These inference problems have different properties indirectly inherited from their commutative semirings: for example, since both operations have inverses, sum-product is a **field** [Pinter 2012]. The availability of inverse for \(\otimes\) operation – i.e., when \(S\) is an abelian group – has an important implication for inference: the expanded form of eq (1) can be normalized, and we may inquire about normalized marginals

\[
p(x) = \bigoplus_{x \in J} p(x)
\]

where

\[
p(x) \overset{\text{def}}{=} \frac{1}{q(y)} \otimes \left( \bigotimes_{i} f_i(x_i) \right)
\]

That is, when working with normalized expanded form and normalized marginals, we always have \(\bigoplus_{x} p(x) = 1\).

**Example 4.3.** Since \(\mathcal{S} = (\mathbb{R}^{>0}, \times)\) and \(\mathcal{S} = (\mathbb{R}, +)\) are both abelian groups, min-sum and sum-product inference have normalized marginals. For min-sum inference this means \(\min_{\sum} p(x) = \frac{1}{\sum} = 0\). However, for min-max inference, since \((\mathbb{S}, \max)\) is not abelian, normalized marginals are not defined.

We can apply the identity and annihilator of a commutative semiring to define constraints. Our interest in constraints as special type of factors is solely because it enables us to establish the complexity of inference in commutative semirings.

**Definition 4.4.** A constraint is a factor \(f_1 : \mathcal{X} \rightarrow \{0, 1\}\) whose range is limited to identity and annihilator of the expansion monoid.

Here, \(f_1(x) = 1\) iff \(x\) is forbidden and \(f_1(x) = 1\) iff it is permissible. A **constraint satisfaction problem** (CSP) is any inference problem on a semiring in which all factors are constraints. Note that this allows definition of the “same” CSP on any commutative semiring. The idea of using different semirings to define CSPs has been studied in the
past [Bistarelli et al. 1999], however its implication about inference on commutative semirings has been ignored.

**THEOREM 4.5.** Inference in any commutative semiring is \( \mathbb{NP} \)-hard under randomized polynomial-time reduction.

**Example 4.6.** Inference on xor-and semiring \((\{\text{TRUE, FALSE}\}, \text{xor}, \land)\), where each factor is disjunction form is called parity-SAT. Parity-SAT asks whether the number of SAT solutions is even or odd. A corollary to theorem 4.5 is that parity-SAT is \( \mathbb{NP} \)-hard under randomized reduction, which is indeed the case [Valiant and Vazirani 1986].

In the following section we review the implications of the second semiring property and establish its limits beyond semirings.

### 5. DISTRIBUTIVE LAW

If the factor-graph is loop free, we can use distributive law to make inference tractable. Assuming \(q(x_k)\) (or \(q(x_i)\)) is the marginal of interest, form a tree with \(K\) (or \(k\)) as its root. Then starting from the leaves using distributive law we can move the \(\oplus\) inside the \(\otimes\) and define “messages” from leaves towards the root as follows:

\[
q_{i \to I}(x_i) = \bigotimes_{j \in \partial I \setminus I} q_{J \to i}(x_i) \tag{14}
\]

\[
q_{I \to i}(x_i) = \bigoplus_{j \in \partial \setminus i} f_i(z_j) \bigotimes_{j \in \partial \setminus i} q_{j \to i}(x_j) \tag{15}
\]

where eq (14) defines the message from a variable to a factor, closer to the root and similarly eq (15) defines the message from factor \(I\) to a variable \(i\) closer to the root. Here, distributive law allows moving the \(\oplus\) over the domain \(x_j \setminus i\) from outside to inside of eq (15) – the same way \(\oplus\) moves its place in \((a \otimes b) \oplus (a \otimes c)\) to give \(a \otimes (b \oplus c)\), where \(a\) is analogous to a message.

By starting from the leaves, and calculating the messages towards the root, we obtain the marginal over the root node as the product of incoming messages

\[
q(x_k) = \bigotimes_{i \in \partial k} q_{i \to k}(x_k) \tag{16}
\]
In fact, we can assume any subset of variables $x_A$ (and factors within those variables) to be the root. Then, the set of all incoming messages to $A$, produces the marginal

$$q(x_A) = \left( \bigotimes_{I \subseteq A} f_I(x_I) \right) \left( \bigotimes_{i \in A, J \in \partial_i, J \not\subseteq A} q_{J \rightarrow i}(x_i) \right)$$  (17)

**Example 5.1.** Consider the joint form represented by the factor-graph of figure 2

$$q(x) = \bigotimes_{A \in \{I, J, K, L, O, T, U, V, W, X, Y, Z\}} f_A(x_A)$$

and the problem of calculating the marginal over $x_{\{i,j,k\}}$ (i.e., the shaded region).

$$q(x_{\{i,j,k\}}) = \bigoplus_{x_{\{i,j,k\}}} \bigotimes_{A \in \{I, J, K, L, O, T, U, V, W, X, Y, Z\}} f_A(x_A)$$

We can move the $\oplus$ inside the $\otimes$ to obtain

$$q(x_{\{i,j,k\}}) = f_I(x_I) \otimes q_{L \rightarrow i}(x_i) \otimes q_{K \rightarrow i}(x_i) \otimes q_{V \rightarrow j}(x_j) \otimes q_{W \rightarrow j}(x_j) \otimes q_{K \rightarrow k}(x_k)$$

where each term $q_{A \rightarrow i}$ factors the summation on the corresponding sub-tree. For example

$$q_{L \rightarrow i} = \bigoplus_{x_I} f_L(x_I)$$

Here the message $q_{W \rightarrow j}$ is itself a computational challenge

$$q_{W \rightarrow j} = \bigoplus_{x_J} \bigotimes_{A \in \{W, U, Y, O, T, Z\}} f_A(x_A)$$

However we can also decompose this message over sub-trees

$$q_{W \rightarrow j} = \bigoplus_{x_J} f_A(x_A) \otimes q_{c \rightarrow W}(x_c) \otimes q_{r \rightarrow W}(x_r)$$

where again using distributive law $q_{c \rightarrow W}$ and $q_{r \rightarrow W}$ further simplify based on the incoming messages to the variable nodes $x_r$ and $x_c$.

This procedure is known as Belief Propagation (BP), which is sometimes prefixed with the corresponding semiring e.g., sum-product BP. Even though BP is guaranteed to produce correct answers when the factor-graph is a tree (and few other cases [Aji et al. 1998; Weiss 2001; Bayati et al. 2005; Weller and Jebara 2013]), it performs surprisingly well when applied as a fixed point iteration to graphs with loops [Murphy et al. 1999; Gallager 1962]. Here, for numerical stability, when the $\otimes$ operator has an inverse, the messages are normalized. We use $\propto$ to indicate this normalization accord-
Revisiting Algebra & Complexity of Inference in Graphical Models

