Approximating the Permanent with Fractional Belief Propagation

Michael Chertkov   CHERTKOV@LANL.GOV
Theory Division & Center for Nonlinear Studies
Los Alamos National Laboratory
Los Alamos, NM 87545, USA

Adam B. Yedidia[*]   ADAMY@MIT.EDU
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139, USA

Abstract

We discuss schemes for exact and approximate computations of permanents, and compare them with each other. Specifically, we analyze the belief propagation (BP) approach and its fractional belief propagation (FBP) generalization for computing the permanent of a non-negative matrix. Known bounds and Conjectures are verified in experiments, and some new theoretical relations, bounds and Conjectures are proposed. The fractional free energy (FFE) function is parameterized by a scalar parameter $\gamma \in [-1; 1]$, where $\gamma = -1$ corresponds to the BP limit and $\gamma = 1$ corresponds to the exclusion principle (but ignoring perfect matching constraints) mean-field (MF) limit. FFE shows monotonicity and continuity with respect to γ. For every non-negative matrix, we define its special value $\gamma_* \in [-1; 0]$ to be the γ for which the minimum of the γ-parameterized FFE function is equal to the permanent of the matrix, where the lower and upper bounds of the γ-interval corresponds to respective bounds for the permanent. Our experimental analysis suggests that the distribution of γ_* varies for different ensembles but γ_* always lies within the $[-1; -1/2]$ interval. Moreover, for all ensembles considered, the behavior of γ_* is highly distinctive, offering an empirical practical guidance for estimating permanents of non-negative matrices via the FFE approach.

Keywords: permanent, graphical models, belief propagation, exact and approximate algorithms, learning flows

1. Introduction

This work is motivated by computational challenges associated with learning stochastic flows from two consecutive snapshots/images of n identical particles immersed in a flow (Chertkov et al., 2008; Chertkov et al., 2010). The task of learning consists in maximizing the permanent of an $n \times n$ matrix, with elements constructed of probabilities for a particle in the first image to correspond to a particle in the second image, over the low-dimensional parametrization of the reconstructed flow. The permanents in this enabling application are nothing but a weighted number of perfect matchings relating particles in the two images.

    In this manuscript we continue the thread of Watanabe and Chertkov (2010) and focus on computations of positive permanents of non-negative matrices constructed from probabilities. The exact

---

[*]. Also at Theory Division & Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

computation of the permanent is difficult, that is, it is a problem of likely exponential complexity, with the fastest known general Algorithm for computing the permanent of a full $n \times n$ matrix based on the formula from Ryser (1963) requiring $O(n2^n)$ operations. In fact, the task of computing the permanent of a non-negative matrix was one of the first problems established to be in the #-P complexity class, and the task is also complete in the class (Valiant, 1979).

Therefore, recent efforts have mainly focused on developing approximate Algorithms. Three independent developments, associated with the mathematics of strict bounds, Monte-Carlo sampling, and graphical models, contributed to this field.

The focus of the mathematics of permanent approach was on establishing rigorous lower and upper bounds for permanents. Many significant results in this line of research are related to the Conjecture of van der Waerden (1926) that the minimum of the permanent over doubly stochastic matrices is $n!/n^n$, and it is only attained when all entries of the matrix are $1/n$. The Conjecture remained open for over 50 years before Falikman (1981) and Egorychev (1981) proved it. Recently, Gurvits (2008) found an alternative, surprisingly short and elegant proof that also allowed for a number of unexpected extensions. See, for example, the discussion of Laurent and Schrijver (2010).

A very significant breakthrough in the Monte-Carlo sampling was achieved with the invention of the fully polynomial randomized Algorithmic schemes (fpras) for the permanent problem (Jerrum et al., 2004): the permanent is approximated in polynomial time, provably with high probability and within an arbitrarily small relative error. The complexity of the fpras of Jerrum et al. (2004) is $O(n^{11})$ in the general case. Even though the scaling was improved to $O(n^4 \log n)$ in the case of very dense matrices (Huber and Law, 2008), the approach is still impractical for the majority of realistic applications.

Belief propagation (BP) heuristics applied to permanent showed surprisingly good performance (Chertkov et al., 2008; Huang and Jebara, 2009; Chertkov et al., 2010). The BP family of Algorithms, originally introduced in the context of error-correction codes by Gallager (1963), artificial intelligence by Pearl (1988), and related to some early theoretical work in statistical physics by Bethe (1935), and Peierls (1936) on tree graphs, can generally be stated for any GM according to Yedidia et al. (2005). The exactness of the BP on any tree, that is, on a graph without loops, suggests that the Algorithm can be an efficient heuristic for evaluating the partition function, or for finding a maximum likelihood (ML) solution of the GM defined on sparse graphs. However, in the general loopy cases, one would normally not expect BP to work very well, making the heuristic results of Chertkov et al. (2008), Huang and Jebara (2009) and Chertkov et al. (2010) somehow surprising, even though not completely unexpected in view of the existence of polynomially efficient Algorithms for the ML version of the problem (Kuhn, 1955; Bertsekas, 1992), which were shown by Bayati et al. (2008) to be equivalent to an iterative Algorithm of the BP type. This raises questions about understanding the performance of BP. To address this challenge (Watanabe and Chertkov, 2010) established a theoretical link between the exact permanent and its BP approximation. The permanent of the original non-negative matrix was expressed as a product of terms, including the BP-estimate and another permanent of an auxiliary matrix, $\beta. * (1 - \beta)$,[1] where β is the doubly stochastic matrix of the marginal probabilities of the links between the particles in the two images (edges in the underlying GM) calculated using the BP approach. (See Theorem 3.) The exact relation of Watanabe and Chertkov (2010) followed from the general loop calculus technique of Chertkov and Chernyak (2006a,b), but it also allowed a simple direct derivation. Combining this

---

1. Here and below we will follow Matlab notations for the component-wise operations on matrices, such as $A. * B$ for the component-wise, Hadamard, product of the matrices A and B.

exact relation with aforementioned results from the mathematics of permanents led to new lower and upper bounds for the original permanent. Moreover this link between the math side and the GM side gained a new level of prominence with the recent proof by Gurvits (2011) of the fact that the variational formulation of BP in terms of the Bethe free energy (BFE) function, discussed earlier by Chertkov et al. (2008), Chertkov et al. (2010) and Watanabe and Chertkov (2010), and shown to be convex by Vontobel (2013), gives a provable lower bound to the permanent. Remarkably, this proof of Gurvits was based on an inequality suggested earlier by Schrijver (1998) for the object naturally entering the exact, loop calculus based, BP formulas, $\text{perm}(\beta. * (1 - \beta))$.

This manuscript contributes two-fold, theoretically and experimentally, to the new synergy developing in the field. On the theory side, we generalize the BP approach to approximately computing permanents, suggesting replacing the BFE function by its FFE generalization in the general spirit of Wiegerinck and Heskes (2003), differing from the BFE function of Yedidia et al. (2005) in the entropy term, and then derive new exact relations between the original permanent and the results of the FFE-based approach (see Theorem 13). The new object, naturally appearing in the theory, is $\text{perm}(\beta. * (1 - \beta).^{-\gamma})$, where $\gamma \in [-1; 1]$. The case of $\gamma = -1$ corresponds to BP and the case of $\gamma = 1$ corresponds to the so-called exclusion principle (Fermi), but ignoring perfect matching constraints, mean field (MF) approximation discussed earlier by Chertkov et al. (2008). Using recent results from the "mathematics of permanents," in particular from Gurvits (2011), we show, that considered as an approximation, the FFE-based estimate of the permanent is a monotonic continuous function of the parameter γ with $\gamma = -1$ and $\gamma = 0$ setting, respectively, the lower bound (achievable on trees) and an upper bound. We also analyze existing and derive new lower and upper bounds. We adopt for our numerical experiments the so-called zero-suppressed binary decision diagrams (ZDDs) approach of Minato (1993) (see, for example, Knuth 2009), which outperforms Ryser's formula for realistic (sparsified) matrices, for exactly evaluating permanents, develop numerical schemes for efficiently evaluating the fractional generalizations of BP, test the aforementioned lower and upper bounds for different ensembles of matrices and study the special, matrix dependent, γ_*, which is defined to be the special γ for which the FFE-based estimate is equal to the permanent of the matrix.[2]

The material in the manuscript is organized as follows: the technical introduction, stating the computation of the permanent as a GM, is explained in Section 2 and Appendix A. The BP-based optimization formulations, approximate methods, iterative Algorithms and related exact formulas are discussed in Section 3 and Appendices B, C, D, E. Section 4 is devoted to permanental inequalities, discussing the special values of γ and Conjectures. Our numerical experiments are presented and discussed in Section 5 and Appendices F, G, H. We conclude and discuss the path forward in Section 6.

2. Technical Introduction

The permanent of a square matrix p, $p = (p_{ij} | i, j = 1, \ldots, n)$, is defined as

$$\text{perm}(p) = \sum_{s \in S_n} \prod_{i=1}^{n} p_{is(i)},$$

---

2. Note that a methodologically similar approach, of searching for the best/special FFE-based coefficient, was already discussed in the literature by Cseke and Heskes (2011) for a Gaussian BP example.
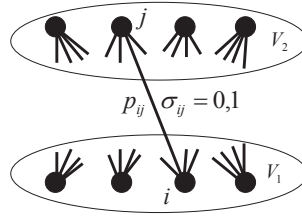
Figure 1: Illustration of graphical model for perfect matchings and permanent.

where S_n is the set of all permutations of the set, $\{1,\ldots,n\}$. Here and below we will only discuss permanents of non-negative matrices, that is, with $\forall i, j = 1,\ldots,n : p_{ij} \geq 0$. We also assume that $\mathrm{perm}(p) > 0$.

An example of a physics problem, where computations of permanents are important, is given by particle tracking experiments and measurements techniques, of the type discussed in Chertkov et al. (2008) and Chertkov et al. (2010). In this case, an element of the matrix, $p = (p_{ij}|i, j = 1,\ldots,n)$, is interpreted as an unnormalized probability that the particle labeled i in the first image moves to the position labeled j in the second image. In its most general formulation, the task of learning a low dimensional parametrization of the flow from two consecutive snapshots consists of maximizing the partition function $Z = \mathrm{perm}(p)$ over the "macroscopic" flow parameters affecting p. Computing the permanent for a given set of values of the parameters constitutes an important subtask, the one we are focusing on in this manuscript.

