Gauges, Loops, and Polynomials for Partition Functions of Graphical Models

Michael Chertkov [1,2], Vladimir Chernyak[3], and Yury Maximov[4,2]
[1] Program in Applied Mathematics, University of Arizona, Tucson, AZ, USA
[2] Skolkovo Institute of Technology, Moscow, Russia
[3] Department of Chemistry, Wayne State University, MI, USA
[4] T-Division, Los Alamos National Laboratory, Los Alamos, NM, USA*
(Dated: September 1, 2020)

Graphical models represent multivariate and generally not normalized probability distributions. Computing the normalization factor, called the partition function, is the main inference challenge relevant to multiple statistical and optimization applications. The problem is #P-hard that is of an exponential complexity with respect to the number of variables. In this manuscript, aimed at approximating the partition function, we consider Multi-Graph Models where binary variables and multivariable factors are associated with edges and nodes, respectively, of an undirected multi-graph. We suggest a new methodology for analysis and computations that combines the Gauge Function technique from [21, 22] with the technique developed in [34] and [5, 67] based on the recent progress in the field of real stable polynomials. We show that the Gauge Function, representing a single-out term in a finite sum expression for the Partition Function which achieves extremum at the so-called Belief-Propagation gauge, has a natural polynomial representation in terms of gauges/variables associated with edges of the multi-graph. Moreover, Gauge Function can be used to recover the Partition Function through a sequence of transformations allowing appealing algebraic and graphical interpretations. Algebraically, one step in the sequence consists in the application of a differential operator over gauges associated with an edge. Graphically, the sequence is interpreted as a repetitive elimination/contraction of edges resulting in Multi-Graph Models on decreasing in size (number of edges) graphs with the same Partition Function as in the original Multi-Graph Model. Even though the complexity of computing factors in the sequence of the derived Multi-Graph Models and respective Gauge Functions grow exponentially with the number of eliminated edges, polynomials associated with the new factors remain *Bi-Stable* if the original factors have this property. Moreover, we show that BP estimations in the sequence do not decrease, each low-bounding the Partition Function.

arXiv:1811.04713v6 [cs.LG] 28 Aug 2020

I. INTRODUCTION

Graphical models (GM) are ubiquitous in natural and engineering sciences where one needs to represent a multivariate distribution function with a structure that is expressed in terms of graphical, statistical or deterministic, relations between the variables [29, 41, 51–53, 56, 59, 76]. Focusing on the so-called Normal Factor Graph representation [28], where binary variables and factors that express relations between the variables are associated with edges and nodes of the graph, respectively, we are interested in resolving the problem of statistical inference, which entails computing the weighted sum over allowed states. Exact evaluation of the sum, called the Partition Function, is known to be #-P-hard [38, 69, 71], that is, of complexity which likely requires an exponential number of steps. Subsequently, deterministic and stochastic approximations were made. In this manuscript, we concentrate primarily on the former. (Stochastic methods for the Partition Function estimations are reviewed in [37, 71]. See also some related discussions in [43] and below.)

The inference problem can also be stated as an optimization. The variational approach to Partition Function computation dates back to Gibbs [31], and possibly earlier. Similar considerations are known in statistics under the name of Kullback–Leibler divergence [44]. The resulting optimization stated in terms of beliefs (i.e., proxies for probabilities of states) is convex but not tractable because of the exponential number of states (and respectively beliefs). Developing relaxations, and more generally approximations, for the Gibbs–Kullback–Leibler variational formulation is the primary research to which this manuscript is contributing.

Theoretical efforts in the field of deterministic estimations of Partition Functions have focused on devising (a) lower and/or upper bounds for GMs of a special type and (b) Fully Polynomial Deterministic Algorithmic Schemes for even more restrictive classes of GMs. (See Section VII for an extensive discussion of the low bounds and related subjects. Section VIII for a brief discussion on unification of these ideas with Fully Polynomial Deterministic Algorithmic Schemes.)

* chertkov@arizona.edu, chernyak@chem.wayne.edu, yury@lanl.gov

Provable lower bounds for Partition Functions are known for Perfect Matching problems over bi-partite graphs [34, 35], independent set problems [16, 65, 79], and Ising models of attractive (log-supermodular) [48, 61, 68] and general [60] types. In a few cases where the exact computation of Partition Functions is polynomial, noticeably GM over tree graphs [11, 28, 29, 56, 57] and also cases where Partition Function becomes a determinant of a polynomial (in the size of the original GM) matrix correspondent to Ising, PM and other specialized models over planar graphs [6, 20, 27, 30, 40, 70], would normally be considered good starting points for analysis of lower bounds on the Partition Functions.

This manuscript contributes to the line of research with roots in the tree-graph and dynamic programming (DP) methodology and also its extensions to loopy multi-graphs. The subject has a distinguished history in physics [11, 52, 57], information theory [28, 29, 59], artificial intelligence and machine learning [41, 51, 56], statistics, and computer science [53, 76]. It culminates in the so-called Belief Propagation (BP) analysis, theory, and algorithms. (The term BP was coined by Pearl, who has pioneered related applications in artificial intelligence and machine learning [56].) Applied to graphs with loops, as first done by Gallager in the context of the Low-Density Parity Check codes [29], BP is a practically successful heuristic algorithm, generally lacking quality assurance. The iterative/algorithmic part of BP was connected to the variational Gibbs–Kullback–Leibler interpretation in [83], where it was shown that (in the case of convergence) the BP algorithm corresponds to a fixed point of the so-called Bethe Free Energy, stated in terms of the marginal beliefs associated with nodes and edges of the GM. (See also Section II A for details.) In the following, we refer to BP as a fixed point (possibly one of many) of the Bethe Free Energy, assuming that it can be found efficiently [64]. We will also generalize the notion of fixed points to the cases when the minimum of the Bethe Free Energy may be achieved at a plaquet/side of the belief polytope over which the Bethe Free Energy is defined (not necessarily within the interior of the polytope).

Heuristic success as well as results claiming exactness of BP for some special optimization problems over loopy graphs (e.g., finding maximum weight perfect matching over bi-partite graphs [10]) have stimulated the design of a number of methods relating results of BP to exact results. These methods include Gauge Transformation and Loop Calculus of [21, 22], the spanning tree approach of [79], the cumulant expansion approach of [80], the graph cover approach of [74], and most recently the Real Stable Polynomial approach of [5, 34, 67]. The first and last approaches are most relevant to this manuscript.

The Gauge Transformation–Loop Calculus method of [21, 22] suggests an exact construct exploring invariance of the Partition Function with respect to special transformations of factors, called gauges, also related to the so-called re-parametrizations of [77] and holographic transformations of [70]. It was shown that BP corresponds to a special choice of gauges, which then lead to expressing the Partition Function in terms of the so-called generalized loops, where each generalized loop contribution is stated explicitly in terms of the underlying BP solution. The Gauge Transformation–Loop Calculus approach was utilized (1) to prove that BP provides a lower bound for attractive Ising models with some additional technical constraints in [81] (it was then shown in [61, 62] through the use of the graph cover approach of [74] that the additional constraints are insignificant); (2) to prove that BP is exact asymptotically for an ensemble of independent set problems [16]; (3) to relate matching models, Fermion models of statistical physics with loop and determinant considerations [18, 19]; (4) to approximate Partition Function in planar GM [23, 32]; (5) to apply Loop Calculus to permanent (Partition Function of perfect matching model over bi-partite graph) [78], to provide a proof that is alternative to the original [34] for the fact that BP results in a lower bound for permanent, and then construct a sequence of fractional-BP approximations for permanents [24]; (6) to build a Fully Polynomial Randomized Approximation Schemes for a subclass of planar GMs [1] by sampling Loop Series; and (7) to construct a provable lower bound for Partition Function in the case when BP fails to provide such guarantees [2] by finding an optimal non-BP gauge that certifies that all terms in the Loop Series are positive.

The Real Stable Polynomials approach to the Partition Function, first developed for permanents in [34] and then generalized to binary GM over (normal) bi-partite graphs with submodular factors in [5, 67], is built on the recent progress in the Real Stable Polynomials theory [12–15]. The essence of the approach is in representing the Bethe Free Energy as a polynomial optimization and then showing that the Partition Function is a result of a sequential application of edge-local differential operators to the Bethe Free Energy/BP estimate of Partition Function. It was shown in [67] that if all polynomials associated with nodes of the GM are Real Stable and the graph is bi-partite, then each application of the edge-local differential operator ensures that the respective Partition Function estimate does not decrease, thus resulting in the statement that BP (the zero term in the sequence) provides a lower bound for the partition function (the last term in the sequence). In a related paper, [5] a polynomial version of the GM statement of [67] was proven for a more general case of the so-called Bi-Stable polynomials over arbitrary graphs.

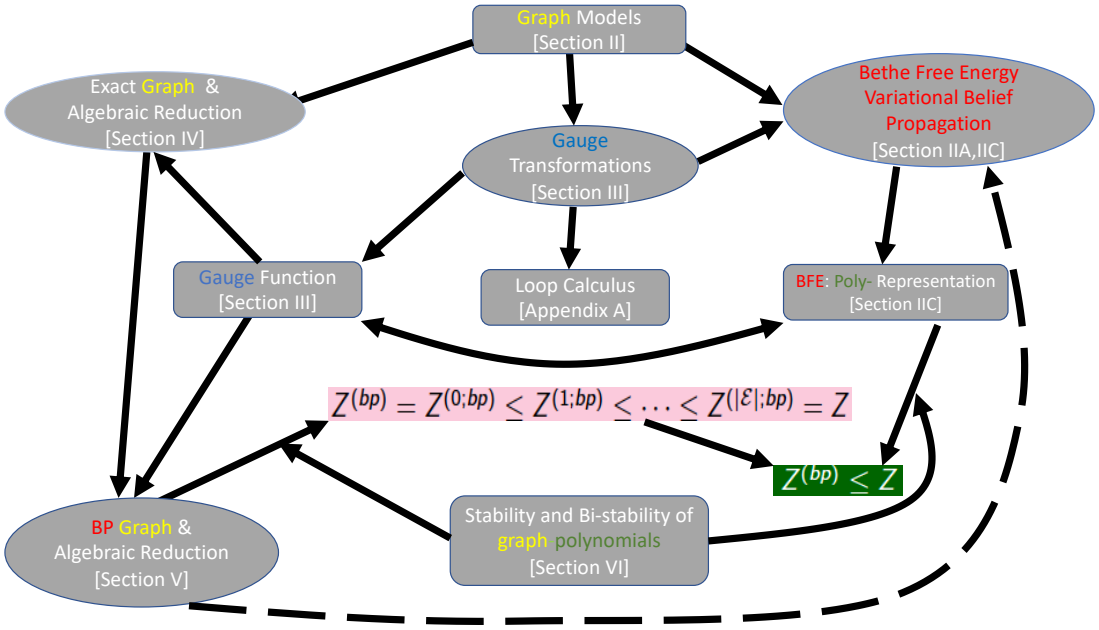


FIG. 1. Relation diagram of steps and concepts of the manuscript with links to respective Sections.

A. Contributions of this manuscript

We complement the Real Stable Polynomials approach of [5, 34, 67] by merging it with the Gauge Transformation approach of [21, 22], and thus in a sense generalize both. Our approach consists of the following steps (see also Fig. (1) presenting a diagram of relations between the manuscript's steps and concepts) :

- Generalize Variational Bethe Free Energy approach (from the case of normal graphical models) to the Multi-Graphical Models. Show that solution of any Soft Multi-Graph Model is attained strictly within a polytope of parameters – so-called Belief Polytope – describing the solution. All further results reported in the paper (unless specified otherwise) apply strictly speaking only to Multi-Graph Models (Multi-GM) which are soft – even though some of the factors may be infinitesimally small. (Section II).