...ing to the mode of inference

\[ \hat{p}_{i \rightarrow j}(x_i) \propto \bigoplus_{x_j} f_j(x_j) \bigotimes_{j \in \partial_i \setminus \{i\}} \hat{p}_{j \rightarrow i}(x_j) \propto P_{i \rightarrow j}(\hat{p}_{\partial_i \setminus \{i\}})(x_i) \] (18)

\[ \hat{p}_{i \rightarrow j}(x_i) \propto \bigotimes_{j \in \partial_i \setminus \{i\}} \hat{p}_{j \rightarrow i}(x_i) \propto P_{i \rightarrow j}(\hat{p}_{\partial_i \setminus \{i\}})(x_i) \] (19)

\[ \hat{p}(x_i) \propto f_i(x_i) \bigotimes_{i \in \partial_i} \hat{p}_{i \rightarrow j}(x_i) \] (20)

\[ \hat{p}(x_i) \propto \bigotimes_{i \in \partial_i} \hat{p}_{i \rightarrow j}(x_i) \] (21)

Here, for general graphs, \( \hat{p}(x_i) \) and \( \hat{p}(x_i) \) are approximations to \( p(x_i) \) and \( p(x_i) \) of eq (13). The functionals \( P_{i \rightarrow j}(\hat{p}_{\partial_i \setminus \{i\}})(x_i) \) cast the BP message updates as an operator on a subset of incoming messages – i.e., \( \hat{p}_{\partial_i \setminus \{i\}} = \{ \hat{p}_{j \rightarrow i} \mid J \in \partial_i \setminus \{i\} \} \). We use these functional notation in presenting the algebraic form of survey propagation in section 6.

5.1. The limits of message passing

By observing the application of distributive law in semirings, a natural question to ask is: can we use distributive law for polynomial time inference on loop-free graphical models over any of the inference problems at higher levels of inference hierarchy or in general any inference problem with more than one marginalization operation? The answer to this question is further motivated by the fact that, when loops exists, the same scheme may become a powerful approximation technique. When we have more than one marginalization operations, a natural assumption in using distributive law is that the expansion operation distributes over all the marginalization operations – e.g., as in min-max-sum, min-max-min, xor-and.

Consider the simplest case with three operators \( 1 \oplus, 2 \oplus \) and \( \otimes \), where \( \otimes \) distributes over both \( 1 \oplus \) and \( 2 \oplus \). Here the integration problem is

\[ q(\emptyset) = \bigoplus_{x_{J_2}} \bigoplus_{x_{J_1}} \prod_{i \in \partial_i} f_i(x_i) \]

where \( J_1 \) and \( J_2 \) partition \( \{1, \ldots, N\} \).

In order to apply distributive law for each pair \( 1 \oplus, \otimes \) and \( 2 \oplus, \otimes \), we need to be able to commute \( 1 \oplus \) and \( 2 \oplus \) operations. That is, we require

\[ \bigoplus_{x_A} \bigoplus_{x_B} g(x_{A \cup B}) = \bigoplus_{x_A} \bigoplus_{x_B} g(x_{A \cup B}) \] (22)

for some \( A \subseteq J_1 \) and \( B \subseteq J_2 \).

Now, consider a simple case involving two binary variables \( x_i \) and \( x_j \), where \( g(x_{(i,j)}) \) is

\[
\begin{array}{c|cc}
0 & a & b \\
1 & c & d \\
\end{array}
\]
Applying eq (22) to this simple case (i.e., \(A = \{i\}, B = \{j\}\)), we require \(\frac{1}{2} \frac{1}{2} \frac{1}{2} \frac{1}{2} = \frac{1}{2} \frac{1}{2} \frac{1}{2} \frac{1}{2} = (a \oplus b) \otimes (c \oplus d)\).

The following theorem leads immediately to a negative result:

**Theorem 5.2.** [Eckmann and Hilton 1962]:

\[
(a \frac{1}{2} b) \frac{1}{2} (c \frac{1}{2} d) = (\frac{2}{2} a \frac{1}{2} b) \frac{1}{2} (\frac{2}{2} c \frac{1}{2} d) \iff \frac{1}{2} \frac{1}{2} = \frac{1}{2} \frac{1}{2}.
\]

which implies that direct application of distributive law to tractably and exactly solve any inference problem with more than one marginalization operation is unfeasible, even for tree structures. This limitation was previously known for marginal MAP inference [Park and Darwiche 2004].

Min and max operations have an interesting property in this regard. Similar to any other operations for min and max we have

\[
\min \max_s (x_{i,j}) \neq \max \min_s (x_{i,j})
\]

However, if we slightly change the inference problem (from pure assignments \(x_{i,j} \in \{\pm 1\}\) to a distribution over assignments; a.k.a. mixed strategies), as a result of the celebrated minimax theorem [Von Neumann and Morgenstern 2007], the min and max operations commute — i.e.,

\[
\min \max_s (x_{i,j}) \sum_{x_{i,j}} s(x_{i}) g(x_{i,j}) s(x_i) = \max \min_s (x_{i,j}) \sum_{x_{i,j}} s(x_{i}) g(x_{i,j}) s(x_i)
\]

where \(s(x_{i,j})\) and \(s(x_i)\) are mixed strategies. This property has enabled addressing problems with min and max marginalization operations using message-passing-like procedures. For example, Ibrahimi et al. [2011] solve this (mixed-strategy) variation of min-max-product inference. Message passing procedures that operate on graphical models for game theory (a.k.a. “graphical games”) also rely on this property [Ortiz and Kearns 2002; Kearns 2007].

### 6. Algebra of Survey Propagation

Survey propagation (SP) was first introduced as a message passing solution to satisfiability [Braunstein et al. 2002a] and was later generalized to general CSP [Braunstein et al. 2002b] and arbitrary inference problems over factor-graphs [Mézard and Montanari 2009]. Several works offer different interpretations and generalizations of survey propagation [Kroc et al. 2002; Braunstein and Zecchina 2003; Maneva et al. 2004]. Here, we propose a generalization based on the same notions that extends the application of SP to arbitrary commutative semirings. Our derivation closely follows and generalizes the variational approach of Mézard and Montanari [2009].

As a fixed point iteration procedure, if BP has more than one fixed points, it may not converge at all. Alternatively, if the messages are initialized properly BP may converge to one of its fixed points. SP equations, take “all” BP fixed points into account. In our algebraic perspective, this accounting of all fixed points is using a third operation \(\otimes\).

In particular, we require that \(\otimes\) also distribute over \(\oplus\), forming a second commutative semiring. We refer to the this new semiring as **SP semiring**.

Let \(\hat{p}_{\rightarrow i} \) be a BP fixed point — that is

\[
\hat{p}_{\rightarrow i} = \{ \forall i, I \in \partial i \hat{p}_{i \rightarrow i} = P_{i \rightarrow i}(\hat{p}_{\partial i \setminus i \rightarrow i}), \hat{p}_{i \rightarrow i} = P_{i \rightarrow i}(\hat{p}_{\partial i \setminus i \rightarrow i}) \}
\]

and denote the set of all such fixed points by \(P\). Each BP fixed point corresponds to an approximation to the \(\mathbb{Q}(\emptyset)\), which we denote by \(Q(\hat{p}_{\rightarrow \emptyset})(\emptyset)\) — using this functional form
is to emphasize the dependence of this approximation on BP messages. Recall that in the original problem, \( \mathcal{X} \) is the domain of assignments, \( q(x) \) is the expanded form and \( \oplus \)-marginalization is (approximately) performed by BP. In the case of survey propagation, \( \mathcal{P} \) is domain of assignments and the integral \( Q(\hat{\mathcal{P}}_{\mathcal{P}})(\emptyset) \) evaluates a particular assignment \( \hat{\mathcal{P}}_{\mathcal{P}} \) to all the messages – i.e., \( Q(\hat{\mathcal{P}}_{\mathcal{P}})(\emptyset) \) is the new expanded form.