2.1 Computation of the Permanent as a Graphical Model Problem

The permanent of a matrix can be interpreted as the partition function Z of a graphical model (GM) defined over a bipartite undirected graph, $\mathcal{G} = (\mathcal{V} = (\mathcal{V}_1, \mathcal{V}_2), \mathcal{E})$, where $\mathcal{V}_1$, $\mathcal{V}_2$ are of equal size, $|\mathcal{V}_1| = |\mathcal{V}_2| = n$, and $\mathcal{V}_1$, $\mathcal{V}_2$, and $\mathcal{E}$ stand for the set of n vertices/labels of particles in the first and second images and the set of edges (possible relations) between particles in the two images, respectively. The basic binary variables, $\sigma_{ij} = 0, 1$, are associated with the edges, while the perfect matchings are enforced via the constraints associated with vertexes, $\forall i \in \mathcal{V}_1 : \quad \sum_{j \in V_2} \sigma_{ij} = 1$ and $\forall j \in \mathcal{V}_2 : \quad \sum_{i \in V_1} \sigma_{ij} = 1$, as illustrated in Figure 1. A non-negative element of the matrix, p_{ij}, turns into the weight associated with the edge (i, j). In summary, the GM relates the following probability to any of $n!$ perfect matchings, σ:

$$\rho(\sigma) = Z(p)^{-1} \prod_{(i,j) \in \mathcal{E}} (p_{ij})^{\sigma_{ij}}, \tag{1}$$

$$\sigma = \left(\sigma_{ij} = 0, 1 \,\middle|\, (i, j) \in \mathcal{E}; \forall i \in \mathcal{V}_1 : \sum_{j \in \mathcal{V}_2} \sigma_{ij} = 1; \forall j \in \mathcal{V}_2 : \sum_{i \in \mathcal{V}_1} \sigma_{ij} = 1 \right),$$

$$Z(p) = \mathrm{perm}(p) = \sum_{\sigma} \prod_{(i,j) \in \mathcal{E}} (p_{ij})^{\sigma_{ij}}. \tag{2}$$

The GM formulation (1) also suggests a variational, Kullback-Leibler (KL) scheme for computing the permanent. The only minimum of the so-called exact free energy (FE) function,

$$F(b|p) = \sum_{\sigma} b(\sigma) \log \left(\frac{b(\sigma)}{\prod_{(i,j) \in \mathcal{E}} (p_{ij})^{\sigma_{ij}}} \right),$$ (3)

computed over $b(\sigma) \geq 0$ for all σ under the normalization condition, $\sum_{\sigma} b(\sigma) = 1$, is achieved at $b(\sigma) = \rho(\sigma)$, and the value of the exact FE function at the minimum over $b(\sigma)$ is $-\log(Z(p))$. Here, the general and the optimal $b(\sigma)$ are interpreted as, respectively, the proxy and the probability of the perfect matching σ.

The relation between the problem of computing the permanent and the problem of finding the most probable (maximum) perfect matching is discussed in Appendix A.1.

2.2 Exact Methods for Computing Permanents

Computing the permanent of a matrix is a #-P hard problem, that is, it is a problem which most likely requires a number of operations exponential in the size of the matrix. In Appendix H, we experiment and compare the performance of the following two exact deterministic ways to evaluate permanents:

- A general method based on zero-suppressed binary decision diagrams (ZDDs), explained in more detail in Knuth (2009). See also detailed explanations below in Appendix G. As argued in Knuth (2009), the ZDD approach may be a very efficient practical tool for computing partition functions in general graphical models. This thesis was illustrated by Yedidia (2009) on the example of counting independent sets and kernels of graphs.

- A permanent-specific method based on Ryser's formula:

$$Z(p) = (-1)^n \sum_{S \subseteq \{1,...,n\}} (-1)^{|S|} \prod_{i=1}^{n} \sum_{j \in S} p_{ij}.$$

We use code from TheCodeProject for implementing the Ryser's formula.

Note that in most practical cases many entries of p are very small and they do not affect the permanent of p significantly. These entries do, however, take computational resources if accounted for in the Algorithm. To make computations efficient we sparsify the resulting matrix p, implementing the heuristic pruning technique explained in Appendix F.

We also verify some of our results against randomized computations of the permanent using the FPRAS from Jerrum et al. (2004), with a specific implementation from Chertkov et al. (2008).

3. Approximate Methods and Exact Relations

We perform an approximate computation of the permanent by following the general BFE approach of Yedidia et al. (2005) and the associated belief propagation/Bethe-Peierls (BP) Algorithm, discussed in detail for the case of permanents of non-negative matrices in Chertkov et al. (2008), Huang and Jebara (2009) and Chertkov et al. (2010). See also Appendix B reproducing the description of Chertkov et al. (2008) and Chertkov et al. (2010) and presented in this manuscript for convenience.

In our BP experiments we implement the Algorithm discussed by Chertkov et al. (2008) with a special type of initialization corresponding to the best perfect matching of p. We also generalize the BP scheme by modifying the entropy term in the BFE.

In the following Subsections we re-introduce the BFE approach, the related but different MF FE approach, and also consider a fractional FE approach generalizing and interpolating between Bethe/BP and MF approaches. Even though these optimization approaches and respective Algorithms can be thought of as approximating the permanent we will show that they also generate some exact relations for the permanent.

3.1 Belief Propagation/Bethe-Peierls Approach

Let us start by defining some useful notation.

Definition 1 (β-polytope) *Call the β-polytope of the non-negative matrix p (or just β-polytope for short) the set of doubly stochastic non-negative matrices with elements corresponding to zero elements of p equal to zero, that is, $\mathcal{B}_p = (\beta_{ij} | \forall i : \sum_{(i,j) \in \mathcal{E}} \beta_{ij} = 1; \forall j : \sum_{(i,j) \in \mathcal{E}} \beta_{ij} = 1; \forall (i,j) \text{ with } p_{ij} = 0 \ \beta_{ij} = 0 \text{ holds})$. We say that β lies in the interior of the β-polytope, $\beta \in \mathcal{B}_p^{(int)}$, if $\forall (i,j)$ with $p_{ij} \neq 0 \ \beta_{ij} \neq 0, 1$ holds.*[3]

In English, the interior solution means that all elements of the doubly stochastic matrices β are non-integer, under exception of the case when $p_{ij} = 0$ and, respectively, $\beta_{ij} = 0$.

Definition 2 (Bethe free energy for the permanent) *The following function of $\beta \in \mathcal{B}_p$*

$$F_{BP}(\beta|p) = \sum_{(i,j)} (\beta_{ij} \log(\beta_{ij}/p_{ij}) - (1-\beta_{ij}) \log(1-\beta_{ij})), \tag{4}$$

conditioned to a given p, is called the Bethe free energy (BFE) or the belief-Propagation/Bethe-Peierls (BP) function (for the permanent).[4]

To motivate the definition above let us briefly discuss the concept of the Bethe FE which was introduced in Yedidia et al. (2005) for the case of a general pair-wise interaction GM (with variables associated with vertices of the graph). Schematically, the logic extended to the case with variables associated with edges of the graph and leading to Equation (4) for the permanent is as follows. (See Watanabe and Chertkov 2010 for a detailed discussion.) Consider a GM with binary variables associated with edges of the graph. If the graph is a tree, then the following exact relation holds, $\rho(\sigma) = \prod_i \rho_i(\sigma_i) / \prod_{(i,j)} \rho_{ij}(\sigma_{ij})$, where $\sigma_i = (\sigma_{ij} = 0, 1|(i,j) \in \mathcal{E})$. Here, $\rho_i(\sigma_i)$ and $\rho_{ij}(\sigma_{ij})$ are marginal probabilities associated, respectively, with vertex i and edge (i,j) of the graph. Replacing the probabilities by their proxies/beliefs, $\rho(\sigma) \to b(\sigma)$, $\rho_i(\sigma_i) \to b_i(\sigma_i)$ and $\rho_{ij}(\sigma_{ij}) \to b_{ij}(\sigma_{ij})$, substituting the ratio of probabilities expression for $b(\sigma)$ in the exact FE function (3), and accounting for relations between the marginal beliefs, one arrives at the general expression for the Bethe FE function. This expression for the Bethe FE function is exact on a tree only, and it is similar in spirit to the one introduced in Yedidia et al. (2005) as an approximation for GM on a graph with loops.

---

3. What we call here "interior" would be mathematically more accurate to call "relative interior".
4. In the following, and whenever Bethe, MF, or fractional FE are mentioned, we will drop the clarifying—for the permanent—as only permanents are discussed in this manuscript.

When the graph is bi-partite with an equal number of nodes in the two parts the BP replacement for $b(\sigma)$ becomes

$$b_{BP}(\sigma) = \frac{\prod_{(i,j)\in\mathcal{E}:\sigma_{ij}=1}\beta_{ij}}{\prod_{(i,j)\in\mathcal{E}:\sigma_{ij}=0}(1-\beta_{ij})}, \tag{5}$$

where $\beta_{ij} = b_{ij}(1)$ is the marginal belief corresponding to finding the edge (i,j) in the matching. Then, substituting $b(\sigma)$ by $b_{BP}(\sigma)$ in Equation (3) results in the Bethe FE expression (4) for the perfect matchings (permanents). Note, that while the exact FE (3) is the sum of $O(n!)$ terms, there are only $O(n^2)$ terms in the Bethe FE (4).

According to the Loop Calculus approach of Chertkov and Chernyak (2006a,b), extended to the case of the permanent in Chertkov et al. (2008), Chertkov et al. (2010) and Watanabe and Chertkov (2010), the BP expression and the permanent are related to each other as follows:

Theorem 3 (Permanent and BP) *If the BP Equations following from minimization of the the BFE (4) over the doubly stochastic matrix β, that is,*

$$\forall(i,j): \quad (1-\beta_{ij})\beta_{ij} = \frac{p_{ij}}{u_i u^j}, \tag{6}$$

where $\log(u_i)$ and $\log(u^j)$ are positive-valued Lagrangian multipliers conjugated to $\sum_{j\in\mathcal{V}_2}\beta_{ij} = 1$ and $\sum_{i\in\mathcal{V}_1}\beta_{ij} = 1$, respectively, have a solution in the interior of the β-polytope, $\beta \in \mathcal{B}_p^{(int)}$, then

$$Z = Z_{BP}(p)\mathrm{perm}(\beta.*(1-\beta))\frac{1}{\prod_{i,j}(1-\beta_{ij})}, \tag{7}$$

where $Z_{BP}(p) = -\log(F_{BP}(\beta|p))$.

The proof of the Theorem 3 also appears in Appendix B.1. An iterative heuristic Algorithm solving BP Equations (6) for the doubly stochastic matrix β efficiently is discussed in Appendix B.2.

Let us recall that the (i,j) element of the doubly stochastic matrix β, β_{ij}, is interpreted as the proxy (approximation) to the marginal probability for the (i,j) edge of the bipartite graph $\mathcal{G}$ to be in a perfect matching, that is, β_{ij}, should be thought of as an approximation for $\rho_{ij} = \sum_{\sigma:\sigma_{ij}=1}\rho(\sigma)$.

Note also that $\log(u_i)$ and $\log(u^j)$ in Equations (6) are the Lagrange multipliers related to the $2n$ double stochasticity (equality) constraints on β.