- Restate the Gauge Transformation expression for Partition Function from [21, 22] as a series of polynomials in variables/gauges. Single out a term from the series, which we call the Gauge Function, relate stationary points of the Gauge Function, so-called BP-gauges and show that the minimum of the Bethe Free Energy is achieved at the maximal BP-gauge. (Section III).

- Introduce a sequence of Multi-GMs, where each new member is a result of an edge contraction (graphically) or summation over respective edge variables (algebraically). Build BP polynomial (principal polynomial evaluated at the optimal BP gauge) for each Multi-GM in the sequence such that the last term is the Partition Function (constant) corresponding to the fully contracted graph. Introduce BP-optimal gauge for each Multi-GM and show that BP-optimal estimation stays exact in the process of contraction of a normal edge, however, it becomes approximate respective contraction of a self-edge. (Section IV.)

- Observe that the Bi-Stability of polynomials correspondent to factors of the original Multi-GM results in the Real Stable Polynomials of each factor in each Multi-GM of the aforementioned contraction sequence. Show that the variational BP solution (correspondent to the minimum of the respective Bethe Free Energy) for each next Multi-GM in the sequence is larger or equal to BP if all factors in the original Multi-GM correspond to Real Stable Polynomials polynomials. A direct corollary of this construction is the desired statement that the BP optimal estimation for the Partition Function of the original Multi-GM low bounds the exact Partition Function. (Section V.)

We present in Section II, for the purpose of setting terminology and self-consistency of the presentation, introductory material for the Bethe Free Energy approach. For the Loop Calculus approach of [21, 22], we present introductory material in Appendix A. Section VIII is reserved for discussions of the results and the path forward.

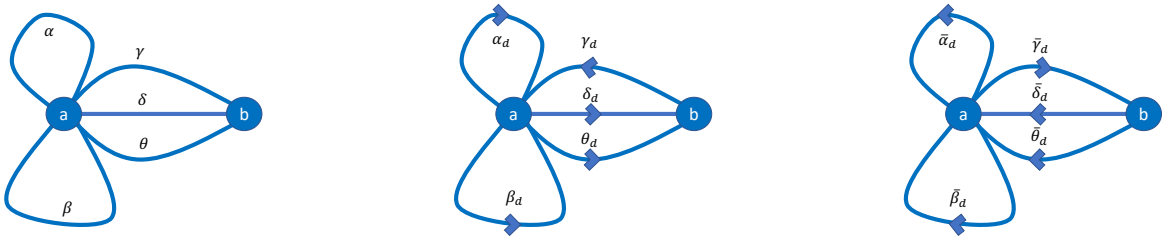


FIG. 2. Multi-Factor Graphical Model notations for undirected (left sub-figure) and directed (middle and right sub-figures) edges are illustrated. $\mathcal{V} = \{a, b\}$, $\mathcal{E} = \{\alpha, \beta, \gamma, \delta, \theta\}$ and $\mathcal{E}_d = \{\alpha_d, \bar{\alpha}_d, \beta_d, \bar{\beta}_d, \gamma_d, \bar{\gamma}_d, \delta_d, \bar{\delta}_d, \theta_d, \bar{\theta}_d\}$ are the sets of nodes, set of undirected edges and set of directed edges, respectively, where $\cdot\bar{\cdot}\cdot$ is the notation used to denote the directed edge $\cdots$ reversal, thus $\bar{\bar{\alpha}}_d = \alpha_d$. $e(a) = \{\alpha, \beta, \gamma, \delta, \theta\}$, $e(b) = \{\gamma, \delta, \theta\}$ and $e_d(a) = \{\alpha_d, \bar{\alpha}_d, \beta_d, \bar{\beta}_d, \gamma_d, \bar{\gamma}_d, \delta_d, \bar{\delta}_d, \theta_d, \bar{\theta}_d\}$, $e_d(b) = \{\gamma_d, \bar{\delta}_d, \bar{\theta}_d\}$ describe functions, $e(\cdots) : \mathcal{V} \to \mathcal{E}$ and $e_d(\cdots) : \mathcal{V} \to \mathcal{E}_d$, mapping a node to the set of directed and undirected edges, respectively, of the multi-graph shown in the figure. $u(e) : \mathcal{E}_d \to \mathcal{E}$ describes function mapping directed edges into respective undirected edges; $u(\alpha_d) = u(\bar{\alpha}_d) = \alpha$. $v(\alpha) : \mathcal{E} \to \mathcal{V}^2$ describes function mapping an undirected edge to its end nodes.

II. PRELIMINARIES

We consider multi-graph generalization of [21, 22]. Following terminology of [50], we may also call it binary Factor-Multi-GM: variables are associated with edges, and factor functions (or simply factors) are associated with nodes of the undirected multi-graph, $\mathcal{G} \doteq (\mathcal{V}, \mathcal{E})$, thus allowing multi-edges between two different nodes and multiple self-edges associated with a single node, where $\mathcal{V}$ and $\mathcal{E}$ are the sets of nodes and edges, respectively. The main reason for dealing with multi-graphs is that certain geometrical graph transformations, namely edge contraction, introduced in Section VII turn simple graphs (no multiple edges, no self-edges) into multi-graphs.

It is also useful to introduce an oriented version of the undirected multi-graph, i.e., a multi-graph, equipped with orientation. We will then denote $\mathcal{E}_d$ the set of directed edges including the original orientation and its reverse. $e_d(a)$, $v(\alpha)$ and $u(\alpha_d)$ will denote, respectively, the set of directed edges associated with the node a, two nodes associated with the undirected edge, α, and undirected edge, α, associated with the directed edge α_d. Also, and abusing notations a bit, α_d may denote the primary oriented edge for previously introduced undirected edge α; $\alpha_d \in e_d(a)$ denotes a variable directed edge picked from the $e_d(a)$ set. See Fig. 2 for clarifying example.

Definition II.1 ((Multi-) Graph Model)**.** *Multi-GM describes factorization for the probability of a binary-component vector, $\sigma \doteq (\sigma_\alpha = 0, 1 | \alpha \in \mathcal{E}) \in \{0,1\}^{|\mathcal{E}|}$, consistent with the (multi)-graph:*

$$p(\sigma) \doteq \frac{f(\sigma)}{Z}, \quad f(\sigma) \doteq \prod_{a \in \mathcal{V}} f_a(\sigma_a), \quad Z \doteq \sum_{\sigma} f(\sigma). \tag{1}$$

Here σ_a is a sub-vector of σ built from all components of the latter containing node a, i.e. $\sigma_a \in \Sigma_a \doteq \{0,1\}^{e_d(a)}$.

Notice that the Partition Function, Z, defined in Eq. (1) as a summation over all configurations, σ, allows a recast in terms of the following exact variational principle.

Theorem II.2 (Gibbs-Kullback-Leibler (Gibbs–Kullback–Leibler) Variational Reformulation for Partition Function, in the spirit of [31, 44])**.** *The Partition Function, Z, defined in Eq. (1) can be computed through the following optimization*

$$-\log Z = \min_{\mathbb{b}} \sum_{\sigma \in \{0,1\}^{|\mathcal{E}|}} \mathbb{b}(\sigma) \log \frac{\mathbb{b}(\sigma)}{\prod_{a \in \mathcal{V}} f_a(\sigma_a)} \Bigg|_{\substack{\forall \sigma : \quad \mathbb{b}(\sigma) \geq 0 \\ \sum_\sigma \mathbb{b}(\sigma) = 1}}, \tag{2}$$

where $\mathbb{b} = (\mathbb{b}(\sigma) | \forall \sigma \in \sigma \in \{0,1\}^{|\mathcal{E}|})$, and $\mathbb{b}(\sigma)$ are beliefs (i.e., proxies for probabilities) of the state σ.

The optimization (2) is convex but not practical because the number of states (and number of respective beliefs) is exponential in the system size (number of edges).

Theorem II.3 (Exact Maximum-A-Posteriori as a Linear Programming)**.** *The Maximum-A-Posteriori versions of Eqs. (1,2) are*

$$E \doteq -\min_{\sigma} \log f(\sigma) = -\min_{\mathbb{b}} \sum_{\sigma} \mathbb{b}(\sigma) \sum_{a\in\mathcal{V}} \log f_a(\sigma_a) \Bigg|_{\substack{\forall \sigma: \quad \mathbb{b}(\sigma)\geq 0 \\ \sum_{\sigma}\mathbb{b}(\sigma)=1}} . \tag{3}$$

Notice that the formulation on the right of Eq. (3) is an LP over (exponentially many) belief variables.

A. Variational Belief Propagation

Belief propagation (BP) is a popular and practical tool that approximates original beliefs via marginal beliefs according to the following dynamic programming (DP) expression [51–53, 59, 76], which gets the following form when stated for the Multi-GM model (1)

$$\mathbb{b}(\sigma) \approx \frac{\prod_{a} b_a(\sigma_a)}{\prod_{\alpha\in\mathcal{E}} \beta_\alpha(1-\beta_\alpha)}, \text{ s.t. } \begin{cases} \forall a\in\mathcal{V},\ \forall \sigma_a \in \Sigma_a^{(0)}:\ b_a(\sigma_a) \doteq \sum\limits_{\sigma\setminus\sigma_a} \mathbb{b}(\sigma); \\ \forall \alpha\in\mathcal{E}:\ \beta_\alpha = \sum\limits_{\sigma\setminus\sigma_\alpha}^{\sigma_\alpha=1} \mathbb{b}(\sigma); \end{cases} \tag{4}$$

which would be exact for a tree-graph (no loops in GM). Reducing description from the exponential in size vector of original beliefs to the linear in the size of the graph (assuming that node degree in the Multi-GM is $O(1)$) vector of marginal beliefs

$$b = (b_a(\sigma_a)|\, \forall a\in\mathcal{V},\ \sigma_a\in\Sigma_a), \quad \beta = (\beta_\alpha|\, \forall \alpha\in\mathcal{E}). \tag{5}$$

considered over the following marginal polytope

$$\Pi \doteq \left\{ (b,\beta)\in[0,1]^{\Sigma}\times[0,1]^{\mathcal{E}} \,\middle|\, \begin{array}{c} \forall a\in\mathcal{V}:\ \sum_{s\in\Sigma_a} b_a(s)=1; \\ \forall a\in\mathcal{V},\ \forall \alpha_d\in\mathcal{E}_{\rm d}(a):\quad \sum_{s\in\Sigma_a}^{s_{u(\alpha_d)}=1} b_a(s) = \beta_{u(\alpha_d)} \end{array} \right\}. \tag{6}$$

and substituting Eq. (4) into Eq. (2), one arrives at the following optimization.

Definition II.4 (Bethe Free Energy and Variational Belief Propagation approximation for the Partition Function, by analogy with [83])**.** *Variational BP estimation for the Partition Function, $Z^{(\rm vbp)}$, is defined according to*

$$F^{(\rm vbp)} \doteq -\log Z^{(\rm vbp)} \doteq \min_{(b,\beta)\in\Pi} F^{(bp)}(b,\beta), \tag{7}$$

$$F^{(bp)}(b,\beta) \doteq E^{(bp)}(b) - S^{(bp)}(b,\beta), \tag{8}$$

$$E^{(bp)}(b) \doteq -\sum_{a\in\mathcal{V}_a}\sum_{\sigma_a\in\Sigma_a} b_a(\sigma_a)\log f_a(\sigma_a), \tag{9}$$

$$S^{(bp)}(b,\beta) \doteq \sum_{a\in\mathcal{V}_a}\sum_{\sigma_a\in\Sigma_a} b_a(\sigma_a)\log b_a(\sigma_a) - \sum_{\alpha\in\mathcal{E}} \left(\beta_\alpha\log\beta_\alpha + (1-\beta_\alpha)\log(1-\beta_\alpha)\right), \tag{10}$$

where $F^{(\rm vbp)}$ is the Variational Bethe Free Energy, $F^{(bp)}(b,\beta)$, $E^{(bp)}(b)$ and $S^{(bp)}(b,\beta)$ are the Bethe Free Energy, Bethe Self Energy and Bethe Entropy functions of marginal beliefs.