In this algebraic perspective, SP efficiently performs a second integral using \( \hat{\oplus} \) over all fixed points:

\[
Q(\emptyset)(\emptyset) = \hat{\oplus}_{\hat{\mathcal{P}}_{\mathcal{P}}} \in \mathcal{P} Q(q_{\mathcal{P}})(\emptyset)
\]  

(23)

Table I summarizes this correspondence.

Our derivation requires \((\mathcal{Y}^*, \otimes)\) to be an abelian group (i.e., every element of \( \mathcal{Y}^* \) has an inverse w.r.t. \( \otimes \)). The requirement for invertability of \( \otimes \) is because we need to work with normalized BP and SP messages. In section 6.4 we introduce another variation of SP that simply counts the BP fixed points and relaxes this requirement.

### 6.1. Decomposition of the integral

In writing the normalized BP equations in section 5, we hid the normalization constant using \( \otimes \) sign. Here we explicitly define the normalization constants or **local integrals** by defining unnormalized messages, based on their normalized version

\[
\tilde{p}_{1\rightarrow i}(x_i) \overset{\text{def}}{=} \bigotimes_{x_i} f_i(x_i) \otimes \tilde{p}_{j\rightarrow i}(x_j) \quad \text{def} = \tilde{p}_{1\rightarrow i}(\tilde{p}_{\mathcal{P}\mathcal{I}\cup\mathcal{I}}(x_i))
\]  

(24)

\[
\tilde{p}_{i\rightarrow 1}(x_i) \overset{\text{def}}{=} \bigotimes_{i \in \partial i} \tilde{p}_{i\rightarrow 1}(x_i) \quad \text{def} = \tilde{p}_{i\rightarrow 1}(\tilde{p}_{\mathcal{P}\mathcal{I}\cup\mathcal{I}}(x_i))
\]  

(25)

\[
\tilde{p}_i(x_i) \overset{\text{def}}{=} \bigotimes_{i \in \partial i} \tilde{p}_{i\rightarrow 1}(x_i) \quad \text{def} = \tilde{p}_i(\tilde{p}_{\mathcal{P}\mathcal{I}\cup\mathcal{I}}(x_i))
\]  

(26)

\[
\tilde{p}_i(x_i) \overset{\text{def}}{=} \bigotimes_{i \in \partial i} \tilde{p}_{i\rightarrow 1}(x_i)
\]  

(27)

where each update also has a functional form on the r.h.s. In each case, the local integrals are simply the integral of unnormalized messages or marginals – e.g., \( \tilde{p}_{i\rightarrow 1}(\emptyset) = \bigoplus_{x_i} \tilde{p}_{i\rightarrow 1}(x_i) \).
Define the functional \( \tilde{P}_{t+1}(\tilde{p}_{t-1}, \tilde{p}_{t+1}) \) as the product of messages from \( i \) to \( I \) and vice versa

\[
\tilde{P}_{t+1}(x_i) \overset{\text{def}}{=} \tilde{p}_{t-1}(x_i) \otimes \tilde{p}_{t+1}(x_i) \overset{\text{def}}{=} \tilde{P}_{t+1}(\tilde{p}_{t-1}, \tilde{p}_{t+1})(x_i)
\]  

(28)

**Theorem 6.1.** If the factor-graph has no loops and \((\mathcal{Y}^*, \otimes)\) is an abelian group, the global integral decomposes to local BP integrals as

\[
q(\theta) = \bigotimes_i \tilde{p}_i(\theta) \bigotimes_i \tilde{p}_i J(\theta) \left( \bigotimes_{i, j \in \partial i} \tilde{P}_{t+1}(\theta) \right)^{-1}
\]  

or in other words \( q(\theta) = Q(\tilde{p}_{\rightarrow \theta})(\theta) \) where

\[
Q(\tilde{p}_{\rightarrow \theta})(\theta) = \bigotimes_i \tilde{P}_i(\theta) \bigotimes_i \tilde{P}_j(\theta) \left( \bigotimes_{i, j \in \partial i} \tilde{P}_{t+1}(\theta) \right)^{-1}
\]  

(29)

(30)

6.2. The new factor-graph and semiring

The decomposition of integral in theorem 6.1 means \( Q(\tilde{p}_{\rightarrow \theta})(\theta) \) has a factored form. Therefore, a factor-graph with \( \tilde{p} \) as the set of variables and three different types of factors corresponding to different terms in the decomposition – i.e., \( \tilde{P}_i(\tilde{p}_{\rightarrow \theta})(\theta) \), \( \tilde{P}_i(\tilde{p}_{j \rightarrow \theta})(\theta) \) and \( \tilde{P}_{t+1}(\tilde{p}_{t-1}, \tilde{p}_{t+1})(\theta)^{-1} \) can represent \( Q(\tilde{p}_{\rightarrow \theta})(\theta) \).

Figure 3 shows a simple factor-graph and the corresponding SP factor-graph. The new factor-graph has one variable per each message in the original factor-graph and three types of factors as discussed above. Survey propagation is simply belief propagation applied to this new factor-graph using the new semiring. As before BP messages are exchanged between variables and factors. But here, we can simplify BP messages by substitution and only keep two types of factor-to-factor messages. We use \( S_{i \rightarrow t} \) and \( S_{j \rightarrow t} \) to denote these two types of SP messages. These messages are exchanged between two types of factors, namely \( \tilde{P}_i(\tilde{p}_{j \rightarrow \theta})(\theta) \) and \( \tilde{P}_i(\tilde{p}_{j \rightarrow \theta})(\theta) \). Since the third type of factors \( \tilde{P}_{t+1}(\tilde{p}_{t-1}, \tilde{p}_{t+1})(\theta)^{-1} \) are always connected to only two variables, \( \tilde{p}_{t-1} \) and \( \tilde{p}_{t+1} \), we can simplify their role in SP message update to get

\[
S_{i \rightarrow t}(\tilde{p}_{t-1}, \tilde{p}_{t+1}) \propto \bigoplus_{\tilde{p}_{t-1}, \tilde{p}_{t+1}} \left( \frac{\tilde{P}_i(\tilde{p}_{j \rightarrow \theta})(\theta)}{\tilde{P}_{t+1}(\tilde{p}_{t-1}, \tilde{p}_{t+1})(\theta)} \right) \otimes J_{i \in \partial i} S_{j \rightarrow t}(\tilde{p}_{j-1}, \tilde{p}_{j+1})
\]  

(31)

\[
S_{j \rightarrow t}(\tilde{p}_{t-1}, \tilde{p}_{t+1}) \propto \bigoplus_{\tilde{p}_{t-1}, \tilde{p}_{t+1}} \left( \frac{\tilde{P}_j(\tilde{p}_{j \rightarrow \theta})(\theta)}{\tilde{P}_{t+1}(\tilde{p}_{t-1}, \tilde{p}_{t+1})(\theta)} \right) \otimes J_{j \in \partial j} S_{j \rightarrow t}(\tilde{p}_{j-1}, \tilde{p}_{j+1})
\]  

(32)

where in all cases we are assuming the messages \( \tilde{p}_{\rightarrow \theta} \in \mathcal{P} \) are consistent with each other – i.e., satisfy BP equations on the original factor-graph. Note that, here again we are using the normalized BP message update and the normalization factor is hidden using \( \times \) sign. This is possible because we assumed \( \otimes \) has an inverse. We can further simplify this update using the following proposition.
Revisiting Algebra & Complexity of Inference in Graphical Models

Fig. 3. Part of a factor-graph (left) and the corresponding SP factor-graph on the right. The variables in SP factor-graph are the messages in the original graph. The SP factor-graph has three type of factors: (I) $\tilde{\mathcal{P}}_I(\cdot)(\emptyset)$, (II) $\tilde{\mathcal{P}}_I(\cdot)(\emptyset)$ and (III) $\tilde{\mathcal{P}}_I(\cdot)(\emptyset)^{-1}$. As the arrows suggest, SP message updates are simplified so that only two type of messages are exchanged: $S_{i\to I}$ and $S_{I\to i}$ between factors of type (I) and (II).