3.1.1 BP AS THE MINIMUM OF THE BETHE FREE ENERGY

Definition 4 (Optimal Bethe free energy) *We define optimal BFE, $F_{o-BP}(p)$, and respective counting factor, $Z_{o-BP}(p)$, according to*

$$-\log(Z_{o-BP}(p)) = F_{o-BP}(p) = \min_{\beta\in\mathcal{B}} F_{BP}(\beta|p), \tag{8}$$

where $F_{BP}(\beta|p)$ is defined in Equation (4).

Considered in the general spirit of Yedidia et al. (2005), $F_{o-BP}(p)$, just defined, should be understood as an approximation to $-\log(\mathrm{perm}(p))$. To derive Equation (8) one needs to replace $b(\sigma)$ by (5). See Watanabe and Chertkov (2010) for more details.

The relation between the optimization formulation (8) and the BP Equations (6) requires some clarifications stated below in terms of the following two Propositions.

Proposition 5 (Partially resolved BP solutions) *Any doubly stochastic matrix* β *solving Equations (6) and lying on the boundary of the* β *polytope, that is,* β ∈ $\mathcal{B}_p$ *but* β ∉ $\mathcal{B}_p^{(int)}$, *can be reduced by permutations of rows and columns of* β *(and p, respectively) to a block diagonal matrix, with one block consisting of* 0, 1 *elements only and corresponding to a partial perfect matching, and the other block having all elements strictly smaller than unity, and nonzero if the respective* $p_{ij} \neq 0$. *We call such a solution of the BP Equations (6) partially resolved solutions, emphasizing that a part of the solution forms a partial perfect matching, and any other perfect matching over this subset is excluded by the solution (in view of the probabilistic interpretation of* β). *A doubly stochastic matrix* β *corresponding to a full perfect matching is called a fully resolved solution of the BP Equations (6).*

Proof This statement follows directly from the double stochasticity of β and from the form of the BP Equations (6), and it was already discussed in Chertkov et al. (2008) and Watanabe and Chertkov (2010) for the fully resolved case. ∎

Proposition 6 (Optimal Bethe FE and BP equations) *The optimal Bethe FE,* $F_{o-BP}(\beta)$ *over* β ∈ $\mathcal{B}_p$, *can only be achieved at a solution of the BP Equations (6), possibly with the Lagrange multipliers* u_i, u^j *taking the value* +∞.

Proof This statement is an immediate consequence of the fact that Proposition 5 is valid for any *p*, and so a continuous change in *p* (capable of covering all possible achievable *p*) can only result in an interior solution for the doubly stochastic matrix β merging into a vertex of the β-polytope, or emerging from the vertex (then respective Lagrangian multipliers take the value +∞), but never reaching an edge of the polytope at any other location but a vertex. Therefore, we can exclude the possibility of achieving the minimum of the Bethe FE anywhere but at an interior solution, partially resolved solution or a fully resolved solution (corresponding to a perfect matching) of the BP equations. ∎

Note, that an example where the minimum in Equation (8) is achieved at the boundary of the β − *polytope* (in fact, at the most probable perfect matching corner of the polytope) was discussed in Watanabe and Chertkov (2010).

Another useful and related statement, made recently in Vontobel (2013), is

Proposition 7 (Convexity of the Bethe FE) *The Bethe FE (4) is a convex function of* β ∈ $\mathcal{B}_p$.

A few remarks are in order. First, the statement above is nontrivial as, considered naively, individual edge contributions in Equation (4) associated with the entropy term, $\beta_{ij} \log \beta_{ij} - (1 - \beta_{ij}) \log(1 - \beta_{ij})$, are not convex for $\beta_{ij} > 1/2$, and the convexity is restored only due to the global (double stochasticity) condition. Second, the convexity means that if the optimal solution is not achieved at the boundary of $\mathcal{B}_p$, then either the solution is unique (general case) or the situation is degenerate and one finds a continuous family of solutions all giving the same value of the Bethe FE. The degeneracy means that *p* should be fine tuned to get into the situation, and addition of almost any small (random) perturbation to *p* would remove the degeneracy. To illustrate how the degeneracy may occur, consider an example of a (2×2) matrix *p* with all elements equal to each other. We first

observe that regardless of p for $n = 2$, the entropy contributions to the Bethe FE are identical to zero for any doubly stochastic (2×2) matrix, $\sum_{(i,j)}^{i,j=1,2}(\beta_{ij}\log\beta_{ij} - (1-\beta_{ij})\log(1-\beta_{ij})) = 0$. Moreover, the remaining, linear in β, contribution to the Bethe FE (which is also called the self-energy in physics) turns into a constant for the special choice of p. Thus one finds that in this degenerate $n = 2$ case,

$$\beta = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix},$$

with any $\alpha \in [0;1]$, is a solution of Equations (6) also achieving the minimum of the Bethe FE. Creating any asymmetry between the four components of the (2×2) p will remove the degeneracy, moving the solution of Equations (6) achieving the minimum of the Bethe FE to one of the two perfect matchings, corresponding to $\alpha = 0$ and $\alpha = 1$, respectively. It is clear that this special "double" degeneracy (first, cancellation of the entropy contribution, and then constancy of the self-energy term) will not appear at all if the doubly stochastic matrix β, solving Equations (6) in the $n > 2$ case has more than two nonzero components in every row and column. Combined with Proposition 7, this observation translates into the following statement.

Corollary 8 (Uniqueness of interior BP solution) *If $n > 2$ and an interior, $\beta \in \mathcal{B}_p^{(int)}$, solution of Equations (6) has more than two nonzero elements in every row and column, then the solution is unique*

In the following, discussing an interior BP solution, $\beta \in \mathcal{B}_p^{(int)}$ and aiming to focus only on the interesting/nontrivial cases, we will be assuming that $n > 2$ and p has more than two nonzero elements in every row and column.

3.2 Mean-Field Approach

Definition 9 (Mean field free energy) *For $\beta \in \mathcal{B}_p$, the MF FE is defined as*

$$F_{MF}(\beta|p) = \sum_{(i,j)}(\beta_{ij}\log(\beta_{ij}/p_{ij}) + (1-\beta_{ij})\log(1-\beta_{ij})). \tag{9}$$

Let us precede discussion of the MF notion/approach by a historical and also motivational remark. MF is normally thought of as an approximation ignoring correlations between variables. Then, the joint distribution function of σ is expressed in terms of the product of marginal distributions of its components. In our case of the perfect matching GM over the bi-partite graph, the MF approximation constitutes the following substitution for the exact beliefs,

$$b(\sigma) \to \prod_{(i,j)\in\mathcal{E}} b_{ij}(\sigma_{ij}), \tag{10}$$

into Equation (3). Making the substitution and relating the marginal edge beliefs to β according to, $\forall(i,j) \in \mathcal{E}: b_{ij}(1) = \beta_{ij}, b_{ij}(0) = 1 - \beta_{ij}$, one arrives at Equation (9). Because of how the perfect matching problem is defined, the two states of an individual variable, $\sigma_{ij} = 0$ and $\sigma_{ij} = 1$, are in the exclusion relation, and so one can also associate the special form of Equation (9) with the exclusion or Fermi- (for Fermi-statistics of physics) principle.

Direct examination of Equation (9) reveals that

Proposition 10 (MF FE minimum is always in the interior) $F_{MF}(\beta|p)$ *is strictly convex and its minimum is achieved at* $\beta \in \mathcal{B}_p^{(int)}$.

Looking for the interior minimum of Equation (9) one arrives at the following MF equations for the (only) stationary point of the MF FE function

$$\forall (i,j) \in \mathcal{E}: \quad \beta_{ij} = \frac{1}{1 + v_i v^j / p_{ij}},$$

where $\log(v_i)$ and $\log(v^j)$ are Lagrangian multipliers enforcing the conditions, $\sum_j \beta_{ij} = 1$ and $\sum_i \beta_{ij} = 1$, respectively. The equations can also be rewritten as

$$\forall (i,j): \quad \frac{\beta_{ij}}{1 - \beta_{ij}} = \frac{p_{ij}}{v_i v^j}, \tag{11}$$

making comparison with the respective BP Equations (6) transparent. An efficient heuristic for solving the MF equations (11) is discussed in Appendix C.2.

Direct examination shows that (unlike in the BP case) β with a single element equal to unity or zero (when the respective p element is nonzero) cannot be a solution of the MF Equations (11) over doubly stochastic matrix β—fully consistently with the Proposition 10 above. Moreover, $-\log(Z_{o-MF}(p))$, defined as the minimum of the MF FE (9), is simply equal to $-\log(Z_{MF}(p)$, defined as $F_{MF}(\beta)$ evaluated at the (only) doubly stochastic matrix solution of Equation (11).

Note also that the MF function (9) cannot be considered as a variational proxy for the permanent, bounding its value from below. This is because the substitution on the right-hand side of Equation (10) does not respect the perfect matching constraints, assumed reinforced on the left-hand side of Equation (10). In particular, the probability distribution function on the right-hand side of Equation (10) allows two edges of the graph adjacent to the same vertex to be in the active, $\sigma_{ij} = 1$, state simultaneously. However, this state is obviously prohibited by the original probability distribution, on the left-hand side of Equation (10) defined only over $n!$ of states corresponding to the perfect matchings. As shown below in Section 4.2, the fact that the MF ignores the perfect matching constraints results in the estimate $Z_{MF}(p)$ upper bounding perm(p), contrary to what a standard MF (not violating any original constraints) would do.

Finally and most importantly (for the MF discussion of this manuscript), the MF approximation for the permanent, Z_{MF}, can be related to the permanent itself as follows:

Theorem 11 (Permanent and MF)

$$Z(p) = \text{perm}(p) = Z_{MF}(p)\text{perm}(\beta./(1-\beta)) \prod_{(i,j) \in \mathcal{E}} (1 - \beta_{ij}), \tag{12}$$

where β *is the only interior minimum of (9).*

The proof of this statement is given in Appendix C.1.

3.3 FFE-based Approach

Similarity between the exact BP expression (7) and the exact MF expression (12) suggests that the two formulas are the limiting instances of a more general relation. We define

Definition 12 (Fractional free energy) *For* $\beta \in \mathcal{B}_p$, *the FFE is defined as*

$$F_f^{(\gamma)}(\beta|p) = -\log(Z_f^{(\gamma)}(\beta|p)) = \sum_{(i,j)} \left(\beta_{ij} \log(\beta_{ij}/p_{ij}) + \gamma(1-\beta_{ij}) \log(1-\beta_{ij})\right). \qquad (13)$$

Then, one finds that

Theorem 13 (FFE-based representation for permanent) *For any non-negative p and doubly stochastic matrix* β *which solves*

$$\forall(i,j): \quad \frac{\beta_{ij}}{(1-\beta_{ij})^{\gamma}} = \frac{p_{ij}}{w_i w^j}, \qquad (14)$$

for $\gamma \in [-1;1]$, *and if the solution found is in the interior of the domain, that is,* $\beta \in \mathcal{B}_p^{(int)}$, *the following relation holds*

$$\mathrm{perm}(p) = Z_f^{(\gamma)}(\beta|p)\mathrm{perm}\left(\frac{\beta.}{(1-\beta).^{\gamma}}\right) \prod_{(i,j)} (1-\beta_{ij})^{\gamma}, \qquad (15)$$

where FFE was defined above in Equation (13) and $\log(w_i)$ *and* $\log(w^j)$ *in Equation (14) are the Lagrangian multipliers enforcing the conditions,* $\sum_j \beta_{ij} = 1$ *and* $\sum_i \beta_{ij} = 1$, *respectively.*

The proof of Equation (13) is given in Appendix D.1. An iterative heuristic Algorithm solving Equations (14) efficiently is described in Appendix D.2.