We further notice that Eq. (7) may be restated as a polynomial optimization [67]. To derive the polynomial representation one, first, rewrites Eq. (7) as

$$\log Z^{(\rm vbp)} = \max_{\beta\in[0,1]^{\mathcal{E}}} \left(S^{(bp-r)}(\beta) - E^{(bp-r)}(\beta) \right), \tag{11}$$

$$S^{(bp-r)}(\beta) \doteq \sum_{\alpha\in\mathcal{E}} \left(\beta_\alpha\log\beta_\alpha + (1-\beta_\alpha)\log(1-\beta_\alpha)\right), \tag{12}$$

$$E^{(bp-r)}(\beta) \doteq \min_{b\in\Pi_r(\beta)} \sum_{a\in\mathcal{V}}\sum_{\sigma_a\in\Sigma_a} b_a(\sigma_a)\log\frac{b_a(\sigma_a)}{f_a(\sigma_a)} \tag{13}$$

$$\Pi_r(\beta) \doteq \left\{ b\in[0,1]^{\Sigma} \,\middle|\, \begin{array}{c} \forall a\in\mathcal{V}:\ \sum_{s\in\Sigma_a} b_a(s)=1; \\ \forall a\in\mathcal{V},\ \forall \alpha_d\in e_d(a):\quad \sum_{s\in\Sigma_a}^{s_{u(\alpha_d)}=1} b_a(s) = \beta_{u(\alpha_d)} \end{array} \right\}. \tag{14}$$

where $S^{(bp-r)}(\beta)$ and $E^{(bp-r)}(\beta)$ are the reduced BP entropy and the reduced BP self-energy functions, respectively, dependent only on the vector of edge probabilities, β. Applying strong duality to the reduced BP self-energy one derives

$$E^{(bp-r)}(\beta) = \sup_{x\in\mathbb{R}_+^{\mathcal{E}_\mathrm{d}}} \left(\sum_{\alpha\in\mathcal{E}} \beta_\alpha \log(x_{\alpha_d} x_{\bar{\alpha}_d}) - \sum_{a\in\mathcal{V}} \log(h_a(x_a)) \right), \tag{15}$$

$$h_a(x_a) \doteq \sum_{s\in\Sigma_a} f_a(s) \prod_{\alpha\in\mathcal{E}_\mathrm{d}(a)} x_\alpha^{s_\alpha}, \tag{16}$$

where, $x \doteq (x_\alpha > 0 | \forall \alpha \in \mathcal{E}_\mathrm{d}) \in \mathbb{R}_+^{\mathcal{E}_\mathrm{d}}$ and $x_a \doteq (x_\alpha > 0 | \forall \alpha \in e_\mathrm{d}(a)) \in \mathbb{R}_+^{e_\mathrm{d}(a)}$. The $\log(x_\alpha)$ components, with $\alpha \in \mathcal{E}_\mathrm{d}$ were introduced as Lagrangian multipliers (dual variables) for the belief consistency conditions in Eq. (9). Combining Eqs. (7,8,11,12,15), one arrives at the following statement.

Theorem II.5 (Polynomial Max-Min Representation for Variational BP, multi-graph version of Theorem 3.1 of [67])**.** *Variational BP, defined Eqs. (7,8,9,10), can also be stated as the following max-min optimization:*

$$Z^{(\mathrm{vbp})} = \sup_{\beta\in[0;1]^{\mathcal{E}}} \min_{x\in\mathbb{R}_+^{\mathcal{E}_\mathrm{d}}} \mathcal{L}(\beta, x), \tag{17}$$

$$\mathcal{L}(\beta, x) \doteq \left(\prod_{\alpha\in\mathcal{E}} \beta_\alpha^{\beta_\alpha} (1-\beta_\alpha)^{1-\beta_\alpha} \right) \prod_{a\in\mathcal{V}} \frac{h_a(x_a)}{\prod\limits_{\alpha_d\in\mathcal{E}_d(a)} x_\alpha^{\beta_{u(\alpha_d)}}}, \tag{18}$$

B. Linear Programming Relaxation for Maximum-A-Posteriori

Definition II.6 (Linear programming–Belief Propagation approximation)**.** *Linear Programming – Belief Propagation approximation for Maximum-A-Posteriori optimization (3) is*

$$E^{(\mathrm{lp-bp})} \doteq - \min_{(b,\beta)\in P_B} \left(\sum_{a\in\mathcal{V}} \sum_{\sigma_a\in\Sigma_a} b_a(\sigma_a) \log f_a(\sigma_a) \right). \tag{19}$$

Notice that Linear Programming – Bellief Propagation is tractable, and it can be considered both as the "entropy-free" version of the optimization (7) and also as a relaxation of the exact LP formulation (3) and therefore results in the following statement.

Theorem II.7 (Lower bounding by Linear Programming – Belief Progagation (see for example [39, 66, 76] and references therein))**.** *Linear Programming – Belief Progagation lower bounds exact self-energy.* $E^{(\mathrm{lp-bp})} \leq E$.

Note that the same Linear Programming – Belief Progagation is known under the name of "Basic Linear Programming Relaxation" in the community analyzing Constrain Satisfaction Problems. See, e.g. [42] and references therein.

C. Variational Belief Propagation in the Soft Model

Definition II.8 (Soft Multi-GM)**.** *If* $\forall a \in \mathcal{V},\ \forall \sigma_a \in \Sigma_a :\ f_a(\sigma_a) > 0$*, the Multi-GM is called soft.*

Theorem II.9 (Variational BP of Soft Multi-GM – in the spirit of Proposition 6 of [83])**.** *Minimum in Eq. (7) is achieved within the interior of* Π *in the case of soft Multi-GM.*

Proof. The theorem is proved in three steps: first, one shows that the minimum in Eq. (7) cannot be achieved at β such that at one edge, α (at least one edge), β_α is exactly zero or one; (b) given (a) one checks explicitly that when all factors are soft (and thus no terms in the multi-linear polynomials h_a are zero) the minimum over x in Eq. (17) is achieved at a finite $x \in \mathbb{R}_+^{\mathcal{E}_d}$; finally, given that the Lagrangian multipliers, x, for the edge and node belief consistency are all finite, the minimum in Eq. (7) can only be achieved at $\forall a \in \mathcal{V},\ \forall \sigma_a \in \Sigma_a :\quad b_a(\sigma_a) \in]0;1[$.

Therefore, only the first step is left to be proven. We present here only a sketch of the proof. Assume that $\exists \alpha \in \mathcal{E}$ such that the minimum in Eq. (7) is achieved at $\beta_\alpha = 0$. Our strategy consists in showing that one can find a direction from the point on the polytope boundary towards interior along which $-\log \mathcal{Z}(b)$ will decrease, thus arriving at a

contradiction. Indeed, when $\beta_\alpha = \epsilon > 0$ with $\epsilon \to 0$, one derives that according to the belief consistency relations, i.e. equality relations between beliefs embedded in the definition (6) of the polytope Π, all $b_a(\sigma_a) = O(\epsilon)$ where $\forall a \in v(\alpha)$, σ_a is consistent with $\sigma_\alpha = 1$. Moreover, one may redistribute the $O(\epsilon)$ perturbations over these $b_a(\sigma_a)$ such that β_γ, where $\gamma \neq \alpha$, do not depend on the ϵ-perturbation at all. On the other hand, ϵ-corrections to $-\log \mathcal{Z}(b)$ are $O(\epsilon \log \epsilon)$. The corrections originate from the entropy contributions associated with $O(\epsilon)$ beliefs of two types — associated with β_α and associated with the respective $b_a(\sigma_a)$. Accurate counting of the contributions results in the overall $\epsilon \log \epsilon$ correction to $-\log \mathcal{Z}(b)$, where $2\epsilon \log \epsilon$ term comes from the two $b_a(\sigma_a) = O(\epsilon)$ contributions and one $-\epsilon \log \epsilon$ term comes from the single β_α contribution. The resulting, $\epsilon \log \epsilon$, is negative and it decreases with increase in ϵ, thus leading to the contradiction. Similar consideration, now with, $\beta_\alpha = 1 - \epsilon$, where $\epsilon > 0$, $\epsilon \to 0$, results in the statement that at the optimum β_α cannot be equal to unity. □

Note that solution of Eq. (7) can be on the boundary of the Π polytope if the Multi-GM is hard. See [46, 78] for discussion of special hard cases, e.g. of the perfect matching problem, where solution is achieved at the boundary of Π.

Theorem II.9 guarantees that an infinitesimally weak softening of a hard model (achieved by adding an infinitesimal positive correction to $f_a(\sigma_a) = 0$ factors) shifts a solution of the optimization (8) into the interior of the polytope. We will use this softening feature of Multi-GM later in Section III B to relate Variation Belief Propagation formulations and solutions discussed in this Section to the Gauge Transformation and Belief Propagation Equations we are switching out attention to in the next Section.

III. GAUGE TRANSFORMATION AND BELIEF PROPAGATION EQUATIONS

A. Gauge Transformation

Definition III.1 (Gauge Transformation, [21, 22]). *Gauge Transformation is a multi-linear transformation of the GM factors:*

$$\forall a \in \mathcal{V}, \quad \forall \sigma_a \in \Sigma_a^{(0)}: \quad f_a(\sigma_a) \mapsto \tilde{f}_a(\sigma_a|G) \doteq \sum_{\varsigma_a \in \Sigma_a} f_a(\varsigma_a) \prod_{\alpha \in \mathcal{E}_{\rm d}(a)} G_\alpha(\sigma_\alpha, \varsigma_\alpha),$$

which keeps the Partition Function invariant; that is,

$$\begin{aligned} \forall G: \quad Z &= \sum_{\sigma \in S} \prod_{a \in \mathcal{V}} f_a(\sigma_a) = \sum_{\sigma \in S^{(0)}} \prod_{a \in \mathcal{V}} \tilde{f}_a(\sigma_a|G) = \sum_{\sigma \in S^{(0)}} z(\sigma|G), \\ z(\sigma|G) &\doteq \sum_{\varsigma \in S^{(0)}} \prod_{a \in \mathcal{V}} f_a(\varsigma_a) \prod_{\alpha \in \mathcal{E}_{\rm d}(a)} G_\alpha(\sigma_\alpha, \varsigma_\alpha), \end{aligned} \tag{20}$$

where $\varsigma_a \in \Sigma_a^{(0)} \doteq \{0,1\}^{\mathcal{E}_{\rm d}(a)}$.

It is straightforward to check that Eq. (20) holds if the following condition is met.

Theorem III.2 (Orthogonality of Gauge Transformation [21, 22]). *Gauge Transformation* 2×2 *(in the case of a binary alphabet) matrices satisfy*

$$\forall \alpha \in \mathcal{E}, \quad G_{\alpha_d}^T * G_{\bar{\alpha}_d} = \mathbb{1}_\alpha, \tag{21}$$

where α_d *and* $\bar{\alpha}_d$ *mark two directed siblings of* α*, and the matrices,* G_{α_d} *and* $G_{\bar{\alpha}_d}$ *are non-singular with real-valued components.*

To lift the gauge-constraint (21), one introduces the following explicit representation for G.

Definition III.3 (Polynomial, x-, Representation of Gauges). *We call the following representation for* G*, polynomial- or* x*-representation.*

$$G = (G_\alpha \,|\, \alpha \in \mathcal{E}) \quad G_\alpha = \frac{1}{(x_{\alpha_d} x_{\bar{\alpha}_d})^{1/4} \sqrt{1 + x_{\alpha_d} x_{\bar{\alpha}_d}}} \begin{pmatrix} \sqrt{x_{\bar{\alpha}_d}} & x_{\alpha_d} \sqrt{x_{\bar{\alpha}_d}} \\ -x_{\bar{\alpha}_d} \sqrt{x_{\alpha_d}} & \sqrt{x_{\alpha_d}} \end{pmatrix}, \tag{22}$$

where the vector x *is positive component-wise, i.e.,* $x_\alpha > 0$, $\forall \alpha \in \mathcal{E}_{\rm d}$.