**Proposition 6.2.** For $\tilde{\mathcal{P}}_{\cdot \to \cdot} \in \mathcal{P}$

$$\frac{\tilde{\mathcal{P}}_{i}(\tilde{\mathcal{P}}_{\partial_i \to i})(\emptyset)}{P_{i\to i}(\tilde{\mathcal{P}}_{\partial_i \to i}, \tilde{\mathcal{P}}_{i\to i})(\emptyset)} = \tilde{\mathcal{P}}_{i\to i}(\tilde{\mathcal{P}}_{\partial_i \to i, i\to i})(\emptyset)$$

and

$$\frac{\tilde{\mathcal{P}}_{i}(\tilde{\mathcal{P}}_{\partial_i \to i})(\emptyset)}{P_{i\to i}(\tilde{\mathcal{P}}_{\partial_i \to i, i\to i})(\emptyset)} = \tilde{\mathcal{P}}_{i\to i}(\tilde{\mathcal{P}}_{\partial_i \to i, i\to i})(\emptyset)$$

The term on the l.h.s. in the proposition above appear in eqs (31) and (32) and the terms on the r.h.s are local message integrals given by eqs (24) and (25). We can enforce $\tilde{\mathcal{P}}_{\cdot \to \cdot} \in \mathcal{P}$, by enforcing BP updates $\tilde{\mathcal{P}}_{i\to i} = P_{i\to i}(\tilde{\mathcal{P}}_{\partial_i \to i, i\to i})$ and $\tilde{\mathcal{P}}_{i\to i} = P_{i\to i}(\tilde{\mathcal{P}}_{\partial_i \to i, i\to i})$ “locally”, during the message updates in the new factor-graph. Combining this constraint with the simplification offered by proposition 6.2 gives us the SP message updates

$$S_{i\to i}(\tilde{\mathcal{P}}_{\cdot\to\cdot}) \propto \bigoplus_{\tilde{\mathcal{P}}_{\partial_i \to i, i\to i}} \left(1(\tilde{\mathcal{P}}_{i\to i} = P_{i\to i}(\tilde{\mathcal{P}}_{\partial_i \to i, i\to i})) \otimes P_{i\to i}(\tilde{\mathcal{P}}_{\partial_i \to i, i\to i})(\emptyset) \bigotimes_{j \in \partial_i \setminus i} S_{j\to i}(\tilde{\mathcal{P}}_{\cdot\to\cdot}) \right)$$

$$S_{I\to i}(\tilde{\mathcal{P}}_{\cdot\to\cdot}) \propto \bigoplus_{\tilde{\mathcal{P}}_{\partial_i \to i, i\to i}} \left(1(\tilde{\mathcal{P}}_{i\to i} = P_{i\to i}(\tilde{\mathcal{P}}_{\partial_i \to i, i\to i})) \otimes P_{i\to i}(\tilde{\mathcal{P}}_{\partial_i \to i, i\to i})(\emptyset) \bigotimes_{j \in \partial_i \setminus i} S_{j\to i}(\tilde{\mathcal{P}}_{\cdot\to\cdot}) \right)$$

where $1(.)$ is the identity function on the SP semiring, where $1(\text{TRUE}) = \top$ and $1(\text{FALSE}) = \bot$.

Here each SP message is a functional over all possible BP messages between the same variable and factor. However, in updating the SP messages, the identity functions ensure that only the messages that locally satisfy BP equations are taken into account. Another difference from the updates of eqs (31) and (32) is that SP messages have a single argument. This is because the new local integrals either depend on $\tilde{\mathcal{P}}_{i\to i}$ or $\tilde{\mathcal{P}}_{i\to i}$, and not both.
Example 6.3. In variational approach, survey propagation comes in two variations: entropic $SP(x)$ and energetic $SP(y)$ [Mézard and Montanari 2009]. For the readers familiar with variational derivation of SP, here we express the relation to the algebraic approach. The entropic SP has an inverse temperature parameter, a.k.a. Parisi parameter, $x \in \mathbb{R}$ and $\xi = 1$ corresponds to $\oplus = +$, $\ominus = +$ and $\otimes = x$ in our algebraic approach. The limits of $\xi \rightarrow \infty$ corresponds to $\oplus = \max$. On the other hand, the limit of $\xi \rightarrow 0$, which is often used to solve CSPs corresponds to counting BP; see section 6.4. The energetic $SP(x)$ corresponds to $\oplus = +$, $\ominus = \min$ and $\otimes = \min$, and the limits of the inverse temperature parameter $y \rightarrow \infty$ is equivalent to $\ominus = \min$, $\ominus = \min$ and $\otimes = \sum$. By taking an algebraic view we can choose between both operations and domains. For instance, an implication of algebraic view is that all the variations of SP can be applied to the domain of complex numbers $\mathbb{Y}^* = \mathbb{C}$.

6.3. The new integral and marginals

Once again we can use theorem 6.1, this time to approximate the $SP$ integral $Q(\emptyset)(\emptyset) = \bigoplus_{(\mathbb{P}, \rightarrow)} Q(\mathbb{P}_{(\rightarrow)})$ using local integral of SP messages.

The $SP$ marginal over each BP message $\mathbb{P}_{j \rightarrow i}$ or $\mathbb{P}_{i \rightarrow j}$ is the same as the corresponding SP message – i.e., $S(\mathbb{P}_{j \rightarrow i}) = S_{j \rightarrow i}(\mathbb{P}_{j \rightarrow i})$. To see this in the factor-graph of figure 3, note that each message variable is connected to two factors, and both of these factors are already contained in calculating one SP messages.