Following the general GM logic and terminology introduced in Wiegerinck and Heskes (2003) and Yedidia et al. (2005), we call $F_f^{(\gamma)}(\beta|p)$ the fractional FE. Obviously the two extremes of $\gamma = -1$ and $\gamma = 1$ correspond to BP and MF limits, respectively. Many features of the BP and MF approaches extend naturally to the FFE-based case. In particular, one arrives at the following statement which appears in Theorem 60 of Vontobel (2013).

Proposition 14 (Fractional convexity) *The FFE-based function defined in Equation (13),* $F_f^{(\gamma)}(\beta|p)$, *is a convex function, convex over* $\beta \in \mathcal{B}_p$ *for any* $\gamma \in [-1;1]$ *and any non-negative p.*

Obviously, this statement generalizes Proposition 7. Also, the following statement becomes a direct consequence of Proposition 14:

Corollary 15 (Uniqueness of the interior fractional minimum) *If the minimum of* $F_f^{(\gamma)}(\beta|p)$ *is realized at* $\beta \in \mathcal{B}_p^{(int)}$, *it is unique.*

3.4 Minimal FFE-based Solution

It is clear that at $\gamma > 0$ the FFE-based Equations (14) cannot have a perfect matching solution, thus suggesting that at least in this case the solution, if it exists, is in the interior, $\beta \in \mathcal{B}_p^{(int)}$. On the other hand general existence (for any p) of such a solution follows immediately from the existence in a special case, for example of p with all elements equal, and then from the continuity of the Equations (14) solution with respect to p.

The case of $\gamma \in]-1;0]$ is a bit trickier. In this case, Equations (14) formally do allow a perfect matching solution. However, for all but degenerate p, that is, one reducible by permutations to a

diagonal matrix, the perfect matching solution is an isolated point. Indeed, let us consider a vicinity of a degenerate p. If one picks (without loss of generality) a diagonal, $p^{(0)} = (a_i \delta_{ij} | (i, j) \in \mathcal{E})$, and consider $p = p^{(0)} + \delta p$, where δp is a small positive matrix, then one observes that Equations (14) do allow a solution, $\beta = 1 + \varepsilon b$, where ε is a small positive scalar and $b = (b_{ij} | (i, j) \in \mathcal{E}; \forall i \in \mathcal{V}_1 : \sum_j b_{ij} = 0; \ \forall j \in \mathcal{V}_2 : \sum_i b_{ij} = 0)$ is a matrix with $O(1)$ elements, if the following scaling relation holds, $|\delta| \sim \varepsilon^{1+\gamma}$. Moreover, one also finds that a solution β is ε-close to a perfect matching only if p is $\varepsilon^{1+\gamma}$-close to a diagonal matrix. Now we apply the same continuity and existence arguments, as used above in the $\gamma > 0$ case arguments, to find out that the following statement holds.

Proposition 16 (FFE-based Minima) *The minimal FFE-based solution, defined by*

$$-\log(Z^{(\gamma)}_{o-f}(p)) = F^{(\gamma)}_{o-f}(p) = \min_{\beta \in \mathcal{B}_p} F^{(\gamma)}_f(\beta|p),$$

can only be achieved for $\gamma > -1$ and general (non-degenerate) p at $\beta \in \mathcal{B}^{(int)}_p$.

Then the following statement follows.

Proposition 17 (γ-monotonicity and continuity) *For any non-negative p, $Z^{\gamma}_{o-f}(p)$ is a monotonically increasing and continuous function of γ in $[-1; 1]$.*

Proof Observe that for any non-negative p and doubly stochastic matrix β, $\sum_{(i,j) \in \mathcal{E}} (1 - \beta_{ij}) \log(1 - \beta_{ij}) < 0$, so for any $\gamma_{1,2} \in [-1; 1]$ such that $\gamma_1 > \gamma_2$, $F^{(\gamma_1)}_f(\beta|p) \leq F^{(\gamma_2)}_f(\beta|p)$. Then according to the definition of $F^{(\gamma)}_{o-f}(p)$, $F^{(\gamma_1)}_{o-f}(p) \leq F^{(\gamma_1)}_f(\beta|p) \leq F^{(\gamma_2)}_f(\beta|p)$, for any doubly stochastic matrix β, in particular for β which is optimal for γ_2. Finally, $F^{(\gamma_1)}_{o-f}(p) \leq F^{(\gamma_2)}_{o-f}(p)$, proving monotonicity. The continuity of $Z^{(\gamma)}_{o-f}(p)$ with respect to γ in $]-1; 1]$ follows from Proposition 16 combined with $F^{(\gamma)}_f(\beta|p)$ continuity with respect to both $\gamma \in [-1; 1]$ and $\beta \in \mathcal{B}^{(int)}_p$. (The intuition with respect to the continuity is as follows: an increase in γ pushes the optimal β away from the boundary of the $\mathcal{B}_p$ polytope.) ∎

4. Permanent Inequalities, Special Value of γ, and Conjectures

We start this Section discussing in Subsection 4.1 the recently derived permanent inequalities related to BP and MF analysis. Then, we switch to describing new results of this manuscript in Subsection 4.2, which are mainly related to the FFE-based generalizations of the inequalities discussed in Subsection 4.1. We also discuss in Subsection 4.2 the special (and p-dependent) value of the fractional coefficient γ for which $\mathrm{perm}(p)$ is equal to $Z^{\gamma}_{o-f}(p)$. Finally, Subsection 4.3 is devoted to discussing Conjectures whose resolutions should help to tighten bounds for the permanent.

4.1 Recently Derived Inequalities

In this Subsection we discuss a number of upper and lower bounds on permanents of positive matrices introduced recently. Our task is two-fold. First, we wish to relate the bounds/inequalities to the BP and MF approaches introduced and discussed in the preceding Section. Some of these relations and interpretations are new. However, we also aim to test these bounds, and specifically to

characterize the tightness of the bounds by testing the gap as a function of advection and diffusion parameters in the 2d diffusion+advection model in Section 5.

The first bound of interest is

Proposition 18 (BP lower bound) *For any non-negative p*

$$\text{perm}(p) \geq Z_{o-BP}(p). \tag{16}$$

This statement, as an experimental but unproven observation, was made in Chertkov et al. (2008). It was stated as a Theorem (theorem # 14) in Vontobel (2010), but the proof was not provided. See also discussion in Vontobel (2013) following Theorem 49/Corollary 50. The statement was proven in Gurvits (2011). Interpreted in terms of the terminology and logic introduced in the preceding Section, the proof of Gurvits (2011) consists (roughly) in combining the inequality by (Schrijver, 1998)

$$\text{perm}(\beta.*(1-\beta)) \geq \prod_{(i,j)}(1-\beta_{ij}), \tag{17}$$

stated for any doubly stochastic matrix β, with some (gauge) manipulations/transformations of the type discussed above in Sections 3.1.1. We give our version of the proof (similar to the one in Gurvits (2011) in spirit, but somewhat different in details) in Appendix E. One direct Corollary of the bound (16) discussed in Gurvits (2011), is that

Corollary 19 *For an arbitrary doubly stochastic matrix ϕ*

$$\text{perm}(\phi) \geq Z_{o-BP}(\phi) \geq \prod_{(i,j)}(1-\phi_{ij})^{1-\phi_{ij}}. \tag{18}$$

Next, the following two lower bounds follow from analysis of Equation (7).

Proposition 20 (BP lower bound #1) *For any non-negative p and doubly stochastic matrix $\beta \in \mathcal{B}_p^{(int)}$ solving Equations (6) (if the solution exists) results in*

$$\text{perm}(p) \geq Z_{BP}(\beta|p)\prod_{(i,j)}(1-\beta_{ij})^{\beta_{ij}-1}\frac{n!}{n^n}.$$

This is the statement of Corollary 7 of Watanabe and Chertkov (2010) valid for any interior point solution of the BP-equations, and it follows from the Gurvits-van der Waerden Theorem of Gurvits (2008) and Laurent and Schrijver (2010), also stated as Theorem 6 in Watanabe and Chertkov (2010).

Proposition 21 (BP lower bound #2) *For any non-negative p and $\beta \in \mathcal{B}_p^{(int)}$ solving Equations (6) (if the solution exists) results in*

$$Z \geq 2Z_{BP}(\beta|p)(\prod_{i,j}(1-\beta_{ij}))^{-1}\prod_{i}\beta_{i\Pi(i)}(1-\beta_{i\Pi(i)}),$$

where Π is an arbitrary permutation.

This statement was made in Theorem 8 in Watanabe and Chertkov (2010) and it is also related to an earlier observation of Engel and Schneider (1973).[5]

Proposition 22 (BP upper bound) *For any non-negative p and* $\beta \in \mathcal{B}_p^{(int)}$ *solving Equations (6) (if the solution exists)*

$$\text{perm}(p) \leq Z_{BP}(\beta|p)(\prod_{(i,j)\in\mathcal{E}}(1-\beta_{ij}))^{-1}\prod_j(1-\sum_i(\beta_{ij})^2),$$

This statement was made in Proposition 9 of Watanabe and Chertkov (2010).

4.2 New Bounds and γ_*

Of the bounds discussed above, three are related to BP and one to MF, while as argued in Section 4.3 the FFE-based approach interpolates between BP and MF. This motivates exploring below new FFE-based generalizations of the previously known (and discussed in the preceding Subsection) BP and MF bounds.

We first derive a new lower bound generalizing Proposition 20 to the FFE-based case.

Proposition 23 *The following is true for any doubly stochastic matrix* β *and any* $\gamma \in [-1;1]$

$$\text{perm}\left(\beta.*(1-\beta).^{-\gamma}\right) \geq \frac{n!}{n^n}\prod_{(i,j)}(1-\beta_{ij})^{-\gamma\beta_{ij}}.$$

Proof This bound generalizes Corollary 7 of Watanabe and Chertkov (2010), and it follows directly from the Gurvits-van der Waerden Theorem of Gurvits (2008) and Laurent and Schrijver (2010) (see also Proposition 8 of Watanabe and Chertkov (2010), where a misprint should be corrected $n^n/n! \rightarrow n!/n^n$), and the inequality, $\sum_j \beta_{ij}(1-\beta_{ij})^{-\gamma}x_j \geq \prod_j \left((1-\beta_{ij})^{-\gamma}x_j\right)^{\beta_{ij}}$. $\blacksquare$

Then, combining Proposition 23 with Theorem 13, one arrives at the following statement, generalizing Proposition 20:

Corollary 24 (FFE-based lower bound) *For any non-negative p and* $\beta \in \mathcal{B}_p^{(int)}$ *solving Equations (14) (if the solution exists), the following lower bound holds for any* $\gamma \in [-1;1]$

$$\text{perm}(p) \geq Z_f^{(\gamma)}(\beta|p)\frac{n!}{n^n}\prod_{(i,j)\in\mathcal{E}}(1-\beta_{ij})^{\gamma(1-\beta_{ij})}. \tag{19}$$

Next, one arrives at the following FFE-based generalization of Proposition 22.