A number of remarks are in order.

- x-representation for G (22) satisfies Eq. (21) automatically (by construction).
- The trivial, $G = \mathbb{1}$, case is recovered in the $(x_{\alpha_d} = x_{\bar{\alpha}_d}) \to 0$ limit.
- Emergence of negative components in the matrix (in the lower left corner of the representation of Eq. (22)) is unavoidable in order to ensure validity of Eq. (21).
- Parameterized according to Eq. (22), $z(\sigma|G)$, defined in Eq. (20), adopts the following form:

$$
\begin{aligned}
z(\sigma|G) = z(\sigma|x) = & \left(\prod_{\beta\in\mathcal{E}} \frac{1}{1+x_{\beta_d}x_{\bar{\beta}_d}}\right) \\
& \times \prod_{a\in\mathcal{V}} \left(\sum_{\varsigma_a\in\Sigma_a^{(0)}} \left(\prod_{\alpha\in\mathcal{E}_{\rm d}(a)} \left(x_\alpha^{\varsigma_{t(\alpha)}} (x_\alpha x_{t(\alpha)})^{(1/2-\varsigma_\alpha)\sigma_\alpha} (-1)^{(1-\varsigma_\alpha)\sigma_\alpha}\right)\right) f_a(\varsigma_a)\right) && (23)\\
= & \left(\prod_{\alpha\in\mathcal{E}} \frac{(x_{\alpha_d}x_{\bar{\alpha}_d})^{\sigma_{\alpha_d}}}{1+x_{\alpha_d}x_{\bar{\alpha}_d}}\right) \prod_{a\in\mathcal{E}} Q_a(x_a;\sigma_a), \text{where} && (24)\\
Q_a(x_a;\sigma_a) \doteq & \sum_{\varsigma\in S^{(0)}} f_a(\varsigma_a) \prod_{\alpha\in\mathcal{E}_{\rm d}(a)} \left(x_\alpha^{\varsigma_\alpha} (-1)^{\sigma_\alpha} (-x_\alpha x_{t(\alpha)})^{-\varsigma_\alpha\sigma_\alpha}\right) && (25)\\
= & \left(\prod_{\beta_d\in\mathcal{E}_{\rm d}(a)}^{\sigma_\beta=1} \frac{1+x_{\beta_d}x_{\bar{\beta}_d}}{x_{\beta_d}x_{\bar{\beta}_d}}\right) \sum_{\varsigma_a\in\Sigma_a^{(0)}} f_a(\varsigma_a) \prod_{\alpha\in\mathcal{E}_{\rm d}(a)} \left(x_\alpha^{\varsigma_\alpha} \left(\varsigma_\alpha - \frac{x_\alpha x_{t(\alpha)}}{1+x_\alpha x_{t(\alpha)}}\right)^{\sigma_\alpha}\right). && (26)
\end{aligned}
$$

  Here, we transitioned in our notation from the general gauges G to the x-representation; $x_a \doteq (x_\alpha \,|\, \alpha \in \mathcal{E}_{\rm d}(a))$.
- $z(\sigma|x)$ is a polynomial in x, up to the factor on the left-hand side of the first raw in Eq. (23) and when stated in terms of x, thus explaining the name chosen for the representation.
- $Z = \sum_\sigma z(\sigma|x)$ is a constant; that is, x-independent, polynomial in x.

We observe in the next sections that a single out σ-term, say $\sigma = 0$ (chosen without loss of generality) and represented as

$$
z(x) \doteq \frac{\prod_{a\in\mathcal{V}} h_a(x_a)}{\prod_{\alpha\in\mathcal{E}} (1+x_{\alpha_d}x_{\bar{\alpha}_d})}, \tag{27}
$$

where h_a is the vertex polynomial, defined in Eq. (16), and $z(x)$ is a short-cut notation for $z(0|x)$ that plays a special role in establishing known and new relations. We call $z(x)$, described by Eq. (27), the Gauge Function of the Multi-GM (1).

B. Belief Propagation Gauges

Definition III.4 (Interior Belief Propagation gauge [21, 22] for Soft Multi-Graph Models)**.** *We call a solution $x^{(\rm bp)} \in \mathbb{R}_+^{\mathcal{E}_{\rm d}}$ of the following stationary-point condition for the Gauge Function (27) equations of the Soft Multi-Graph Models*

$$
\forall a \in \mathcal{V}, \quad \forall \alpha \in \mathcal{E}_{\rm d}(a) : \partial_{x_\alpha} z(x)|_{x=x^{(\rm bp)}} = 0, \tag{28}
$$

a BP gauge.

A relation between the Variational BP and BP gauge approaches is established by the following straightforward corollary of Theorem II.9.

Corollary III.5 (Soft Multi-GM BP gauge optimality)**.** *In the case of Soft Multi-GM there exists a BP-gauge, x, solving Eq. (28)*

$$
Z^{(\rm vbp)} = z(x), \tag{29}
$$

where the Variational BP estimate for the partition function was defined in Eq. (7).

Given that $z(x)$ is differentiable in x, BP-equations (28) are well defined. Explicit version of Eq. (28), derived from Eq. (27), is:

$$\forall a, \quad \forall \alpha_d \in e_d(a): \quad \sum_{\varsigma_a} f_a(\varsigma_a) \left(\prod_{\beta_d \in e_d(a)} (x_{\beta_d}^{(bp)})^{\varsigma_{\beta_d}} \right) \left(\frac{x_{\alpha_d}^{(bp)} x_{\bar{\alpha}_d}^{(bp)}}{1 + x_{\alpha_d}^{(bp)} x_{\bar{\alpha}_d}^{(bp)}} - \varsigma_\alpha \right) = 0. \tag{30}$$

Theorem III.6 (BP-Solution as the "No Loose Coloring" Condition [21, 22])**.** *The BP Eq. (28), or equivalently Eq. (30) can be restated in terms of the following conditions:*

$$\forall a, \quad \sum_{\alpha_d \in e_d(a)} \sigma_{\alpha_d} = 1: \quad Q_a(x_a; \sigma_a) = 0. \tag{31}$$

This result follows from the definition of Q in Eq. (25) or Eq. (26). With regard to the name chosen for the Theorem III.6, it emphasizes that rewriting BP equations in the form of Eq. (31) highlights interpretation of BP in terms of the "edge coloring". Indeed, Eq. (31) enforces cancellation (exact zero) for all $z(\sigma|x) = z(\sigma|G)$ terms in the series on the right-hand side of Eq. (20), where at least one node, a, has one of its neighboring edges, say $\alpha_d \in e_d(a)$, "colored"; that is, it is set to $\sigma_{\alpha_d} = 1$, whereas all other neighboring edges of the node, $\beta_d \in e_d(a),\ \beta_d \neq \alpha_d$, remain "uncolored", that is, set to, $\sigma_{\beta_d} = 0$.

Solutions of the BP Eq. (30) (or Eq. (31)) also allow a transparent interpretation in terms of the edge-consistent (but graph-globally not consistent) probability distributions. Expressions for the node- and edge- marginal probabilities evaluated at a BP gauge are

$$P_a^{(bp)}(\varsigma_a) \doteq \frac{f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}^{(bp)}(0, \varsigma_{\alpha_d})}{\sum_{\varsigma_a} f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} G_{\alpha_d}^{(bp)}(0, \varsigma_{\alpha_d})} = \frac{f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} (x_{\alpha_d}^{(bp)})^{\varsigma_{\alpha_d}}}{\sum_{\varsigma_a} f_a(\varsigma_a) \prod_{\alpha_d \in e_d(a)} (x_{\alpha_d}^{(bp)})^{\varsigma_{\alpha_d}}}, \tag{32}$$

$$P_\alpha^{(bp)}(\varsigma_{\alpha_d}) \doteq G_{\alpha_d}^{(bp)}(0, \varsigma_{\alpha_d}) G_{\bar{\alpha}_d}^{(bp)}(0, \varsigma_{\alpha_d}) = \frac{\left(x_{\alpha_d}^{(bp)} x_{\bar{\alpha}_d}^{(bp)} \right)^{\varsigma_{\alpha_d}}}{1 + x_{\alpha_d}^{(bp)} x_{\bar{\alpha}_d}^{(bp)}}. \tag{33}$$

Edge-consistency of the marginal beliefs means

$$\forall a, \quad \forall \alpha_d \in e_d(a), \quad \forall \varsigma_{\alpha_d} = \{0, 1\}:$$
$$P_\alpha^{(bp)}(\varsigma_{\alpha_d}) = \frac{\left(x_{\alpha_d}^{(bp)} x_{\bar{\alpha}_d}^{(bp)} \right)^{\varsigma_{\alpha_d}}}{1 + x_{\alpha_d}^{(bp)} x_{\bar{\alpha}_d}^{(bp)}} = \frac{\sum_{\varsigma_a \setminus \varsigma_{\alpha_d}} f_a(\varsigma_a) \prod_{\beta_d \in e_d(a)} (x_{\beta_d}^{(bp)})^{\varsigma_{\beta_d}}}{\sum_{\varsigma_a} f_a(\varsigma_a) \prod_{\gamma_d \in e_d(a)} (x_{\gamma_d}^{(bp)})^{\varsigma_{\gamma_d}}} = \sum_{\varsigma_a \setminus \varsigma_{\alpha_d}} P_a^{(bp)}(\varsigma_a). \tag{34}$$

Multiplying Eq. (34) on ς_{α_d} and summing it up over $\varsigma_{\alpha_d} = 0, 1$, one arrives at the already introduced system of BP Eqs. (30). We also note for consistency with earlier notations that

$$\forall \alpha \in \mathcal{E}: \beta_\alpha^{(bp)} = P_\alpha^{(bp)}(1), \quad \beta^{(bp)} \doteq \left(\beta_\alpha^{(bp)} \middle| \, \alpha \in \mathcal{E} \right). \tag{35}$$

Notice that, consistently with the Corollary III.5, Soft Multi-GM may have multiple (more then one) BP-gauges solving Eqs. (28).