Moreover, from the SP marginals over messages we can recover the SP marginals over BP marginals which we denote by $S(\mathbb{P}(x_i))$. For this, we simply need to enumerate all combinations of BP messages that produce a particular marginal

$$S(\mathbb{P})(x_i) \propto \bigoplus_{\mathbb{P}_{j \rightarrow i}} 1(\mathbb{P}(x_i) = \mathbb{P}(\mathbb{P}_{j \rightarrow i}))(x_i) \bigotimes_{I \in \partial i} S_{I \rightarrow i}(\mathbb{P}_{I \rightarrow i}) \tag{37}$$

6.4. counting survey propagation

Previously we required the $\otimes$ operator to have an inverse, so that we can decompose the BP integral $q(\emptyset)$ into local integrals. Moreover, for a consistent decomposition of the BP integral, SP and BP semiring previously shared the $\otimes$ operation.\(^5\)

Here, we lift these requirements by discarding the BP integrals altogether. This means SP semiring could be completely distinct from BP semiring and $\mathbb{Y}^*$, $\otimes$ does not have to be an abelian group. This setting is particularly interesting when the SP semiring is sum-product over real domain

$$S_{j \rightarrow i}(\mathbb{P}_{j \rightarrow i}) \propto \sum_{\mathbb{P}_{j \rightarrow i}} 1(\mathbb{P}_{j \rightarrow i} = \mathbb{P}_{j \rightarrow i}(\mathbb{P}_{j \rightarrow i})) \prod_{j \in \partial j \setminus i} S_{j \rightarrow i}(\mathbb{P}_{j \rightarrow i})$$

$$S_{i \rightarrow j}(\mathbb{P}_{i \rightarrow j}) \propto \sum_{\mathbb{P}_{i \rightarrow j}} 1(\mathbb{P}_{i \rightarrow j} = \mathbb{P}_{i \rightarrow j}(\mathbb{P}_{j \rightarrow i})) \prod_{j \in \partial j \setminus i} S_{j \rightarrow i}(\mathbb{P}_{j \rightarrow i})$$

Here, the resulting SP integral $Q(\mathbb{P}_{(\rightarrow)}) = \sum_{\mathbb{P}_{(\rightarrow)}} 1(\mathbb{P}_{(\rightarrow)} \in \mathbb{P})$ simply “counts” the number of BP fixed points and SP marginals over BP marginals (given by eq (37)).

\(^5\)This is because if the expansion operation $\otimes$ was different from the expansion operation of BP, $\otimes$, the expanded form $Q(\mathbb{P}_{(\rightarrow)})$ in the SP factor-graph would not evaluate the integral $q(\emptyset)$ in the BP factor-graph, even in factor-graphs without any loops.
approximates the frequency of a particular marginal. The original survey propagation equations in [Braunstein et al. 2002a], that are very successful in solving satisfiability correspond to counting SP.

Example 6.4. Interestingly, in all min-max problems with discrete domains $\mathcal{X}$, min-max BP messages can only take the values that are in the range of factors — i.e., $\mathcal{Y}^* = \mathcal{Y}$. This is because any ordered set is closed under min and max operations. Here, each uniform SP message $S_{i \rightarrow I}(\hat{p}_{i \rightarrow I}) : \mathcal{Y}^{|X_i|} \rightarrow \mathbb{R}$ is a discrete distribution over all possible min-max BP messages. This means counting survey propagation where the BP semiring is min-max is computationally “tractable”. In contrast (counting) SP, when applied to sum-product BP over real domains is not tractable. This is because in this case each SP message is a distribution over an uncountable set: $S_{i \rightarrow I}(\hat{p}_{i \rightarrow I}) : \mathbb{R}^{|X_i|} \rightarrow \mathbb{R}$.

In practice, (uniform) SP is only interesting if it remains tractable. The most well-known case corresponds to uniform SP when applied to or-and semiring. In this case the factors are constraints and the domain of SP messages is $\{\text{TRUE, FALSE}\}^{|X_i|}$. Our algebraic perspective extends this set of tractable instances. For example, it show that counting SP can be applied to xor-and or min-max semiring.

Conclusion
This paper builds on previous work to addresses three basic questions about inference in graphical models: (I) “what is an inference problem in a graphical model?” We use the combination of commutative semigroups and a factor-graph to answer this question in a broad sense, generalizing variety of previous models, while keeping the model free of semantics. (II) “How difficult is inference?” By confining inference to four operations of min, max, sum and product that easily lend themselves to models of computation, we build an inference hierarchy that is complete for PSPACE and organizes inference problems into complexity classes with increasing level of difficulty. Only a few of these problems are previously studied and only a handful of them have a variational interpretation to this day. Moreover, we prove that inference for “any” commutative semiring is NP-hard under randomized reduction, which generalizes previous results for particular semirings. (III) “When does distributive law help?” After reviewing the algebraic form of belief propagation, and the conditions that allow its normalized form, we show that application of distributive law in performing exact inference is limited to inference problems with one marginalization operation. Finally we propose an algebraic interpretation of survey propagation that extends its theoretical application to new settings.

Proofs
PROOF. (claim 3.2) We show this for an integration problem. Since the integral can be obtained from a polynomial-sized marginal, the exponential representation of the integral implies the exponential representation of the marginal. For example if $q(x_i)$ has a polynomial representation in $N$, then the same should hold for $q(\emptyset) = \bigoplus_{x_i} q(x_i)$.

To see why this integral has an exponential representation in $N$, consider its simplified form

$$q(\emptyset) = \prod_{x_i} q(\underline{x}_i)$$

where $q(\underline{x}_i)$ is the result of inference up to the last marginalization step, which is product and $X_1$ grows exponentially with $N$. Since $q(\underline{x}_i)$ for each $\underline{x}_i \in X_1$ has a constant size,
say \( c \), the size of \( q(\emptyset) \) is

\[
\log(q(\emptyset)) = \sum_{z \in \mathcal{X}} c = c|\mathcal{X}| \]

which is exponential in \( N \). The exponential representation of the query also means the corresponding computation is outside \( \text{PSPACE} \).

**Proof.** (proposition 3.3) To show that these inference problems are in \( \mathbb{P} \), we provide polynomial-time algorithms for them:

- **sum − sum** is short for

  \[
  q(\emptyset) = \sum_{z} \sum_{I} f_{I}(x_{I})
  \]

  which asks for the sum over all assignments of \( x \in \mathcal{X} \), of the sum of all the factors. It is easy to see that each factor value \( f_{I}(x_{I}) \) for all \( I \) is counted \( |\mathcal{X}_{I}| \) times in the summation above. Therefore we can rewrite the integral above as

  \[
  q(\emptyset) = \sum_{I} |\mathcal{X}_{I}| \left( \sum_{z} f_{I}(x_{I}) \right)
  \]

  where the new form involves polynomial number of terms and therefore is easy to calculate.

- **min − min** (similar for max − max) is short for

  \[
  q(\emptyset) = \min_{z} \min_{I} f_{I}(x_{I})
  \]

  where the query seeks the minimum achievable value of any factor. We can easily obtain this by seeking the range of all factors and reporting the minimum value in polynomial time.

**Proof.** (proposition 3.4) Given \( x \) it is easy to verify the decision problem so, **max-min-decision** belongs to \( \text{NP} \). To show \( \text{NP} \)-completeness, we reduce the 3-SAT to a max-min inference problem, such that 3-SAT is satisfiable iff the max-min value is \( q(\emptyset) \geq 1 \) and unsatisfiable otherwise.

Simply define one factor per clause of 3-SAT, such that \( f_{I}(x_{I}) = 1 \) if \( x_{I} \) satisfies the clause and any number less than one otherwise. With this construction, the max-min value \( \max_{x} \min_{I} f_{I}(x_{I}) \) is one iff the original SAT problem was satisfiable, otherwise it is less than one. This reduces 3-SAT to **Max-Min-decision**.