Corollary 25 (FFE-based upper bound #1) *For any non-negative p and* $\beta \in \mathcal{B}_p^{(int)}$ *solving Equations (14) (if the solution exists), the following upper bound holds for any* $\gamma \in [-1;1]$

$$\text{perm}(p) \leq Z_f^{(\gamma)}(\beta|p)(\prod_{(i,j)\in\mathcal{E}}(1-\beta_{ij})^{\gamma})\prod_j\sum_i\beta_{ij}(1-\beta_{ij})^{-\gamma}.$$

---

5. The proof of the Theorem 8 in Watanabe and Chertkov (2010) contained a misprint that was corrected in the erratum available at `https://sites.google.com/site/mchertkov/publications/mypapers/91_erratum.pdf`.

This upper bound follows from combining Theorem 13, with the simple (and standard) upper bound, $\text{perm}(A) \leq \prod_j(\sum_i A_{i,j})$ applied to $A = \beta. * (1 - \beta).^{-\gamma}$.

Note that Corollary 25, applied to the $\gamma = 0$ case and reinforced by the observation, that for $\gamma \geq 0$ the minimum of the FFE-based function (13) is achieved in $\beta \in \mathcal{B}_p^{(int)}$, translates into

$$\text{perm}(p) \leq Z_{o-f}^{(\gamma=0)}(p). \tag{20}$$

Combined with Proposition 17, Equation (20) results in the following:

Corollary 26 (FFE-based upper bound #2) *For any non-negative p*

$$\forall \gamma \geq 0: \quad \text{perm}(p) \leq Z_{o-f}^{(\gamma)}(p).$$

This completes the list of inequalities we were able to derive generalizing the BP and MF inequalities stated in the preceding Subsection for the FFE-based case. These generalizations are valid (at least) for any $\gamma \in [0;1]$. Therefore, one may hope to derive somewhat stronger statements reinforcing the continuous family of inequalities with the mononotonicity of the FFE-based approach stated in Proposition 17.

Indeed, combining Equations (16) with Propositions 17,26 one arrives at

Proposition 27 (Special γ_*) *For any non-negative p there exists a special $\gamma_* \in [-1;0]$, such that* $\text{perm}(p) = Z_{o-f}^{(\gamma_*)}(p)$*, and the minimal FFE-based solution upper (lower) bounds the permanent at* $0 \geq \gamma > \gamma_*$ *($-1 \leq \gamma < \gamma_*$).*

proposition 27 motivates our experimental analysis of the $\gamma_*(p)$ dependence discussed in Section 5.

Note also that due to the monotonicity stated in Proposition 17, the $\gamma = 0$ upper bound on the permanent is tighter than the MF, $\gamma = 1$, upper bound. However, and as discussed in more details in the next Subsection, even the $\gamma = 0$ upper bound on the permanent is not expected to be tight.

4.3 Conjectures

It was Conjectured in Vontobel (2010) that

$$\text{perm}(p) \leq Z_{o-BP}(p) * f(n), \tag{21}$$

and also that $f \sim \sqrt{n}$. The second part of the Conjecture was disproved by Gurvits (2011) with an explicit counter-example. The inequality in Equation (21) turns into the equality $f(n) = \sqrt{2}^n$ when p is doubly stochastic matrix and block diagonal, with all the elements in the 2×2 blocks equal to $1/2$.[6] Then it was Conjectured in Gurvits (2011) that

Conjecture 28 (BP upper bound Gurvits, 2011) *For any non-negative p, $f(n)$ in Equation (21) is $\sim \sqrt{2}^n$.*

Another related (but not identical) Conjecture of Gurvits (2011) is as follows:

---

6. Note that this special form of the 2×2 block corresponds to the "double degeneracy" discussed in the paragraph preceding Corollary 8.

Conjecture 29 *The following inequality holds for any doubly stochastic $n \times n$ matrix ϕ:*

$$\text{perm}(\phi) \leq \sqrt{2}^n \prod_{(i,j)} (1 - \phi_{ij})^{(1-\phi_{ij})}. \tag{22}$$

Note that if Equation (22) is true it implies according to Linial et al. (2000) a deterministic polynomial-time Algorithm to approximate the permanent of $n \times n$ nonnegative matrices within the relative factor $\sqrt{2}^n$.

It can be verified directly that the condition (22) is achieved (i.e., inequality is turned into equality) for the special matrix built of the "doubly degenerate" blocks. This special case results in $Z_f^{(\gamma)}$ with $\gamma = -1/2$ on the right-hand side of Equation (22). Therefore one reformulates Conjecture 28 as

Conjecture 30 *The following inequality holds for any non-negative p*

$$\text{perm}(p) \leq Z_{o-f}^{\gamma=-1/2}(p).$$

We refer an interested reader to Vontobel (2013) for discussion of some other Conjectures related to permanents.

5. Experiments

We experiment with deterministic and random (drawn from an ensemble) non-negative matrices.

Our simple deterministic example is of the matrices with elements taking two different values such that all the diagonal are the same and such that all the off-diagonal elements are the same (Watanabe and Chertkov, 2010).

In our experiments with stochastic matrices we consider the following four different ensembles

- $(\lambda_{\text{in}}, \lambda_{\text{out}})$*:* Ensemble of matrices motivated by Chertkov et al. (2008) and Chertkov et al. (2010) and corresponding to a mapping between two consecutive images in 2d flows parameterized by the vector $\lambda = (a, b, c, \kappa)$, where κ is the diffusion coefficient and (a, b, c) stand for the three parameters of the velocity gradient tensor (stretching, shear and rotation, respectively—see Chertkov et al. 2010 for details). In generating such a matrix p we need to construct two sets of λ parameters. The first one, λ_{in}, is used to generate an instance of particle positions in the second image, assuming that particles are distributed uniformly at random in the first image. The second one, λ_{out}, corresponds to an instance of the guessed values of the parameters in the learning problem, where computation of the permanent is an auxiliary step. (Actual optimal learning consists in computing the maximum of the permanent over λ_{out}.) In our simulations we test the quality of the permanent approximations in the special case, when $\lambda_{\text{in}} = \lambda_{\text{out}}$, and also in other cases when the guessed values of the parameters do not coincide with the input ones, $\lambda_{\text{in}} \neq \lambda_{\text{out}}$.

- $[0; \rho]$*-uniform:* In this case one generates elements of the matrix independently at random and distributed uniformly within the $[0; \rho]$-range.

- δ*-exponential:* In this case one generates elements of the matrix independently at random. Any element is an exponentially distributed random variable with mean δ.

(a)   log(Right hand side of Equation (19)/perm(p)) vs. T at $n = 20$, $w = 2$ and different values of γ

(b) γ_* vs. n at different values of T and $w = 2$
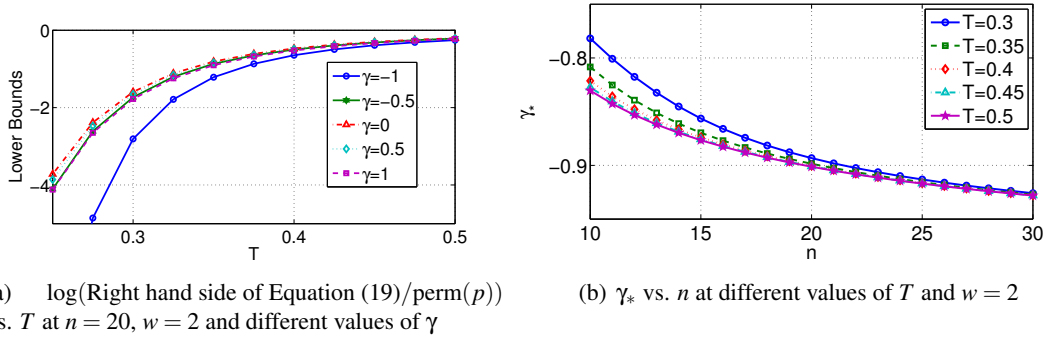
Figure 2:   Illustration for the case of the deterministic matrix (23). Figure (a) shows the gap between the exact permanent and its lower bound estimate by Equation (24). Figure (b) shows dependence of the special γ_* on the parameters.

- $[0;\rho]$-*shifted:* We generate the block diagonal matrix with $\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ blocks and add independent random and uniformly distributed in the $[0;\rho]$ interval components to all elements of the matrix. The choice of this ensemble is motivated by the special role played by the (doubly degenerate) block-diagonal matrix in the Gurvits Conjecture discussed in Section 4.3.

To make the task of the exact computation of the permanent of a random matrix tractable we consider sparsified versions of the ensembles defined above. To achieve this goal we either prune full matrix from the bare (i.e., not yet pruned) ensemble, according to the procedure explained in Appendix F, or in the case of the $[0;\rho]$-uniform ensemble we first generate a sufficiently sparse sub-graph of the fully connected bipartite graph (for example picking a random subgraph of fixed $O(1)$ degree) and then generate nonzero elements corresponding only to edges of the sub-graph.

5.1 Deterministic Example

We consider a simple example which was already discussed in Watanabe and Chertkov (2010). The permanent of the matrix p with elements

$$p_{ij} = \begin{cases} w^{1/T}, & i = j \\ 1, & i \neq j \end{cases},$$

where $w > 1$ and $T > 0$, can be evaluated through the recursion,

$$\sum_{k=0}^{n} W^{(n-k)/T} \binom{n}{k} D_k,$$

$D_0 = 1$, $D_1 = 0$, and $\forall k \geq 2, D_k = (k-1)(D_{k-1} + D_{k-2})$. On the other hand, seeking for solution of the FFE-based Equations (14) in the form of a doubly stochastic matrix β, where

$$\beta_{ij} = \begin{cases} 1 - \varepsilon(n-1), & i = j \\ \varepsilon, & i \neq j \end{cases}, \tag{23}$$

one finds that ε should satisfy the following transcendental equation,

$$(1 - \varepsilon(n-1))(1-\varepsilon)^\gamma = w^{1/T}(n-1)^\gamma \varepsilon^{1+\gamma}.$$

At $T \to \infty$ this equation has a unique uniform, $\varepsilon \to 1/(n-1)$, solution. An interior, $\varepsilon > 0$, solution of Equation (23) exists, and it is also unique, at any finite T for $\gamma > -1$. According to Watanabe and Chertkov (2010), the interior solution does not exists at $\gamma = -1$ and $T < \log \omega / \log(n-1)$.