IV. ELIMINATION OF EDGES

Assume some ordering of the graph edges,

$$m = 1, \cdots, |\mathcal{E}|: \alpha^{(1)}, \cdots, \alpha^{(|\mathcal{E}|)} \tag{36}$$

and consider summing up over (i.e., eliminating) edges in the expression for the Partition Function (1) one-by-one according to this order, therefore naturally arriving at the sequence of graphs, $\mathcal{G}^{(0)} = (\mathcal{V}^{(0)}, \mathcal{E}^{(0)}), \cdots, \mathcal{G}^{(|\mathcal{E}|)} = (\mathcal{V}^{(|\mathcal{E}|)}, \mathcal{E}^{(|\mathcal{E}|)})$ such that $\forall m = 1, \cdots, |\mathcal{E}|: \quad \mathcal{E}^{(m)} \doteq \mathcal{E}^{(m-1)} \setminus \alpha^{(m)}$. Each next graph in the sequence has one less number of edges and the same or one less number of nodes than its predecessor, see Fig. (3) for illustration. (If the number of nodes at elementary step of the sequence decreases by one, then one of the two merged nodes, chosen

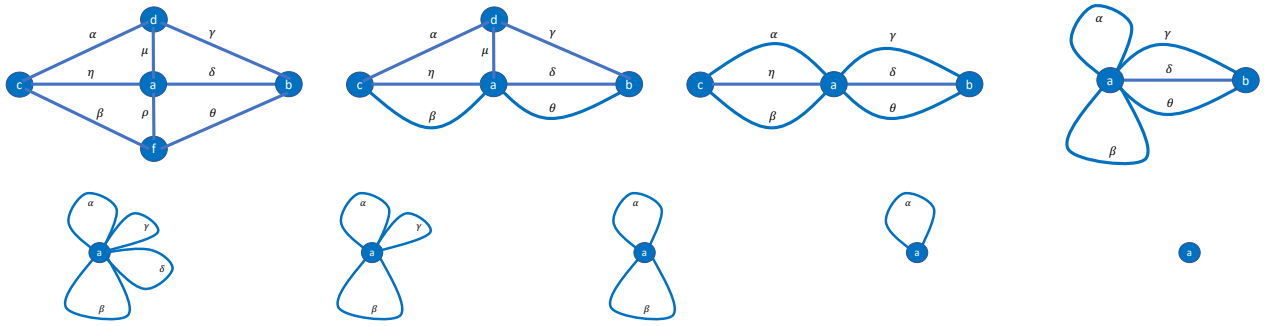


FIG. 3. Graph transformation via elimination of edges is demonstrated through a sequence of sub-figures (from left to right, top to bottom). Each step results in elimination of an edge. Eliminations of edges of two types are possible: (1) elimination of an edge connecting two nodes, e.g., as seen eliminating edges ρ, μ, η and δ in the first four steps in the sequence, and (2) elimination of self-edges, as seen eliminating edges $\delta, \gamma, \beta, \alpha$.

arbitrarily is removed from the resulting/new set of nodes.) Then we define a sequence of Multi-GMs, on the sequence of graphs just defined, as follows

$$
\begin{aligned}
m = 0: \quad & \sigma^{(0)} = \sigma, \quad p^{(0)}(\sigma^{(0)}) \doteq p(\sigma), \\
\forall m = 1, \cdots, |\mathcal{E}|: \quad & \sigma^{(m)} = \sigma^{(m-1)} \setminus \sigma_{\alpha^{(m)}}, \quad p^{(m)}(\sigma^{(m)}) \doteq \frac{f^{(m)}(\sigma^{(m)})}{Z}, \quad f^{(m)}(\sigma^{(m)}) \doteq \prod_{a \in \mathcal{V}^{(m)}} f_a^{(m)}(\sigma_a^{(m)}), \\
& \forall a \in \mathcal{V}^{(m)},\ a \notin v(\alpha^{(m)}): \quad f_a^{(m)}(\sigma_a^{(m)}) \doteq f_a^{(m-1)}(\sigma_a^{(m-1)}), \\
& a \in \mathcal{V}^{(m)}, v(\alpha^{(m)}), \mathcal{V}^{(m-1)}: \quad f_a^{(m)}(\sigma_a^{(m)}) \doteq \sum_{\sigma_{\alpha^{(m)}}} \prod_{b \in v(\alpha^{(m)})} f_b^{(m-1)}(\sigma_b^{(m-1)})
\end{aligned}
\tag{37}
$$

where $v(\alpha^{(m)})$ is defined as a set of nodes in $\mathcal{V}^{(m-1)}$ associated with the edge $\alpha^{(m)}$ (there may be one or two of these depending on if the edge is a self-edge or edge linking two distinct nodes). Notice also that the elimination procedure just introduced is such that the node picked in the last condition (last line) in Eq. (37) is defined uniquely.

In the following, we state a number of remarks about the elimination sequence.

- Notice that by construction, the Partition Function stays the same for all the Multi-GMs in the sequence. The following relations elucidate this point, $m = 1, \cdots, |\mathcal{E}|$

$$
Z = \sum_{\sigma} f(\sigma) = \sum_{\sigma_{\alpha(|\mathcal{E}|)}} \cdots \sum_{\sigma_{\alpha(1)}} \prod_{a \in \mathcal{V}} f_a(\sigma_a) = \sum_{\sigma_{\alpha(|\mathcal{E}|)}} \cdots \sum_{\sigma_{\alpha(m)}} \prod_{a \in \mathcal{V}} f_a(\sigma_a) = \sum_{\sigma^{(m)}} f^{(m)}(\sigma^{(m)}). \tag{38}
$$

- As illustrated in Fig. (3), the elimination procedure may lead to double-, triple-, and in general multiple-edges connecting the same nodes, and it may also result in self-edges, even if the original graph is a normal/simple graph. In fact, this observation explains why we choose to work in this manuscript with the general Multi-GMs.

- Even though we are free to choose any edge-elimination sequence, it is reasonable to eliminate, first, all normal edges (having two distinct nodes associated with the edge), as in the illustrative example of Fig. (3). This results in a "bouquet" (of self-edges) graph containing a single node and multiple self-edges. Notice that if the original graph is a tree the resulting bouquet graph contains no self-edges, and then in this case the elimination sequence is completed. It is straightforward to check that the number of self-edges in the bouquet is invariant of the the normal portion of the edge elimination procedure, i.e. it does not depend on the order of the normal edge eliminations. Moreover this number is exactly equal to the number of edge cuts one needs to apply to the original graph to turn it into a tree. In the following (and unless specify otherwise) we will be assuming that the normal edges are eliminated in the elimination sequence first. Significance of the bouquet graph for our procedure will become clear in the following when we analyze application of the BP procedure to Multi-GMs from the sequence.

- Obviously the exact elimination (37) is not practical because the factor degree and most importantly the complexity of computing the factors grow exponentially with m.

- A number of approximate elimination schemes was introduced in the past to bypass the hardness of the exact computations of Z. Of these approximations the mini-bucket elimination schemes [3, 4, 25, 49] are arguably the most popular and also related to the gauge Gauge Transformation and BP subjects. See Section VIII for additional discussions.
- Even though the sequence of Multi-GMs is not tractable, it is still of a theoretical interest (e.g., in relation to analyzing the class of Multi-GMs where one can derive tractable bounds on Gauge Function, to analyze an intermediate Multi-GM in the sequence from the perspective of the Gauge Transformation and BP). This approach is explored in the following section.

V. FROM GAUGE FUNCTION, $z(x)$, TO PARTITION FUNCTION, Z

Theorem V.1 ("Differentiate+marginalize"). *Exact Partition Function, Z, of the GM can be recovered from the Gauge Function, $z(x)$, defined in Eq. (27), via application of the following mixed-derivative operator:*

$$\begin{aligned}
m=0: \quad & x^{(0)} \doteq x, \quad \mathrm{Z}^{(0)}(x) \doteq z(x), && (39)\\
m=1,\cdots,|\mathcal{E}|: \quad & x^{(m)} \doteq x^{(m-1)} \setminus \{x_{\alpha_d^{(m)}}, x_{\bar\alpha_d^{(m)}}\},\\
& \mathrm{Z}^{(m)}(x^{(m)}) \doteq \left(1+\partial_{x_{\alpha_d^{(m)}}}\partial_{x_{\bar\alpha_d^{(m)}}}\right)\left(\left(1+x_{\alpha_d^{(m)}}x_{\bar\alpha_d^{(m)}}\right)\mathrm{Z}^{(m-1)}\right)\Big|_{x_{\alpha_d^{(m)}}=x_{\bar\alpha_d^{(m)}}=0}, && (40)\\
m=|\mathcal{E}|: \quad & x^{(|\mathcal{E}|)}=\emptyset, \quad \mathrm{Z}^{(|\mathcal{E}|)}=Z, && (41)
\end{aligned}$$

where ordering of edges according to Eq. (36) is assumed.

The statement of Theorem V.1 expressed in Eqs. (39,40,41) is a direct consequence of the following observation:

$$\left(1+\partial_{x_{\alpha_d}}\partial_{x_{\bar\alpha_d}}\right)\left(x_{\alpha_d}^{\varsigma_{\alpha_d}}x_{\bar\alpha_d}^{\varsigma_{\bar\alpha_d}}\right)\Big|_{x_{\alpha_d}=x_{\bar\alpha_d}=0}=\delta\left(\varsigma_{\alpha_d},\varsigma_{\bar\alpha_d}\right), \tag{42}$$

where $\delta\left(\varsigma_{\alpha_d},\varsigma_{\bar\alpha_d}\right)$ stands for the Kronecker symbol, which returns unity when $\varsigma_{\alpha_d}=\varsigma_{\bar\alpha_d}$ and is zero otherwise.

Furthermore, comparing sequence of transformations in the Theorem V.1 with the edge-elimination sequence described in Section IV one arrives at the following statement.

Theorem V.2 (Equivalence of the Algebraic (differentiate+marginalize) and Graphical (edge elimination) transformations). *$\mathrm{Z}^{(m)}(x^{(m)})$, defined in Eq. (40), is the Gauge Function introduced in Eq. (27) however applied to the m-th GM in the edge-elimination sequence defined in Eq. (37):*

$$\mathrm{Z}^{(m)}(x^{(m)})=\frac{h^{(m)}(x^{(m)})}{\prod_{\alpha\in\mathcal{E}^{(m)}}\left(1+x_{\alpha_d}x_{\bar\alpha_d}\right)}, \tag{43}$$

$$h^{(m)}(x^{(m)}) \doteq \prod_{a\in\mathcal{V}^{(m)}}\left(\sum_{\varsigma_a\in\Sigma_a^{(m)}} f_a^{(m)}(\varsigma_a)\prod_{\alpha\in\mathcal{E}_{\mathrm{d}}^{(m)}(a)} x_\alpha^{\varsigma_\alpha}\right). \tag{44}$$

The sequence of the graph-algebraic transformations from the Gauge Function, $z(x)$, to the Partition Function, Z, introduced and discussed in this and preceding sections is the main technical point of this manuscript.

In the next section we relate this sequence, and each step in the sequence of edge eliminations resulting in the mapping, to BP estimations of Multi-GMs in the sequence.

VI. BP-ELIMINATION

Let us take advantage of the rational, in x, structure of the Gauge Function, $z(x)$, and represent $h(x)$ in Eq. (27), as a generic quadratic function of x_{α_d} and $x_{\bar\alpha_d}$, where the edge α of the original graph, $\mathcal{G}$ is selected arbitrarily,

$$h(x) \doteq h^{(0,0)}+h^{(1,0)}x_{\alpha_d}+h^{(0,1)}x_{\bar\alpha_d}+h^{(1,1)}x_{\alpha_d}x_{\bar\alpha_d}, \tag{45}$$

the coefficients of the expansion are non-negative, $\forall i,j=0,1, \quad h^{(i,j)}\geq 0$, and also dependent on $x^{(1)}=x\setminus\{x_{\alpha_d},x_{\bar\alpha_d}\}$ (the dependence is dropped here and also in some formulas below to avoid bulky expressions). Substituting Eq. (45)

into the BP-equations (28) for the edge α, one arrives at a quadratic equations for x_{α_d} and $x_{\bar\alpha_d}$, which results in two roots of which only one is physical, i.e. consistent with respective positive marginal probabilities:

$$x^{(\alpha-\mathrm{bp})}_{\alpha_d} = \frac{h^{(1,1)} - h^{(0,0)} + \sqrt{\left(h^{(1,1)} - h^{(0,0)}\right)^2 + 4h^{(0,1)}h^{(1,0)}}}{2h^{(1,0)}}, \tag{46}$$

$$x^{(\alpha-\mathrm{bp})}_{\bar\alpha_d} = \frac{h^{(1,1)} - h^{(0,0)} + \sqrt{\left(h^{(1,1)} - h^{(0,0)}\right)^2 + 4h^{(0,1)}h^{(1,0)}}}{2h^{(0,1)}}, \tag{47}$$

The value of $h(x)/(1+x_{\alpha_d}x_{\bar\alpha_d})$ evaluated at the physical α-BP-gauge is

$$h^{(\alpha-\mathrm{bp})} = \left.\frac{h(x)}{1+x_{\alpha_d}x_{\bar\alpha_d}}\right|_{x_{\alpha_d}=x^{(\alpha-bp)}_{\alpha_d},x_{\bar\alpha_d}=x^{(\alpha-bp)}_{\bar\alpha_d}} = \frac{h^{(1,1)} + h^{(0,0)} + \sqrt{\left(h^{(1,1)} - h^{(0,0)}\right)^2 + 4h^{(0,1)}h^{(1,0)}}}{2}. \tag{48}$$

Notice that the expression (45) for, h, simplifies when α is a normal edge (not a self-edge). In this special case h is a product of two polynomials of the first order over x_{α_d} and $x_{\bar\alpha_d}$, respectively, i.e. in this case $h^{(1,1)}h^{(0,0)} = h^{(0,1)}h^{(1,0)}$. Then Eqs. (46,47) and Eq. (48) transform to

$$\text{If } \alpha \text{ is a normal edge:}\quad x^{(\alpha-\mathrm{bp})}_{\alpha_d} = \frac{h^{(1,1)}}{h^{(1,0)}},\quad x^{(\alpha-\mathrm{bp})}_{\bar\alpha_d} = \frac{h^{(1,1)}}{h^{(0,1)}},\quad h^{(\alpha-\mathrm{bp})} = h^{(1,1)} + h^{(0,0)}. \tag{49}$$

Comparing Eq. (49) with the first step of the exact edge elimination (first step of the mixed derivative application to the Partition Function) described in Section V, observing that the result is equivalent to the exact elimination, as long as our Multi-GM has a maximal BP gauge, which happens to be interior. Therefore, sequential BP-elimination of edges will be still equivalent to exact elimination, as long as the edge to be eliminated is normal, and the Multi-GM has a maximal BP gauge, which is interior, on each step of the elimination procedure. A formal statement is as follows.