**Proof.** (proposition 3.5) Recall that \( \text{PP} \) is the class of problems that are polynomially solvable using a non-deterministic Turing machine, where the acceptance condition is that the majority of paths accept. To see that \( \sum_{z} \min_{I} f_{I}(x_{I}) \geq q \) is in \( \text{PP} \), enumerate all \( x \in \mathcal{X} \) non-deterministically and for each assignment calculate \( \min_{I} f_{I}(x_{I}) \) in polynomial time (where each path accepts iff \( \min_{I} f_{I}(x_{I}) = 1 \)) and accept iff \( q \) of the paths accept.

Given a matrix \( A \in \{0,1\}^{N \times N} \) the problem of calculating its permanent

\[
\text{perm}(A) = \sum_{z \in S_{N}} \prod_{i=1}^{N} A_{i,z}
\]

where \( S_{N} \) is the set of permutations of \( 1, \ldots, N \) is \( \#P \)-complete and the corresponding decision problem is \( \#P \)-complete [Valiant 1979]. To show completeness wrt \( \text{PP} \) it is
enough to reduce the problem of computing the matrix permanent to sum-min inference in a graphical model.

The problem of computing the permanent has been reduced to sum-product inference in graphical models [Huang and Jebara 2009]. However, when \( f_I(\mathbf{x}_I) \in \{0, 1\} \forall I \), sum-product is isomorphic to sum-min. This is because \( y_1 \times y_2 = \min(y_1, y_2) \forall y_i \in \{0, 1\} \).

Therefore, the problem of computing the permanent for such matrices reduces to sum-min inference in the factor-graph of [Huang and Jebara 2009].

\[ \text{PROOF. (theorem 3.7)} \]

To prove that a problem is \( \text{PSPACE-complete} \), we have to show that 1) it is \( \text{PSPACE} \) and 2) a \( \text{PSPACE}- \)complete problem reduces to it. We already saw in section 3.3 that QSAT, which is \( \text{PSPACE}- \)complete, reduces to the inference hierarchy. But it is not difficult to show that inference hierarchy is contained in \( \text{PSPACE} \).

Let \( q(\mathbf{x}_{I_0}) = \bigoplus_{\mathbf{x}_{I_M}} \bigoplus_{\mathbf{x}_{I_{M-1}}} \ldots \bigoplus_{\mathbf{x}_{I_2}} \bigotimes_{\mathbf{x}_{I_1}} f_I(\mathbf{x}_I) \)
be any inference problem in the hierarchy. We can simply iterate over all values of \( \mathbf{x} \in \mathcal{X} \) in nested loops or using a recursion. Let \( j(i) : \{1, \ldots, N\} \rightarrow \{1, \ldots, M\} \) be the index of the marginalization that involves \( x_i \) – that is \( i \in J_{j(i)} \). Moreover let \( i_1, \ldots, i_N \) be an ordering of variable indices such that \( j(i_k) \leq j(i_{k+1}) \). Algorithm 1 uses this notation to demonstrate this procedure using nested loops. Note that here we loop over individual domains \( \mathcal{X}_{i_k} \) rather than \( \mathcal{X}_{I_m} \) and track only temporary tuples \( q_{i_k} \), so that the space complexity remains polynomial in \( N \).

\[ \text{PROOF. (theorem 4.5)} \]

To prove that inference in any semiring \( S = (\mathcal{Y}^*, \oplus, \otimes) \) is \( \text{NP-hard} \) under randomized polynomial reduction, we reduce “unique satisfiability”(USAT) to an inference problems on any semiring. USAT is a so-called “promise problem”, that asks whether a satisfiability problem that is promised to have either zero or one satisfying assignment is satisfiable. Valiant and Vazirani [1986] prove that a polynomial time randomized algorithm (R\#P) for USAT implies a \( \text{R\#P=NP} \).

For this reduction consider a set of binary variables \( \mathbf{x} \in \{0, 1\}^N \), one per each variable in the given instance of USAT. For each clause, define a constraint factor \( f_I \) such that \( f_I(\mathbf{x}_I) = \otimes \) if \( \mathbf{x}_I \) satisfies that clause and \( f_I(\mathbf{x}_I) = 1 \) otherwise. This means, \( \mathbf{x} \) is a satisfying assignment for USAT iff \( q(\mathbf{x}) = \bigotimes_{I} f_I(\mathbf{x}_I) = \otimes \). If the instance is unsatisfiable, the integral \( q(\emptyset) = \bigoplus_{\mathbf{x}} \otimes = \otimes \) (by definition of \( \otimes \)). If the instance is satisfiable there is only a single instance \( \mathbf{x}^* \) for which \( q(\mathbf{x}^*) = \otimes \), and therefore the integral evaluates to \( \otimes \). Therefore we can decide the satisfiability of USAT by performing inference on any semiring, by only relying on the properties of identities. The satisfying assignment can be recovered using a decimation procedure, assuming access to an oracle for inference on the semiring.

\[ \text{PROOF. (theorem 6.1)} \]

For this proof we build a tree around an root node \( r \) that is connected to one factor. (Since the factor-graph is a tree such a node always exists.)
**ALGORITHM 1:** inference in PSPACE

**Input:** \( \bigoplus_{x_i} M_{i_1} \bigoplus_{x_{i-1}} \cdots \bigoplus_{x_1} f_i(x_i) \)

**Output:** \( q(x_{j_0}) \)

**For each** \( x_{j_0} \in X_{j_0} \) do // loop over the query domain

**For each** \( x_i \in X_i \) do // loop over \( X_i \)

**For each** \( z_{i_1} \in X_{i_1} \) do // loop over \( X_{i_1} \)

\[
q_{i_1}(z_{i_1}) := \bigotimes_{f_i(z_{i_1})};
\]

**End**

\[
q_{i_2}(z_{i_2}) := \bigoplus_{x_{i_1}} q_{i_1}(x_{i_1})
\]

\[
q_{i_3}(z_{i_3}) := \bigoplus_{x_{i_2}} q_{i_2}(x_{i_2})
\]

\[
q_{i_N}(z_{i_N}) := \bigoplus_{x_{i_{N-1}}} q_{i_{N-1}}(x_{i_{N-1}})
\]

**End**

\[
q(z_{j_0}) := \bigoplus_{x_{i_N}} q_N(x_{i_N})
\]

Send BP messages from the leaves, up towards the root \( r \) and back to the leaves. Here, any message \( q_{i_1}(x_i) \), can give us the integral for the sub-tree that contains all the nodes and factors up to node \( i \) using \( q_{i_1}(\emptyset) = \bigoplus_{x_i} q_{i_1}(x_i) \). Noting that the root is connected to exactly one factor, the global integral is

\[
\bigoplus_{x_r} q(x_r) = \bigoplus_{x_r} \bigotimes_{I \in \partial r} q_{i_1}(x_r) = q_{i_1}(\emptyset) \tag{38}
\]

On the other hand, We have the following relation between \( q_{i_1} \) and \( \hat{p}_{i_1} \) (also corresponding factor-to-variable message)

\[
q_{i_1}(x_i) = \hat{p}_{i_1}(x_i) \otimes q_{i_1}(\emptyset) \quad \forall i, I \in \partial i \tag{39}
\]

\[
q_{i_1}(x_i) = \hat{p}_{i_1}(x_i) \otimes q_{i_1}(\emptyset) \quad \forall i, I \in \partial i \tag{40}
\]