To test the gap between the exact expression for the permanent and the FFE-based lower bound of Corollary 24, we fix $w = 2, n = 20$ and vary the temperature parameter, T. The results are shown in Figure 2a. One finds that the gap depends on γ with $\gamma = 0$ resulting in the best lower bound for all the tested temperatures. One also observes that the γ-dependence of the gap decreases with increase in T. Figure 2b shows dependence of the special γ_*, defined in Proposition 27, on n and T at $w = 2$. One finds that, consistently with the Conjecture 30, γ_* is always smaller than $-1/2$ and it also decreases with increase in either n or T.

5.2 Random Matrices. Special γ_*.

We search for the special $\gamma = \gamma_*$, defined in Proposition 27, by calculating the permanent of a full matrix, p, of size $n \times n$, with $n = 3, \ldots, 14$, and of a pruned matrix with $n = 10, \ldots, 40$, and then comparing it with the FFE-based value $Z_f^{(\gamma)}(\beta, p)$,[7] where the doubly stochastic matrix β solves Equations (14) for given p, for different γ. By repeatedly evaluating the FFE-based approximation for different values of γ and then taking advantage of the $Z_f^{(\gamma)}$ monotonicity and continuity with respect to γ and performing a search we find the special γ for a specific p.

In general we observed that the special γ_* for tested matrices was always less than or equal to $-1/2$, which is consistent with Conjecture 30. We also observed, estimating or extrapolating the approximate value of the special γ_* for a given matrix, that it might be possible to estimate the permanent of a matrix efficiently and very accurately for some ensembles.

5.2.1 THE $(\lambda_{\text{in}}, \lambda_{\text{out}})$ ENSEMBLE

In this Subsection we describe experiments with several of the $(\lambda_{\text{in}}, \lambda_{\text{out}})$ ensembles. We are interested in studying the dependence of the special γ_*, defined in Proposition 27, on the matrix size and other parameters of the ensemble. We consider here a variety of cases.

Figure 3 shows the results of experiments with full but (relatively) small matrices and different values $\lambda_{\text{in}}, \lambda_{\text{out}}$. The results are presented in the form of a scatter plot, showing results for different matrix instances from the same ensemble.

As can be seen from the grouping of the first five plots in Figure 3, the dependence of the special γ_* on the matrix size at $\lambda_{\text{in}} = \lambda_{\text{out}}$ is largely sensitive to the diffusion parameter κ and it is not so dependent on the advection parameters a, b, c. Indeed, Figures 3 (a,b) are similar to each other, as are Figures 3(c-e), despite having different values of a, b, c.

Figures 3(a-e) also demonstrate an interesting feature: the lower κ, the more erratic the behavior of the special γ_*.

Analyzing the three last cases in Figure 3 with $\lambda_{\text{in}} \neq \lambda_{\text{out}}$, we observed that the larger the value of κ_{out}, the more regular the resulting behavior.

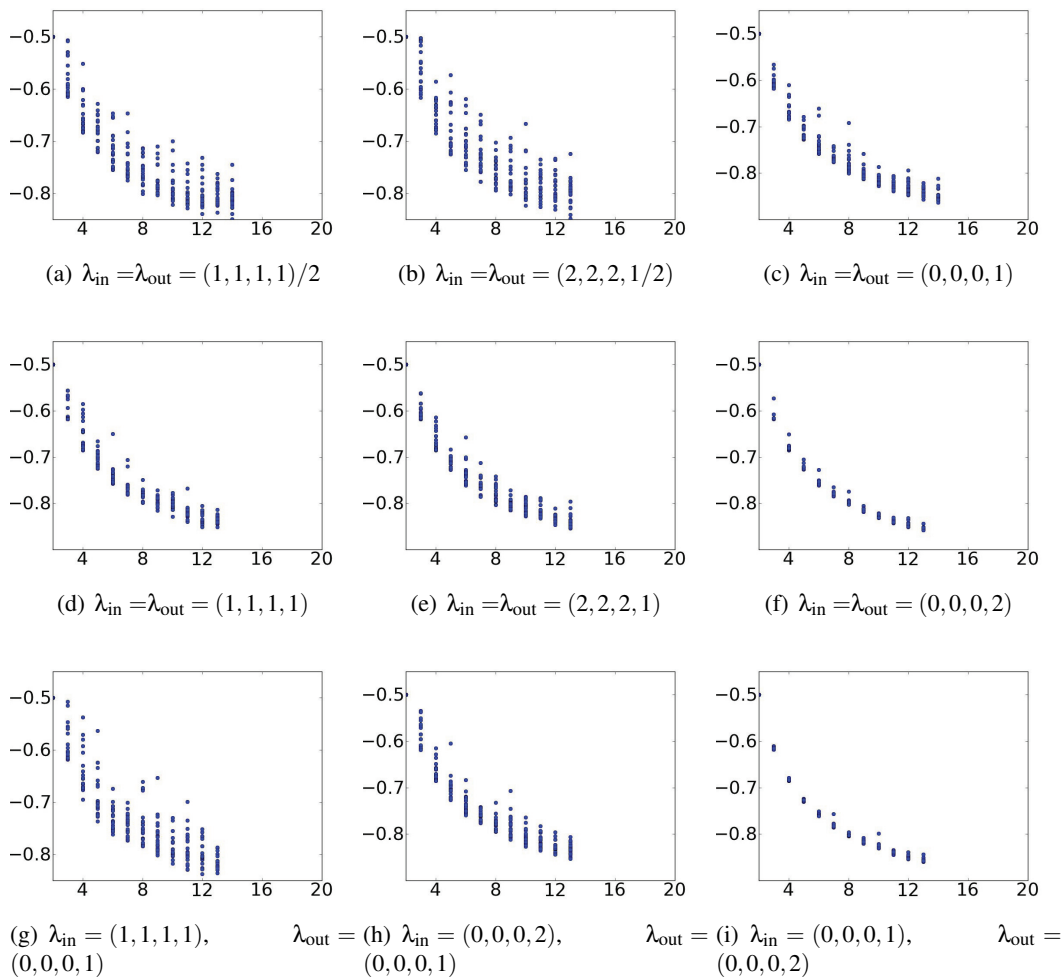7. In the following we will use the shorter notation, $Z_f^{(\gamma)}$ for this object.

(a) $\lambda_{\text{in}} = \lambda_{\text{out}} = (1,1,1,1)/2$

(b) $\lambda_{\text{in}} = \lambda_{\text{out}} = (2,2,2,1/2)$

(c) $\lambda_{\text{in}} = \lambda_{\text{out}} = (0,0,0,1)$

(d) $\lambda_{\text{in}} = \lambda_{\text{out}} = (1,1,1,1)$

(e) $\lambda_{\text{in}} = \lambda_{\text{out}} = (2,2,2,1)$

(f) $\lambda_{\text{in}} = \lambda_{\text{out}} = (0,0,0,2)$

(g) $\lambda_{\text{in}} = (1,1,1,1)$, $\lambda_{\text{out}} = (0,0,0,1)$

(h) $\lambda_{\text{in}} = (0,0,0,2)$, $\lambda_{\text{out}} = (0,0,0,1)$

(i) $\lambda_{\text{in}} = (0,0,0,1)$, $\lambda_{\text{out}} = (0,0,0,2)$

Figure 3: Scatter plot of the special γ_* calculated for instances from the $(\lambda_{\text{in}}, \lambda_{\text{out}})$ ensemble and varying the matrix size within the $2 \div 14$ range (no pruning).

Figure 4 shows the same scatter plots as in Figure 3, observed for larger but sparser (90% pruned) matrices. We observed the general tendency for the average special γ_* to decrease with increasing n; however, it is not clear from the observations if the resulting level of fluctuations decreases with the increase in n or remains the same.

Summarizing, for the $(\lambda_{\text{in}}, \lambda_{\text{out}})$ ensemble, we found that the behavior of the special γ_* with respect to matrix size to be largely dependent on κ_{out}, the diffusion coefficient used to generate the matrix, while the dependence on other factors is significantly less pronounced. The average special γ_* decreases with increasing n, while respective variance remains roughly the same.
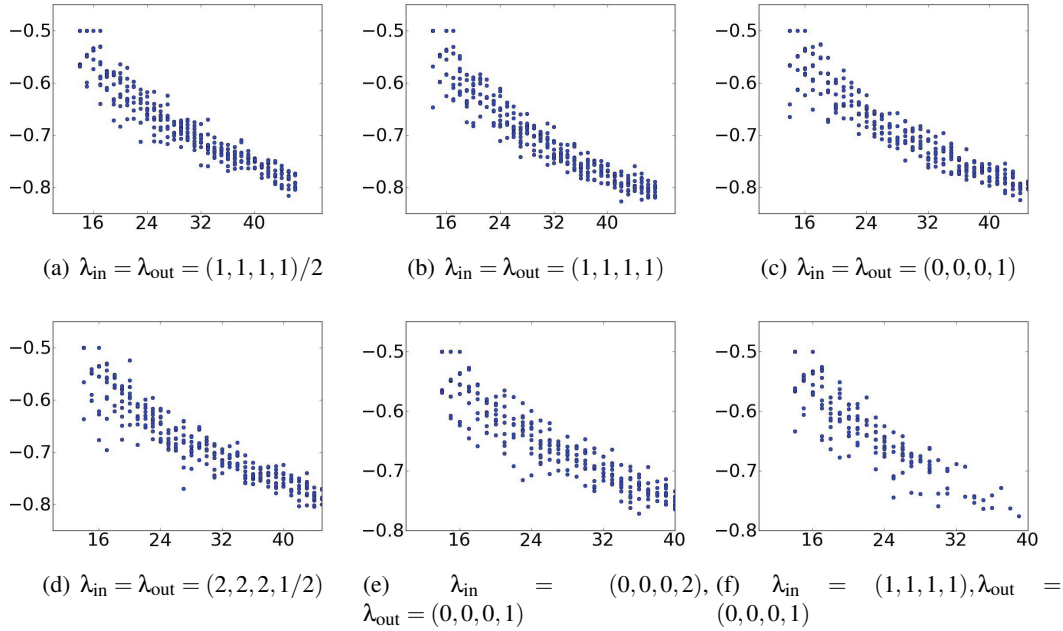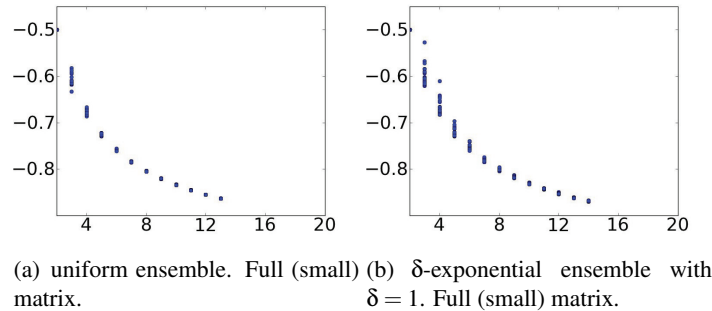
(a) $\lambda_{in} = \lambda_{out} = (1,1,1,1)/2$ (b) $\lambda_{in} = \lambda_{out} = (1,1,1,1)$ (c) $\lambda_{in} = \lambda_{out} = (0,0,0,1)$

(d) $\lambda_{in} = \lambda_{out} = (2,2,2,1/2)$ (e) $\lambda_{in} = (0,0,0,2)$, (f) $\lambda_{in} = (1,1,1,1), \lambda_{out} = (0,0,0,1)$ $\lambda_{out} = (0,0,0,1)$

Figure 4: Scatter plot of the estimated special γ_* calculated for instances from the $(\lambda_{in}, \lambda_{out})$ ensemble and varying the matrix size within the $15 \div 40$ range (with 90% pruning).