Theorem VI.1 (BP-to-bouquet)**.** *Elimination of normal edges in Multi-GM via application of the sequential edge-by-edge BP-gauge procedure, resulting in Multi-GM for the bouquet graph, is exact, i.e. it is equivalent to the exact elimination via summation over binary variables associated with the eliminated edges.*

Notice that the sequential BP-elimination approach does not generalize to self-edges, because in this case expression (48) for $h^{(\alpha-bp)}$ returns a fractional function of $h^{(i,j)}$, thus resulting in a fractional (not polynomial) function over the remaining gauge variables, $x \setminus \{x_{\alpha_d}, x_{\bar\alpha_d}\}$. However, the general fractional relations (46,47,48) still results in a number of useful statements.

Theorem VI.2 (BP-saddle)**.** *Any BP gauge solution of Eq. (28) is a saddle-point of the Gauge Function, $z(x)$, defined in Eq. (27), over any pair of the edge-gauges, $x_{\alpha_d}, x_{\bar\alpha_d}$.*

Proof. We only need to discuss here the case of an interior gauge, extending it to the (generic) case of a BP gauge following the logic of Section III B. Expanded in the Taylor series over deviations from the BP-gauge, the rational expression, $h(x)/(1+x_{\alpha_d}x_{\bar\alpha_d})$, representing the $x_{\alpha_d}, x_{\bar\alpha_d}$-dependent part of $z(x)$, where $h(x)$ is from Eq. (45), becomes

$$\begin{aligned} h(x) - h^{(\alpha-\mathrm{bp})} - \text{ cubic corrections} &= d\left(x_{\alpha_d} - x^{(\mathrm{bp})}_{\alpha_d}\right)\left(x_{\bar\alpha_d} - x^{(\mathrm{bp})}_{\bar\alpha_d}\right) \\ &= \frac{d}{4}\left(\left(x_{\alpha_d} + x_{\bar\alpha_d} - x^{(\mathrm{bp})}_{\alpha_d} - x^{(\mathrm{bp})}_{\bar\alpha_d}\right)^2 - \left(x_{\alpha_d} - x_{\bar\alpha_d} - x^{(\mathrm{bp})}_{\alpha_d} + x^{(\mathrm{bp})}_{\bar\alpha_d}\right)^2\right), \end{aligned} \tag{50}$$

where $d > 0$ (we skip presenting here bulky but explicit expression for d), thus completing the proof. □

The following technical statement, proven through a straightforward algebraic manipulation, introduces another useful feature of Eq. (48).

Lemma VI.3 (BP- vs exact- reductions)**.** *Consider a generic polynomial, $h(x)$, representing Multi-GM and stated in the form of expansion (45) over variables x_{α_d} and $x_{\bar\alpha_d}$ associated with the edge, α. Then, condition that for all values of the variables remaining after contraction of the edge α_d, the BP-reduced function, defined according to Eq. (48), is less or equal then the exact-reduced function,*

$$\forall x^{(1)}: \quad \left.\frac{h(x)}{1+x_{\alpha_d}x_{\bar\alpha_d}}\right|_{x_{\alpha_d}=x^{(\alpha-\mathrm{bp})}_{\alpha_d},x_{\bar\alpha_d}=x^{(\alpha-bp)}_{\bar\alpha_d}} \le h^{(1,1)}(x^{(1)}) + h^{(0,0)}(x^{(1)}), \tag{51}$$

holds if

$$\forall x^{(1)}: \quad h^{(0,1)}(x^{(1)})h^{(1,0)}(x^{(1)}) \le h^{(0,0)}(x^{(1)})h^{(1,1)}(x^{(1)}). \tag{52}$$

The following optimization version of the Lemma VI.3 was also introduced (and proven) in [5, 67].

Lemma VI.4 (BP- vs exact- reductions: variational version)**.** *Given representation (45) of $h(x)$ as the polynomial in $x_{\alpha_d}, x_{\bar{\alpha}_d}$, an arbitrarily chosen marginal belief, $\beta_\alpha \in [0,1]$, and Eq. (52) satisfied, guarantees that*

$$\forall x^{(1)}: \quad (\beta_\alpha)^{\beta_\alpha}(1-\beta_\alpha)^{1-\beta_\alpha} \inf_{x_{\alpha_d}, x_{\bar{\alpha}_d}} \frac{h(x)}{\left(x_{\alpha_d} x_{\bar{\alpha}_d}\right)^{\beta_\alpha}} \leq h^{(1,1)}(x^{(1)}) + h^{(0,0)}(x^{(1)}). \tag{53}$$

VII. BI-STABILITY AND MONOTONICITY OF BP ELIMINATION

Remarkably the condition (52) was shown to hold generically if $h(x)$ is Bi-Stable [5], where the stability and bi-stability of a polynomial are defined as follows.

Definition VII.1 (Real Stable Polynomial and Bi-Stable Polynomial. See [5])**.** *A nonzero polynomial, $g(x) \in \mathbb{R}[x_1, \cdots, x_N]$, with real coefficients is Real Stable if none of its roots $z = (z_1, \cdots, z_N) \in \mathbb{C}^N$ (i.e., solutions of $g(z) = 0$) satisfies: $Im(z_i) > 0$ for every $i = 1, \cdots, N$. A polynomial $h(x_{\alpha_d^{(1)}}, x_{\bar{\alpha}_d^{(1)}}; x_{\alpha_d^{(2)}}, x_{\bar{\alpha}_d^{(2)}} \cdots)$ is Bi-Stable if $h(x_{\alpha_d^{(1)}}, -x_{\bar{\alpha}_d^{(1)}}; x_{\alpha_d^{(2)}}, -x_{\bar{\alpha}_d^{(2)}} \cdots)$.*

Therefore we arrive at the following powerful statement.

Theorem VII.2 (Monotonicity of Variational BP)**.** *Consider an Multi-GM over graph $\mathcal{G}$ and with the factors correspondent to a Bi-Stable polynomial, $h(x)$, build a sequence of Multi-GMs, $m = 0, \cdots, |\mathcal{E}|$, starting with the original Multi-GM and getting next Multi-GM in the sequence by contraction of an edge, and denote (according to notations of the preceding Sections), graph, vector of gauge variables, polynomial and Variational BP estimation for Partition Function evaluated at the m-th step of the hierarchy, $\mathcal{G}^{(m)}, x^{(m)}, h^{(m)}(x^{(m)})$ and $Z^{(k;\mathrm{vbp})}$, respectively. Then*

(1) Each polynomial, $h^{(m)}(x^{(m)})$, in the sequence is Bi-Stable.

(2) Value of the Variational BP estimation for Partition Function does not decrease with elimination and therefore

$$Z^{(\mathrm{vbp})} = Z^{(0;\mathrm{vbp})} \leq Z^{(1;\mathrm{vbp})} \ldots \leq Z^{(|\mathcal{E}|;\mathrm{vbp})} = Z. \tag{54}$$

Proof. The first step of exact contraction, applied to $h(x)$, consists of applying a differential operator $\left(1 + \partial_{x_{\alpha_d^{(1)}}} \partial_{x_{\bar{\alpha}_d^{(1)}}}\right)$, followed by setting both $x_{\alpha_d^{(1)}}$ and $x_{\bar{\alpha}_d^{(1)}}$ to zero. The composite operator preserves Bi-Stability. This follows from a standard argument that involves characterization of linear operators T that preserve real stability of polynomials in terms of their algebraic symbols [12, 13]. The algebraic symbol of the above composite operator is easily computed, and the stability of its symbol is obvious, therefore $h^{(1)}(x^{(1)})$ is Bi-Stable. Applying the logic sequentially, we conclude that all polynomials in the sequence: $m = 1, \cdots, |\mathcal{E}|, \quad h^{(m)}(x^{(m)})$ are Bi-Stable. Statement (1) of the Theorem VII.2 is proven.

Applying Lemma VI.4 to each elimination in the sequence one writes

$$\begin{aligned} &m = 1, \cdots, |\mathcal{E}| \quad \forall \beta^{(m-1)} \in [0,1], \quad \forall x^{(m)}: \\ &(\beta_{\alpha^{(m)}})^{\beta_{\alpha^{(m)}}}(1-\beta_{\alpha^{(m)}})^{1-\beta_{\alpha^{(m)}}} \inf_{x_{\alpha_d^{(m)}}, x_{\bar{\alpha}_d^{(m)}} > 0} \frac{h^{(m-1)}(x^{(m-1)})}{\left(x_{\alpha_d^{(m)}} x_{\bar{\alpha}_d^{(m)}}\right)^{\beta_{\alpha^{(m)}}}} \leq h^{(m)}(x^{(m)}). \end{aligned} \tag{55}$$

Next one multiplies both sides of Eq. (55) on,

$$\frac{\prod_{\alpha \in \mathcal{E}^{(m)}} (\beta_\alpha)^{\beta\alpha}(1-\beta_\alpha)^{1-\beta\alpha}}{\prod_{\alpha \in \mathcal{E}^{(m)}} (x_{\alpha_d} x_{\bar{\alpha}_d})^{\beta_\alpha}}, \tag{56}$$

and observe that $\inf_{x^{(m)}>0}$ applied to the left hand side of the resulting inequality is less or equal to the $\inf_{x^{(m)}>0}$ applied to the right hand side of the inequality. Finally, similar application of the $\max_{\beta^{(m-1)}}$ operation to the two sides of the inequality obtained at the previous step results in the desired Eq.(54). (2) is proven. □

Notice that related technical statements and proofs were reported in [5] and [67].