Substituting this into BP eqs (14) and (15) we get

\[
q_{i_1}(x_i) = \bigotimes_{J \in \partial i \setminus I} q_{i_1}(\emptyset) \hat{p}_{j_{i_1}}(x_i) \tag{41}
\]

\[
q_{i_1}(x_i) = \bigotimes_{J \in \partial i \setminus I} f_i(x_i) \bigotimes q_{j_{i_1}}(\emptyset) \hat{p}_{j_{i_1}}(x_j) \tag{42}
\]
By summing over both l.h.s and r.h.s in equations above and substituting from eq (25) we get

\[
\bigoplus_{x_i} q_{i\rightarrow l}(x_i) = \left( \bigotimes_{j \in \partial l \setminus l} q_{j\rightarrow i}(\emptyset) \right) \otimes \left( \bigoplus_{x_i} \prod_{j \in \partial l \setminus l} \tilde{p}_{j\rightarrow i}(x_i) \right) \rightarrow
q_{i\rightarrow l}(\emptyset) = \tilde{p}_{i\rightarrow l}(\emptyset) \otimes q_{j\rightarrow i}(\emptyset)
\] (43)

and similarly for eq (42) using integration and substitution from eq (24) we have

\[
\bigoplus_{x_i} q_{l\rightarrow i}(x_i) = \left( \bigotimes_{j \in \partial i \setminus i} q_{j\rightarrow l}(\emptyset) \right) \otimes \left( \bigoplus_{x_i} f_i(x_i) \bigotimes_{j \in \partial i \setminus i} \tilde{p}_{j\rightarrow l}(x_j) \right) \rightarrow
q_{l\rightarrow i}(\emptyset) = \tilde{p}_{l\rightarrow i}(\emptyset) \bigotimes_{j \in \partial i \setminus i} q_{j\rightarrow l}(\emptyset)
\] (44)

Eqs (43) and (44) are simply recursive integration on a tree, where the integral up to node \( i \) (i.e., \( q_{i\rightarrow l}(\emptyset) \) in eq (43)) is reduced to integral in its sub-trees. By unrolling this recursion we see that \( q_{i\rightarrow l}(\emptyset) \) is simply the product of all \( \tilde{p}_{i\rightarrow l}(\emptyset) \) and \( \tilde{p}_{j\rightarrow i}(\emptyset) \) in its sub-tree, where the messages are towards the root. Eq (38) tells us that the global integral is not different. Therefore, eqs (43) and (44) we can completely expand the recursion for the global integral. For this, let \( \uparrow i \) restrict the \( \partial i \) to the factor that is higher than variable \( i \) in the tree (i.e., closer to the root \( r \)). Similarly let \( \uparrow 1 \) be the variable that is closer to the root than \( i \). We can write the global integral as

\[
q(\emptyset) = \bigotimes_{i, i = \uparrow i} \tilde{p}_{i\rightarrow l}(\emptyset) \bigotimes_{l, i = \uparrow l} \tilde{p}_{l\rightarrow i}(\emptyset)
\] (45)

Proposition 6.2 shows that these local integrals can be written in terms of local integrals of interest – i.e.,

\[
\tilde{p}_{i\rightarrow l}(\emptyset) = \frac{\tilde{p}_{l}(\emptyset)}{\tilde{p}_{l\rightarrow l}(\emptyset)} \quad \text{and} \quad \tilde{p}_{l\rightarrow i}(\emptyset) = \frac{\tilde{p}_{i}(\emptyset)}{\tilde{p}_{i\rightarrow i}(\emptyset)}
\]

Substituting from the equations above into eq (45) we get the equations of theorem 6.1.

PROOF. (proposition 6.2) By definition of \( \tilde{p}_{l}(x_i) \) and \( \tilde{p}_{l\rightarrow i}(x_i) \) in eqs (24) and (26)

\[
\tilde{p}_{l}(x_i) = \tilde{p}_{l\rightarrow i}(x_i) \otimes \tilde{p}_{l\rightarrow i}(x_i) \rightarrow \bigoplus_{x_i} \tilde{p}_{l}(x_i) = \bigoplus_{x_i} \tilde{p}_{l\rightarrow i}(x_i) \otimes \tilde{p}_{l\rightarrow i}(x_i) \rightarrow
\tilde{p}_{l}(\emptyset) = \tilde{p}_{l\rightarrow i}(\emptyset) \bigotimes_{x_i} \tilde{p}_{l\rightarrow i}(x_i) \rightarrow \tilde{p}_{i}(\emptyset) = \tilde{p}_{i\rightarrow l}(\emptyset) \bigotimes_{x_i} \tilde{p}_{i\rightarrow l}(x_i)
\]

where in the last step we used eq (28).

Similarly for the second statement of the proposition we have

\[
\tilde{p}_{i}(x_i) = \tilde{p}_{i\rightarrow l}(x_i) \otimes \tilde{p}_{i\rightarrow l}(x_i) \rightarrow \bigoplus_{x_i} \tilde{p}_{i}(x_i) = \bigoplus_{x_i} \tilde{p}_{i\rightarrow l}(x_i) \otimes \tilde{p}_{i\rightarrow l}(x_i) \rightarrow
\tilde{p}_{l}(\emptyset) = \tilde{p}_{i\rightarrow l}(\emptyset) \bigotimes_{x_i} \tilde{p}_{i\rightarrow l}(x_i) \rightarrow \tilde{p}_{i}(\emptyset) = \tilde{p}_{i\rightarrow l}(\emptyset) \bigotimes_{x_i} \tilde{p}_{i\rightarrow l}(x_i)
\]