(a) uniform ensemble. Full (small) matrix. (b) δ-exponential ensemble with $\delta = 1$. Full (small) matrix.

Figure 5: Scatter plot of the estimated special γ calculated for instances from a random matrix ensemble and varying the matrix size.

5.2.2 UNIFORM AND δ-EXPONENTIAL ENSEMBLES

Figure 5 shows scatter plots for examples of the (a) $[0;1]$-uniform ensemble,[8] and (b) δ-exponential ensemble. Here we found a very impressive decrease in variance with increase in the matrix size. Besides, we observe that in spite of their difference, the two ensembles show qualitatively similar behavior of γ_* as a function of n. This indicates that for large matrices, whose entries are inde-

---

8. The simulations results of Section 4 of Huang and Jebara (2009) and the Conjectures in Section VII.B of Vontobel (2013) apply to the same ensemble.
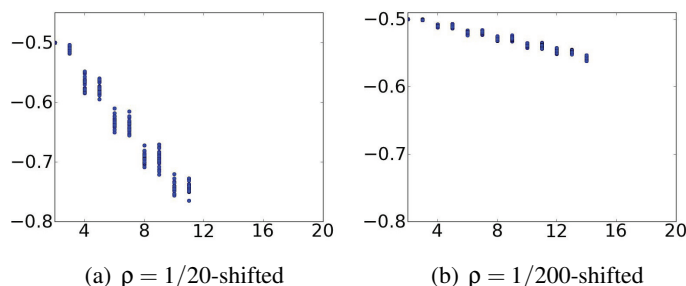
(a) $\rho = 1/20$-shifted (b) $\rho = 1/200$-shifted

Figure 6: Scatter plot of the estimated special γ_* calculated for instances from two examples of the $[0;\rho]$-shifted ensembles and varying the matrix size.

pendent random variables, we could achieve excellent accuracy by extrapolating on the special γ_* and estimating $Z_f(\gamma_*) = \text{perm}(p)$. Indeed, the rapidly decreasing variance suggests that most of the error would come from the extrapolation of the average special γ_* for a given matrix size and not the error of $Z_f(\gamma_{*,\text{average}}) - Z_f(\gamma_{*,\text{actual}})$, since $\gamma_{*,\text{average}}$ and $\gamma_{*,\text{actual}}$ will be very close in value.

5.2.3 THE $[0;\rho]$-SHIFTED ENSEMBLE

We also studied the special γ_* in "badly-behaved" cases such as the one brought up earlier, with 2×2 squares of $1/2$'s positioned along the diagonal. (See discussion in Section 4.3.) It can be easily shown that the special value of γ_* of the bare block-diagonal matrix is $-1/2$. Unsurprisingly, our experiments, documented in Figure 6, showed that: (a) the resulting γ is always smaller than $-1/2$, and (b) as more noise was introduced, the special γ_* decreased in value faster with respect to matrix size. However, this decrease with n towards smaller γ_* was much slower than in other ensembles, particularly for low noise.

5.3 Random Matrices: Testing Inequalities and Conjectures

Figures 7,8 and Figures 9,10, showing the average behavior and scatter plots for smaller and larger (pruned) matrices, respectively (see Figure captions for explanations), present experimental verification to the variety of inequalities discussed in Section 4.1. The ensemble used for these plots was the ensemble $\lambda_{\text{in}} = \lambda_{\text{out}} = (1,1,1,1)/2$. The data suggests that neither of the bounds are actually tight, and moreover the values of the gaps, between the exact expression and respective estimates tested, fluctuate more strongly with increasing matrix size. We also observe from Figures 7 and Figures 8f, that Equation (21) has $f(n)$ growing faster with n than $\sim \sqrt{n}$ even on average. In the case of larger pruned matrices we removed the two expressions $\log((Z_{BP}(\prod_{(i,j)\in\mathcal{E}}(1-\beta_{ij}))^{-1}\prod_j(1-\sum_i(\beta_{ij})^2))/Z)$ and $\log((2Z_{BP}(\prod_{i,j}(1-\beta_{ij}))^{-1}\prod_i\beta_{i\Pi(i)}(1-\beta_{i\Pi(i)}))/Z)$. We removed the former because in the case of pruning the resulting β is often partially-resolved (with some elements of β equal to one) and in this case the inequality does not carry any restriction. We removed the latter because, in the pruned case and for a randomly chosen permutation, it is very likely that at least one element of β is zero, making the bound discussed unrestrictive.

Figure 11 shows that the bounds given by the corollaries 24, 25 do not depend much on γ and that they in practice depend more on matrix size and on peculiarities of individual matrices. There
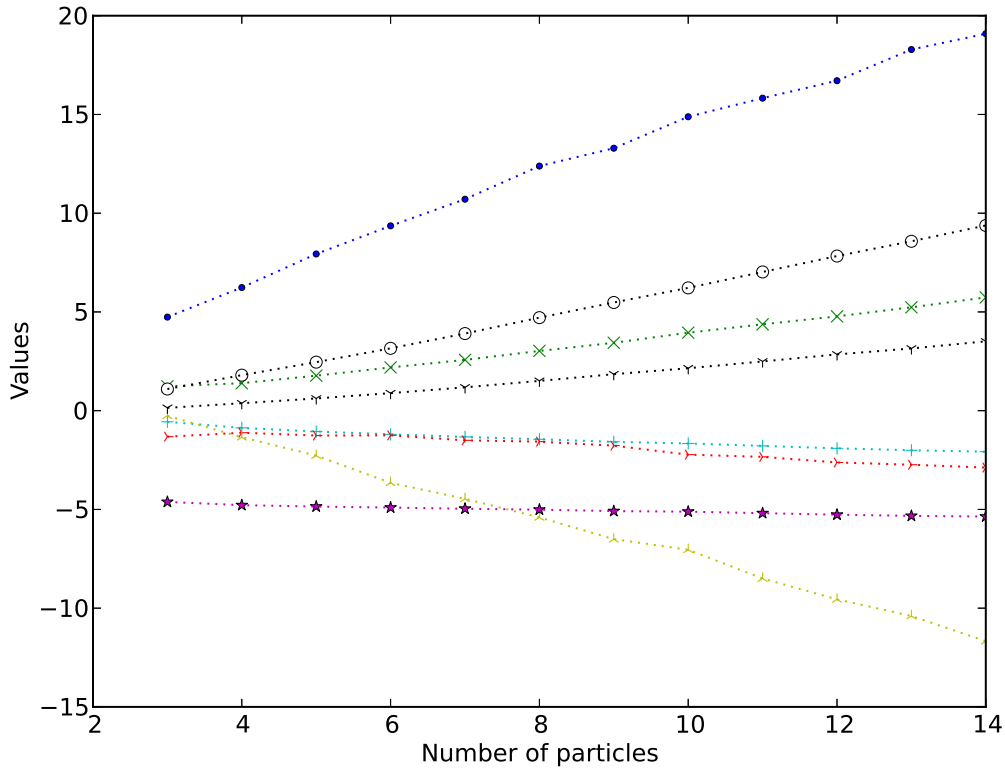
Figure 7: This Figure describes the gap between the actual value of the permanent and the values of the various theoretical upper and lower bounds described in the paper and averaged over different simulation trials. Each color corresponds to an upper or lower bound as follows: Blue dots corresponds to the MF approximation to the permanent, Z_{MF}. Green X's correspond to $\log(Z_{BP}(\prod_{(i,j)\in\mathcal{E}}(1-\beta_{ij}))^{-1}\prod_j(1-\sum_i(\beta_{ij})^2)/Z)$. Red sideways Y's correspond to $\log(Z_{BP}\prod_{i,j}(1-\beta_{ij})^{\beta_{ij}-1}\frac{n!}{n^n}/Z)$. Cyan "+" symbols correspond to $\log(Z_{BP}/Z)$. Yellow inverted Y's correspond to $\log(2Z_{BP}(\prod_{i,j}(1-\beta_{ij}))^{-1}\prod_i\beta_{ii}(1-\beta_{ii})/Z)$. Purple stars correspond to $\log(0.01Z_{BP}\sqrt{n}/Z)$. Black circles correspond to $\log(Z_f^{\gamma=0}/Z)$. Black Y's correspond to $\log(Z_f^{\gamma=-0.5}/Z)$. To make a data point, 100 instances, each corresponding to a new matrix are drawn, and the log of the ratio of the bound to the actual permanent is recorded. The data shown corresponds to matrices from the $(\lambda_{\text{in}},\lambda_{\text{out}})$ ensemble.

is a slight change for values of γ near -1, but otherwise the plot is nearly flat, so it seems that unfortunately little tightening of the bounds can be achieved by tweaking γ. Another noteworthy observation is that a higher upper bound implies a higher lower bound, and vice-versa.