VIII. DISCUSSION AND PATH FORWARD

Inspired by [5, 67], we began this manuscript by generalizing the Bethe Free Energy approach from normal GM to multi-GM. Then we reformulate gauge representation of [21, 22] for computing Partition Function of an Multi-GM in terms of polynomials. According to [21, 22], picking up a Gauge Function, which is a term in the gauge-transformed series, and making it least sensitive to the gauge transformations (looking for stationary point of the Gauge Function over gauges) results in the BP gauge and subsequently in the Loop Series expression for the Partition Function, where each term is an explicit functional of the BP gauge. One may say that the algebraic essence of the Loop Series approach is in reconstructing exact Partition Function from its tractable BP approximation by summing the Loop Series terms. The main construct of this manuscript is an alternative map, suggested by analogy with the polynomial construct of [5, 33, 67] from the Gauge Function to the Partition Function, Z. Now, this is possible via a sequence of differentiation of the Gauge Function over gauge variables, each associated with a directed edge of the graph. We show that a differentiation step in the sequence can be interpreted graphically as contraction/elimination of an edge, which results in a new Multi-GM with one less edge and one less node. Partition Function of each Multi-GM in the sequence is exactly equal to Partition Function of the original Multi-GM. (Note in passing that (a) construction is similar to an elementary transformation step in the graph minor theory [26]; and (b) even if the original GM is normal, i.e. it contains only normal edges and no self-edges, one eventually arrives advancing in the sequence at an Multi-GM, containing self-edges, therefore justifying discussion of the most general Multi-GM setting.) Evaluating minimum of the Bethe Free Energy, or equivalently specially defined optimum of the respective Gauge Function, for each Multi-GM in the sequence we get an optimal BP estimation for each Multi-GM in the sequence. We observe that BP transformation is exact for contraction of a normal edge but approximate for contraction of a self-edge. Then, utilizing the power of the Real Stable Polynomials theory [58, 72, 75], we showed that (a) all polynomials associated with factors of the contracted Multi-GMs are Real Stable Polynomials if all polynomials associated with factors of the original Multi-GM are Real Stable Polynomials; (b) optimal BP estimation for Partition Function of an Multi-GM in the sequence upper bounds optimal BP estimation for Partition Function of the preceding Multi-GM (in the sequence). Corollary of the latter statement is a new proof (also generalization from GM to Multi-GM) that the optimal BP estimation of the original Multi-GM low bounds the exact Partition Function. The original proof was made for the special case of bi-partite GM in [67], when Real Stable Polynomials is reduced to Real Stable, while the polynomial version of our results is a particular case of the relation presented in [5].

Synthesis of the two approaches, GM/gauges/BP/loops and Real Stable Polynomials, is far from explored by this and preceding [5, 34, 67] manuscripts. Therefore, we find it useful to combine in the remainder of this section some remarks, that follow from the manuscript results, with speculations about future research directions.

- **Bi-Stable examples**: Linear (degree one) real polynomials, $a + \sum_i b_i z_i$, with $a > 0$ and $\forall i, b_i > 0$, correspondent to generic matching (monomer–dimer) models over bi-partite graphs, is the main example of a GM represented by Bi-Stable Polynomials/Real Stable Polynomials. A non-bi-partite Bi-Stable Polynomials example can be derived from the bi-partite case by contraction of (a number of) edges described in Section IV. Other known classes of BiStable Polynomialss are also to be explored in GMs. In particular, determinantal polynomials, $\det(B + \sum_i z_i A_i)$, with positive semi-definite matrices, $\forall i: \ A_i > 0$, and Hermitian matrix B (all matrices are quadratic of the same dimensionality) is another (and arguably the most popular example in the Real Stable Polynomials theory) that may also have interesting relations/consequences for Fermion GM of statistical and quantum physics, see [18, 19] and references therein. All statements made in this manuscript (e.g., on the ordering of the Partition Function estimates for the contracted sequence of Multi-GMs) would apply to the special Multi-GM with the underlying Bi-Stable Polynomials structure.

- **Improving BP approximation**: The elimination scheme of Section IV has a significant approximation potential, both theoretically and empirically. On the theoretical side, one may attempt to seek a more restrictive class of polynomials, for example, models which are Bi-Stable Polynomials locally and not globally in the upper-half planes for each complex variable (associated with a directed edge). Approached from an empirical/algorithmic stand point, the elimination can be carried over and then checked post-factum (if it results in an increase or decrease of the Partition Function of the contracted graphs). Given that the complexity of the contracted Multi-GM evaluations will be increasing exponentially with the elimination steps, one may consider approximate methods in the spirit of the mini-bucket elimination schemes [3, 4, 25, 49]. Therefore, developing new mini-bucket schemes based on the polynomial stability properties is another promising direction for the future. Besides, it will be important to take advantage of the polynomial structure in creating synthetic practical algorithms mixing BP/Gauge Transformation/Loop Calculus ideas with random sampling ideas; for example, in the spirit of Fully Polynomial Randomized Approximation Schemes and empirical schemes a-la [1], and the mini-bucket elimination schemes a-la [3, 4].

- **Efficient computation of BP gauge**: A comment in Section 3.3 of [67] suggests that some algorithmic improvements for computing $Z^{(bp)}$ based on techniques from the theory of stable polynomials are possible. In general, an Real Stable feature of the node polynomials does not guarantee convexity of the Bethe Free Energy (7), even though for some special cases and noticeably for the case of perfect matching [46, 47, 73], the convexity may be guaranteed. Moreover, an optimal solution of the Bethe Free Energy may be achieved at the boundary of the belief polytope, thus not satisfying the BP Eq. (30). Since results of this manuscript are dependent on the existence of a valid solution of BP Eq. (30), it is imperative for future progress to develop Real Stable Polynomials theory-based schemes answering the question of existence and discovering solution(s) of BP equations efficiently. It will also be important to generalize the analysis of this manuscript to the case when solution of the BP Eqs. (30) is found outside of the feasibility domain (outside of the BP polytope).

- **Higher alphabets and higher-degree polynomials**: Both the gauge transformation and the Real Stable Polynomials theory extend, in principle, to the case of higher alphabets and related higher-degree polynomials. The loop tower approach of [17] and alternative approach of [54] build generalizations of the Gauge Transformation and Loop Series for the case of higher (than binary) alphabets. We conjecture that choosing polynomial gauge parametrization for the non-binary cases, generalizing elimination/differentiation procedure such that it would result in a Multi-GM sequence with the desired non-decreasing BP estimates for the Partition Function, is possible.

- **Synthesis with Fully Polynomial Deterministic Algorithmic Schemes**: A number of Fully Polynomial Deterministic Algorithmic Schemes that apply ideas from the theory of graph polynomials to Partition Function were recently developed. Some of the most recent results have focused on estimating (a) Partition Function of cliques over graphs [7, 8]; (b) permanents of some complex matrices [8, 9]; (c) complex-valued graph polynomials on finite degree graphs, including Tutte polynomials, independence polynomials, as well as Partition Function of complex valued spin and edge coloring models [55]; and (d) Partition Function of attractive Ising models of bounded degree [48]. All of the manuscripts just mentioned rely on results of advance complex analysis initiated by studies of phase transitions for infinite systems in statistical physics [36, 45, 63, 82], which allowed zeros of graph polynomials to be located and satisfy certain properties. Polynomials considered in the studies are special but also different from Real Stable polynomials considered in [67] and in this manuscript. Besides, and as was emphasized in [67], there exist models that are not of the class explained by Real Stable Polynomials, notably attractive Ising models [61, 62, 68], even though their Partition Functions are bounded from below by respective BP estimates. (Note in passing that light may be shed on the relation between the two seemingly unrelated statements of BP validity as a lower bound via mapping of a general Ising model to a matching model suggested in [36].) It will be important to reconcile and unify these sister subjects. We conjecture that a combination of methods from the gauge and graph transformations, loop series, and analysis of Partition Function zeros in the complex domain of parameters will be imperative for making the progress towards unification of the existing approaches.

IX. ACKNOWLEDGEMENTS

MC and YM are grateful to organizers and participants of the EPFL, Bernoulli center, workshops on “Introduction to Partition Functions” in July of 2018 and "Applications of the Partition Functions" in November of 2018, where this work was initiated and where its first version was criticized, respectively. We are particularly indebted to Nisheeth Vishnoi for many discussions and useful explanations, to Peter Csikvari for attracting our attention to inconsistency in our early notes and to Nima Anari explaining to us the notion of bi-stability. We are also thankful to Pascal Vontobel, Jinwoo Shin, and Marc Lelarge for help with references and useful comments.

Appendix A: Loop Series [21, 22] restated in the polynomial form

With gauge x chosen to satisfy the BP Eqs. (28), or equivalently Eqs. (31), of Soft Multi-GM, thus denoted $x^{(bp)}$, consistently with notations introduced in the main part of the manuscript, one derives from Eq. (20) the Loop Series, expression for Z:

$$Z = \sum_{\sigma \in \Sigma_{glp}} z(\sigma | x^{(bp)}), \tag{A1}$$

where Σ_{glp} stands for the set of σ vectors corresponding to the so-called Generalized Loops , $\sigma \in \Sigma_{glp}$ iff $\forall a \in \mathcal{V}$, $\sum_{\alpha \in e(a)} \sigma_\alpha \neq 1$. Note that an empty set, $\sigma = 0^{|\mathcal{E}|}$ is included in Σ_{glp}. A Soft Multi-GM can also be thought of as a subgraph of $\mathcal{G}$, $\mathcal{G}^{(\sigma)} = (\mathcal{V}^{(\sigma)}, \mathcal{E}^{(\sigma)}) \subseteq \mathcal{G}$, constructed by coloring edges of the graph (setting respective σ_α to unity) according to the following rules: each node neighboring an edge of the Soft Multi-GM set contains at least two edges colored, i.e. $V^{(\sigma)} \doteq (a \in \mathcal{V} | \sum_{\alpha \in e(a)} \sigma_\alpha > 1)$ and $E^{(\sigma)} \doteq (\alpha \in \mathcal{E} | \sigma_\alpha = 1)$.

Each Soft Multi-GM contribution in Eq. (A1) is expressed via a BP solution as follows:

$$\forall \sigma \in \Sigma_{glp}: \quad z(\sigma|x^{(bp)}) = z(x^{(bp)}) \frac{\prod\limits_{a \in \mathcal{V}^{(\sigma)}} \mu_a^{(bp)}}{\prod\limits_{\alpha \in \mathcal{E}^{(\sigma)}} \beta_\alpha^{(bp)} (1 - \beta_\alpha^{(bp)})}, \tag{A2}$$

$$\forall a \in \mathcal{V}^{(\sigma)}: \quad \mu_a^{(bp)} \doteq \frac{\sum_{\varsigma_a} f_a(\varsigma_a) \prod_\alpha \left((x_\alpha^{(bp)})^{\varsigma_\alpha} \left(\varsigma_\alpha - \beta_\alpha^{(bp)} \right)^{\sigma_\alpha} \right)}{\sum_{\varsigma_a} f_a(\varsigma_a) \prod_{\alpha \in e^{(\sigma)}(a)} (x_\alpha^{(bp)})^{\varsigma_\alpha}}, \tag{A3}$$

$$\forall \alpha \in \mathcal{E}^{(\sigma)}: \quad \beta_\alpha^{(bp)} \doteq \frac{x_{\alpha_d} x_{\bar{\alpha}_d}}{1 + x_{\alpha_d} x_{\bar{\alpha}_d}}, \tag{A4}$$

where $e_d^{(\sigma)}(a)$ marks the set of edges of $\mathcal{E}^{(\sigma)}$ associated with the node a of $\mathcal{V}^{(\sigma)}$.

[1] S. Ahn, M. Chertkov, and J. Shin. Synthesis of MCMC and Belief Propagation. *Neural Information Processing Systems (NIPS)*, 2016.

[2] S. Ahn, M. Chertkov, and J. Shin. Gauging Variational Inference. *Neural Information Processing Systems (NIPS)*, 2017.

[3] S. Ahn, M. Chertkov, A. Weller, and J. Shin. Bucket Renormalization for Approximate Inference. *arXiv:1803.05104*, 2018.

[4] S. Ahn, M. Chertkov, J. Shin, and A. Weller. Gauged Mini-Bucket Elimination for Approximate Inference. *arXiv preprint arXiv:1801.01649*, 2018.

[5] N. Anari and S. Oveis Gharan. A Generalization of Permanent Inequalities and Applications in Counting and Optimization. *arXiv:1702.02937*, February 2017.