\( \square \)
REFERENCES
S Aji and R McEliece. 2001. The Generalized distributive law and free energy minimization. In Allerton Conf.
Srinivas M Aji, Gavin B Horn, and Robert J McEliece. 1998. On the Convergence of iterative decoding on graphs with a single cycle. In Proc 1998 IEEE Int Symp Information Theory.
S M Aji and R J McEliece. 2000. The generalized distributive law. Information Theory, IEEE Transactions on 46, 2 (2000), 325–343.
Fahiem Bacchus. 1991. Representing and reasoning with probabilistic knowledge: a logical approach to probabilities. MIT press.
Mohsen Bayati, Devavrat Shah, and Mayank Sharma. 2005. Maximum weight matching via max-product belief propagation. In ISIT. IEEE, 1763–1767.
Stefano Bistarelli, Ugo Montanari, Francesca Rossi, Thomas Schiex, Gérard Verfaillie, and Helene Fargier. 1999. Semiring-based CSPs and valued CSPs: Frameworks, properties, and comparison. Constraints 4, 3 (1999), 199–240.
Lucas Bordeaux and Eric Monfroy. 2002. Beyond NP: Arc-consistency for quantified constraints. In Principles and Practice of Constraint Programming-CP 2002. Springer, 371–386.
A Braunstein, M Mezard, M Weigt, and R Zecchina. 2002b. Constraint Satisfaction by Survey Propagation. Physics 2 (2002), 8.
A Braunstein, M Mezard, and R Zecchina. 2002a. Survey propagation: an algorithm for satisfiability. Random Structures and Algorithms 27, 2 (2002), 19.
A. Braunstein and R. Zecchina. 2003. Survey Propagation as local equilibrium equations. ArXiv:cond-mat/0312485 (Dec. 2003). doi: http://dx.doi.org/10.1088/1742-5468/2004/06/P06007 J. Stat. Mech., P06007 (2004).
Peter Clifford. 1990. Markov random fields in statistics. Disorder in physical systems (1990), 19–32.
G Cooper. 1990. The computational complexity of probabilistic inference using Bayesian belief networks. Artificial Intelligence 42 (1990).
Brian A Davey and Hilary A Priestley. 2002. Introduction to lattices and order. Cambridge university press.
Rina Dechter and David Larkin. 2001. Hybrid processing of beliefs and constraints. In Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 112–119.
Beno Eckmann and Peter J Hilton. 1962. Group-like structures in general categories I multiplications and comultiplications. Math. Ann. 145, 3 (1962), 227–255.
G. David Forney Jr. 2001. Codes on graphs: normal realizations. Information Theory, IEEE Transactions on 47, 2 (2001), 520–548.
Y Fu and P W Anderson. 1986. Application of statistical mechanics to NP-complete problems in combinatorial optimisation. Journal of Physics A: Mathematical and General 19, 9 (June 1986), 1605–1620. DOI: http://dx.doi.org/10.1088/0305-4470/19/9/033
Robert G Gallager. 1962. Low-density parity-check codes. Information Theory, IRE Transactions on 8, 1 (1962), 21–28.
Ronald A Howard and James E Matheson. 2005. Influence diagrams. Decision Analysis 2, 3 (2005), 127–143.
Bert Huang and Tony Jebara. 2009. Approximating the permanent with belief propagation. ArXiv preprint arXiv:0908.1769 (2009).
Morteza Ibrahimi, Adel Javanmard, Yashodhan Kanoria, and Andrea Montanari. 2011. Robust max-product belief propagation. In Proceedings of the Forty Fifth Asilomar Conference on. IEEE, 43–49.
Edwin T Jaynes. 2003. Probability theory: the logic of science. Cambridge university press.
Michael Kearns. 2007. Graphical games. Algorithmic game theory 3 (2007), 159–180.
Lukas Kroc, A Sabharwal, and Bart Selman. 2002. Survey Propagation Revisited. 23rd UAI (2002), 217–226.
F R Kschischang and B J Frey. 2001. Factor graphs and the sum-product algorithm. Information Theory, IEEE (2001).
Steffen L Lauritzen and Finn Verner Jensen. 1997. Local computation with valuations from a commutative semigroup. Annals of Mathematics and Artificial Intelligence 21, 1 (1997), 51–69.
Michael L Littman, Stephen M Majercik, and Toniann Pitassi. 2001. Stochastic boolean satisfiability. Journal of Automated Reasoning 27, 3 (2001), 251–296.
E N. Maneva, E Mossel, and M J Wainwright. 2004. A New Look at Survey Propagation and its Generalizations. ArXiv:cs/0409012 (Sept. 2004).
M. Mézard and A. Montanari. 2009. *Information, physics, and computation*. Oxford.
Marc Mézard, Georgio Parisi, and M. A. Virasoro. 1987. *Spin Glass Theory and Beyond*. Singapore: World Scientific.
M. M´ezard, G. Parisi, and R. Zecchina. 2002. Analytic and Algorithmic Solution of Random Satisfiability Problems. *Science* 297, 5582 (Aug. 2002), 812–815. DOI: http://dx.doi.org/10.1126/science.1073287
D Mitchell, B Selman, and H Levesque. 1999. Hard and easy distributions of SAT problems. In *AAAI*, Vol. 92. 459–465.
Rémi Monasson, Riccardo Zecchina, Scott Kirkpatrick, Bart Selman, and Lidror Troyansky. 1999. Determining computational complexity from characteristic phase transitions. *Nature* 400, 6740 (1999), 133–137.
Kevin Murphy, Yair Weiss, and Michael Jordan. 1999. Loopy-belief Propagation for Approximate Inference: An Empirical Study. In *IS*. 467–475.
Luis E Ortiz and Michael Kearns. 2002. Nash propagation for loopy graphical games. In *Advances in Neural Information Processing Systems*. 793–800.
Christos H Papadimitriou. 2003. *Computational complexity*. John Wiley and Sons Ltd.
James D Park and Adnan Darwiche. 2004. Complexity Results and Approximation Strategies for MAP Explanations. *J. Artif. Intell. Res.(JAIR)* 21 (2004), 101–133.
Judea Pearl. 1985. Bayesian networks: a model of self-activated memory for evidential reasoning. (1985).
J Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Representation and Reasoning, Vol. 88. Morgan Kaufmann.
Charles C Pinter. 2012. *A book of abstract algebra*. Courier Dover Publications.
Cédric Pralet, Gérard Verfaillie, and Thomas Schiex. 2007. An Algebraic Graphical Model for Decision with Uncertainties, Feasibilities, and Utilities. *J. Artif. Intell. Res.(JAIR)* 28 (2007), 421–489.
S Ravanbakhsh, C Srinivasa, B Frey, and R Greiner. 2014. Min-Max Problems on Factor-Graphs. *ICML* (2014).
Hartley Rogers. 1987. *Theory of recursive functions and effective computability*. (1987).
Dan Roth. 1993. On the hardness of approximate reasoning. In *Proceedings of the 13th international joint conference on Artificial intelligence-Volume 1*. Morgan Kaufmann Publishers Inc., 613–618.
Thomas Schiex, Helene Fargier, and Gerard Verfaillie. 1995. Valued constraint satisfaction problems: hard and easy problems. In *Proceedings of the 14th international joint conference on Artificial intelligence-Volume 1*. Morgan Kaufmann Publishers Inc., 631–637.
Prakash P Shenoy. 1992. Valuation-based systems: A framework for managing uncertainty in expert systems. In *Fuzzy logic for the management of uncertainty*. John Wiley & Sons, Inc., 83–104.
Prakash P Shenoy and Glenn Shafer. 1990. Axioms for probability and belief-function propagation. In *Uncertainty in Artificial Intelligence*.
Solomon Eyal Shimony. 1994. Finding MAPs for belief networks is NP-hard. *Artificial Intelligence* 68, 2 (1994), 389–410.
Larry J Stockmeyer. 1976. The polynomial-time hierarchy. *Theoretical Computer Science* 3, 1 (1976), 1–22.
Seinosuke Toda. 1991. PP is as hard as the polynomial-time hierarchy. *SIAM J. Comput.* 20, 5 (1991), 865–877.
L G. Valiant. 1979. The complexity of computing the permanent. *Theoretical computer science* 8, 2 (1979), 189–201.
Leslie G Valiant and Vijay V Vazirani. 1986. NP is as easy as detecting unique solutions. *Theoretical Computer Science* 47 (1986), 85–93.
John Von Neumann and Oskar Morgenstern. 2007. *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton university press.
Klaus W Wagner. 1986. The complexity of combinatorial problems with succinct input representation. *Acta Informatica* 23, 3 (1986), 325–356.
Martin J Wainwright and Michael I Jordan. 2007. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends in Machine Learning* 1, 12 (2007), 1–305. DOI: http://dx.doi.org/10.1561/2200000001
Yair Weiss. 2001. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural computation* June (2001).
Adrian Weller and Tony S Jehara. 2013. On MAP inference by MWSS on Perfect Graphs. *arXiv preprint arXiv:1309.6872* (2013).
JS Yedidia and WT Freeman. 2001. Generalized belief propagation. *Advances in neural information* (2001).