[6] F. Barahona. On the computational complexity of Ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241, 1982.

[7] A. Barvinok. Computing the partition function for cliques in a graph. *Theory of Computing*, 11(13):339–355, 2015.

[8] A. Barvinok. Combinatorics and complexity of partition functions. In *Algorithms and combinatorics*, 2016.

[9] A. Barvinok. Computing the permanent of (some) complex matrices. *Found. Comput. Math.*, 16(2):329–342, April 2016.

[10] M. Bayati, D. Shah, and M. Sharma. Max-product for maximum weight matching: Convergence, correctness, and lp duality. *IEEE Transactions on Information Theory*, 54(3):1241–1251, 2008. Proc. IEEE Int. Symp. Information Theory, 2006.

[11] H.A. Bethe. Statistical theory of superlattices. *Proceedings of Royal Society of London A*, 150:552, 1935.

[12] J. Borcea and P. Brändén. The Lee-Yang and Pólya-Schur Programs. II. Theory of Stable Polynomials and Applications. *arXiv:0809.3087*, September 2008.

[13] J. Borcea and P. Brändén. The Lee-Yang and Pólya-Schur programs. I. Linear operators preserving stability. *Inventiones Mathematicae, arXiv:0809.0401*, 177:541–569, March 2009.

[14] J. Borcea, P. Brändén, and T. M. Liggett. Negative dependence and the geometry of polynomials. *arXiv:0707.2340*, 2007.

[15] P. Brändén. Polynomials with the half-plane property and matroid theory. *arxiv:math/0605678*, 2006.

[16] V. Chandrasekaran, M. Chertkov, D. Gamarnik, D. Shah, and J. Shin. Counting independent sets using the Bethe approximation. *SIAM Journal on Discrete Mathematics*, 25(2):1012–1034, 2011.

[17] V. Y. Chernyak and M. Chertkov. Loop Calculus and belief propagation for q-ary alphabet: Loop tower. In *2007 IEEE International Symposium on Information Theory*, pages 316–320, June 2007.

[18] V. Y. Chernyak and M. Chertkov. Fermions and loops on graphs: I. Loop Calculus for determinants. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(12):P12011, 2008.

[19] V. Y. Chernyak and M. Chertkov. Fermions and loops on graphs: Ii. a monomer-dimer model as a series of determinants. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(12):P12012, 2008.

[20] V. Y. Chernyak and M. Chertkov. Planar graphical models which are easy. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(11):P11007, 2010.

[21] M. Chertkov and V. Y. Chernyak. Loop Calculus in statistical physics and information science. *Phys. Rev. E*, 73:065102, Jun 2006.

[22] M. Chertkov and V. Y. Chernyak. Loop series for discrete statistical models on graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(06):P06009, 2006.

[23] M. Chertkov, V. Y. Chernyak, and R. Teodorescu. Belief propagation and loop series on planar graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(05):P05003, 2008.

[24] M. Chertkov and A. B. Yedidia. Approximating the permanent with fractional belief propagation. *J. Mach. Learn. Res.*, 14(1):2029–2066, January 2013.

[25] R. Dechter and I. Rish. Mini-buckets: A general scheme for bounded inference. *Journal of the ACM (JACM)*, 50(2):107–153, 2003.

[26] E. D. Demaine, M. T. Hajiaghayi, and K. Kawarabayashi. Algorithmic graph minor theory: Decomposition, approximation, and coloring. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 637–646, Oct 2005.

[27] M. E. Fisher. On the dimer solution of planar Ising models. *Journal of Mathematical Physics*, 7(10):1776–1781, 1966.

[28] G. D. Forney. Codes on graphs: normal realizations. *IEEE Transactions on Information Theory*, 47(2):520–548, Feb 2001.

[29] R.G. Gallager. *Low Density Parity Check Codes*. MIT Press, Cambridge, MA, 1963.

[30] A. Galluccio, M. Loebl, and J. Vondrák. New algorithm for the Ising problem: partition function for finite lattice graphs. *Physical review letters*, 84:5924–7, 2000.

[31] J. W. Gibbs. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*. Cambridge Library Collection - Mathematics. Cambridge University Press, 1902; reprinted in 2010.

[32] V. Gómez, H.J. Kappen, and M. Chertkov. Approximate inference on planar graphs using loop Loop Calculus and belief propagation. *J. Mach. Learn. Res.*, 11:1273–1296, August 2010.

[33] L. Gurvits. A Polynomial-Time Algorithm to Approximate the Mixed Volume within a Simply Exponential Factor. *Discrete & Computational Geometry*, 41(4):533–555, 2009.

[34] L. Gurvits. Unharnessing the power of Schrijver's permanental inequality. *ArXiv e-prints*, June 2011.

[35] L. Gurvits and A. Samorodnitsky. Bounds on the permanent and some applications. In *Proceedings of the 2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, FOCS '14, pages 90–99, WashinGauge Transformationon, DC, USA, 2014. IEEE Computer Society.

[36] O.J. Heilmann and E.H. Lieb. Theory of monomer-dimer systems. *Communications in Mathematical Physics*, 25(3):190–232, Sep 1972.

[37] M. Jerrum and A. Sinclair. The Markov chain Monte Carlo method: an approach to approximate counting and integration. In *Approximation algorithms for NP-hard problems*, pages 482–520. PWS Publishing Co., 1996.

[38] M. R. Jerrum, L. G. Valiant, and V. V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, 43:169âĂŞ–188, 1986.

[39] J. K. Johnson. *Convex relaxation methods for graphical models: Lagrangian and maximum entropy approaches*. PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2008.

[40] P. W. Kasteleyn. Dimer statistics and phase transitions. *Journal of Mathematical Physics*, 4(2):287–293, 1963.

[41] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009.

[42] V. Kolmogorov, J. Thapper, and S. Zivny. The power of linear programming for general-valued CSPS. *SIAM Journal on Computing*, 44(1):1–36, 2015.

[43] V. Kolmogorov. A faster approximation algorithm for the Gibbs partition function. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 228–249. PMLR, 06–09 Jul 2018.

[44] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 03 1951.

[45] T. D. Lee and C. N. Yang. Statistical theory of equations of state and phase transitions. ii. lattice gas and ising model. *Phys. Rev.*, 87:410–419, Aug 1952.

[46] M. Lelarge. Loopy annealing belief propagation for vertex cover and matching: convergence, LP relaxation, correctness and Bethe approximation. *arXiv:1401.7923*, 2014.

[47] M. Lelarge. Counting matchings in irregular bi-partite graphs and random lifts. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2230–2237. Society for Industrial and Applied Mathematics, 2017.

[48] J. Liu, A. Sinclair, and P. Srivastava. The Ising partition function: zeros and deterministic approximation. *FOCS2017*, pages 986–997, 2017.

[49] Q. Liu and A. T. Ihler. Bounding the partition function using Hölder's inequality. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 849–856, 2011.

[50] H. A. Loeliger. An introduction to factor graphs. *IEEE Signal Processing Magazine*, 21(1):28–41, Jan 2004.

[51] D. J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, New York, NY, USA, 2002.

[52] M. Mezard and A. Montanari. *Information, Physics, and Computation*. Oxford University Press, Inc., New York, NY, USA, 2009.

[53] C. Moore and S. Mertens. *The Nature of Computation*. Oxford University Press, Inc., New York, NY, USA, 2011.

[54] R. Mori. Loop Calculus for nonbinary alphabets using concepts from information geometry. *IEEE Transactions on Information Theory*, 61(4):1887–1904, April 2015.

[55] V. Patel and G. ReGauge Transformations. Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials. *Electronic Notes in Discrete Mathematics*, 61:971 – 977, 2017. The European Conference on Combinatorics, Graph Theory and Applications (EUROCOMB'17).

[56] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[57] H.A. Peierls. Ising's model of ferromagnetism. *Proceedings of Cambridge Philosophical Society*, 32:477–481, 1936.

[58] R. Pemantle. Hyperbolicity and stable polynomials in combinatorics and probability. *Current Developments in Mathematics*, 2011(1), 2011.

[59] T. Richardson and R. Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008.

[60] A. Risteski. How to calculate partition functions using convex programming hierarchies: provable bounds for variational methods. In *Conference on Learning Theory*, pages 1402–1416, 2016.

[61] N. Ruozzi. The Bethe partition function of log-supermodular graphical models. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pages 117–125, USA, 2012. Curran Associates Inc.

[62] N. Ruozzi. A Lower Bound on the Partition Function of Attractive Graphical Models in the Continuous Case. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1048–1056, Fort Lauderdale, FL, USA, 20–22 Apr 2017. PMLR.

[63] A. D. Scott and A. D. Sokal. The repulsive lattice gas, the independent-set polynomial, and the lovász local lemma. *Journal of Statistical Physics*, 118(5):1151–1261, Mar 2005.

[64] J. Shin. The complexity of approximating a Bethe equilibrium. *IEEE Transactions on Information Theory*, 60(7):3959–3969, July 2014.

[65] A. Sly and N. Sun. Counting in two-spin models on d -regular graphs. *Ann. Probab.*, 42(6):2383–2416, 11 2014.

[66] D. Sontag. *Approximate Inference in Graphical Models using LP Relaxations*. PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2010.

[67] D. Straszak and N. K. Vishnoi. Belief propagation, Bethe approximation and polynomials. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 666–671, Oct 2017.

[68] E. B. Sudderth, M. J. Wainwright, and A. S. Willsky. Loop series and Bethe variational bounds in attractive graphical models. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, NIPS'07, pages 1425–1432, USA, 2007. Curran Associates Inc.

[69] L. G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8:189–201, 1979.

[70] L. G. Valiant. Holographic algorithms. *SIAM J. Comput.*, 37(5):1565–1594, February 2008.

[71] V. Vazirani. *Approximation Algorithms*. Springer, 2003.

[72] N.K. Vishnoi. Zeros of polynomials and their applications to theory: a primer. *FOCS 2013 Workshop on Zeros of Polynomials and their Applications to Theory*, 2013.

[73] P. O. Vontobel. The Bethe permanent of a nonnegative matrix. *IEEE Transactions on Information Theory*, 59:1866–1901, 2010.

[74] P. O. Vontobel. Counting in graph covers: A combinatorial characterization of the Bethe entropy function. *IEEE Trans. Inf. Theor.*, 59(9):6018–6048, September 2013.

[75] D. Wagner. Multivariate stable polynomials: theory and applications. *Bulletin of the American Mathematical Society*, 48(1):53–84, 2011.

[76] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.*, 1(1-2):1–305, January 2008.

[77] M.J. Wainwright, T.S. Jaakkola, and A.S. Willsky. Tree-based reparametrization framework for approximate estimation on graphs with cycles. *Information Theory, IEEE Transactions on*, 49(5):1120–1146, 2003.

[78] Y. Watanabe and M. Chertkov. Belief propagation and loop calculus for the permanent of a non-negative matrix. *Journal of Physics A: Mathematical and Theoretical*, 43(24):242002, 2010.

[79] D. Weitz. Counting independent sets up to the tree threshold. In *In Proceedings of the Annual ACM Symposium on Theory of Computing*, pages 140–149, 2006.

[80] M. Welling, A. E. Gelfand, and A. Ihler. A cluster-cumulant expansion at the fixed points of belief propagation. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, UAI'12, pages 883–892, ArlinGauge Transformationon, Virginia, United States, 2012. AUAI Press.

[81] A. S. Willsky, E. B. Sudderth, and M. J. Wainwright. Loop series and Bethe variational bounds in attractive graphical models. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 1425–1432. Curran Associates, Inc., 2008.

[82] C. N. Yang and T. D. Lee. Statistical theory of equations of state and phase transitions. i. theory of condensation. *Phys. Rev.*, 87:404–409, Aug 1952.

[83] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *Information Theory, IEEE Transactions on*, 51(7):2282–2312, 2005.