Figure 12 is related to discussions of Corollary 19 for the permanent of a doubly stochastic matrix. We generate an instance of a doubly stochastic matrix and calculate the respective BP expression in three steps (this is the procedure of Knopp and Sinkhorn (1967), also discussed by
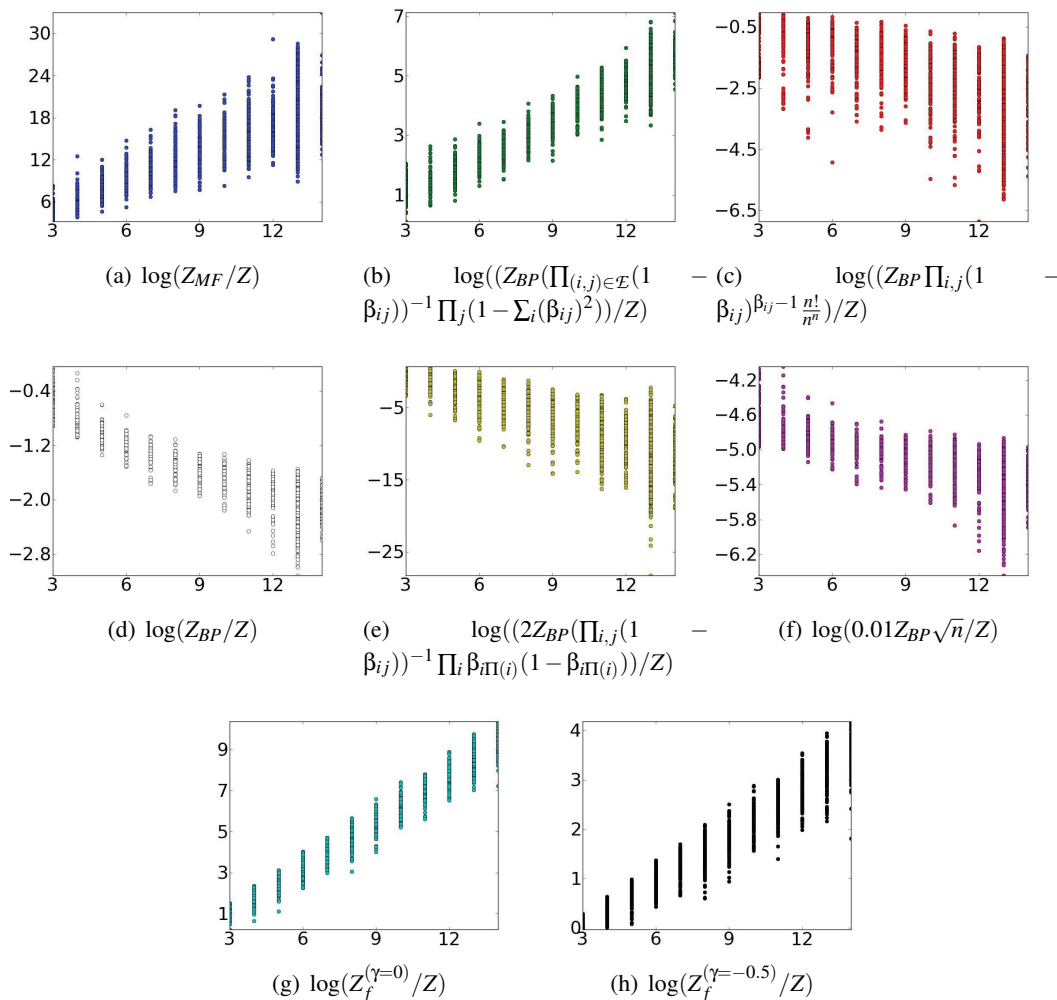
(a) $\log(Z_{MF}/Z)$

(b) $\log((Z_{BP}(\prod_{(i,j)\in\mathcal{E}}(1-\beta_{ij}))^{-1}\prod_j(1-\sum_i(\beta_{ij})^2))/Z)$

(c) $\log((Z_{BP}\prod_{i,j}(1-\beta_{ij})^{\beta_{ij}-1}\frac{n!}{n^n})/Z)$

(d) $\log(Z_{BP}/Z)$

(e) $\log((2Z_{BP}(\prod_{i,j}(1-\beta_{ij}))^{-1}\prod_i\beta_{i\Pi(i)}(1-\beta_{i\Pi(i)}))/Z)$

(f) $\log(0.01Z_{BP}\sqrt{n}/Z)$

(g) $\log(Z_f^{(\gamma=0)}/Z)$

(h) $\log(Z_f^{(\gamma=-0.5)}/Z)$

Figure 8: Scatter plots for the data shown in Figures (7). For better presentation the data is split into 8 sub-figures. The vertical axis of each scatter plot is specific to the behavior of the expression with respect to the matrix size and the color coding for different objects tested coincides with that used in Figure 7.

Huang and Jebara 2009): (a) generate a non-negative matrix from the $[0;1]$ ensemble; (b) re-scale rows and columns of the matrix iteratively to get a respective doubly stochastic matrix;[9] and (c) apply the BP- ($\gamma=-1$) procedure to evaluate the Z_{o-BP} estimate for the resulting doubly stochastic matrix. In agreement with Equation (18), the average value of the log corresponding to the BP-lower bound is positive and smaller than the respective expression for the average of the log of the explicit expression on the right-hand side of Equation (18). (The hierarchical relation obviously holds as well for any individual instance of the doubly stochastic matrix β from the generated ensemble.) We

9. The rescaling is a key element of Linial et al. (2000), and we can also think of the procedure as of a version of the $\gamma=0$ iterative Algorithm.
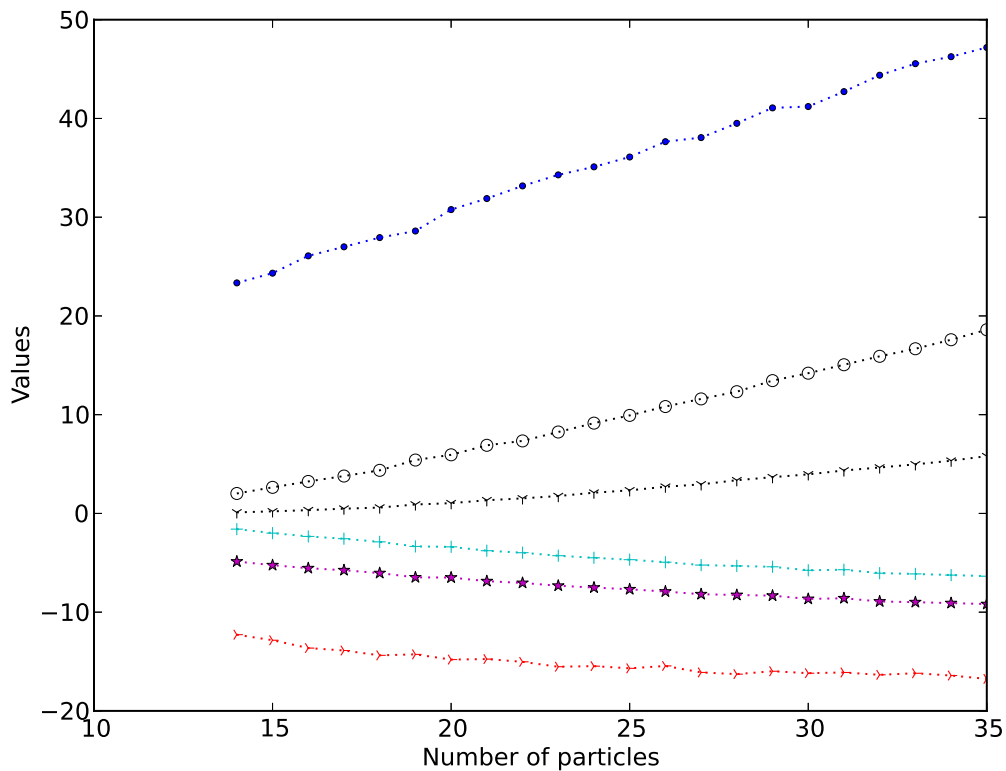
Figure 9: The data is shown like in Figure 7, but for large, sparsified matrices. Less meaningful expressions were removed from the plot. Each color/symbol pair corresponds to a mathematical expression, as follows, blue dots: $\log(Z_{MF}/Z)$, red sideways Y's: $\log(Z_{BP}\prod_{i,j}(1-\beta_{ij})^{(\beta_{ii}-1)} \cdot n!/(n^n Z))$, cyan "+" symbols: $\log(Z_{BP}/Z)$, purple stars: $\log(Z_{BP}\sqrt{n}/(100Z))$, black circles: $\log(Z_f(\gamma=0)/Z)$, black Y's: $\log(Z_f(\gamma=-1/2)/Z)$; where $Z = \text{perm}(p)$.

also observe that the average values of the curves show a tendency to saturate, while the standard deviation decreases dramatically, suggesting that for large n this random ensemble may be well approximated by either BP or, even more simply, by its explicit lower bound from the right-hand side of Equation (18), the latter being in the agreement with the proposal of Gurvits (2011).

6. Conclusions and Path Forward

The main message of this and other related recent papers by Chertkov et al. (2008), Huang and Jebara (2009), Chertkov et al. (2010), Watanabe and Chertkov (2010), Vontobel (2010), Vontobel (2013) and Gurvits (2011) is that the BP approach and improvements not only give good heuristics for computing permanents of non-negative matrices, but also provide theoretical guarantees and thus reliable deterministic approximations. The main highlights of this manuscript are

(a) $\log(Z_{MF}/Z)$

(b) $\log((Z_{BP} \prod_{i,j}(1 - \beta_{ij})^{\beta_{ij}-1} \frac{n!}{n^n})/Z)$

(c) $\log(Z_{BP}/Z)$

(d) $\log(0.01 Z_{BP}\sqrt{n}/Z)$

(e) $\log(Z_f^{(\gamma=0)}/Z)$

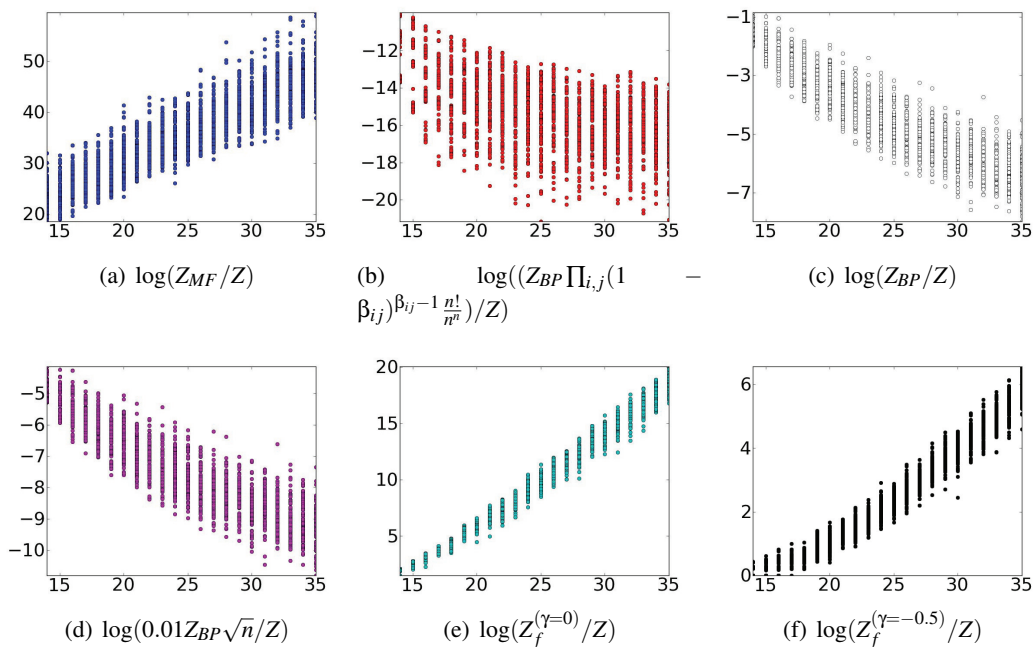(f) $\log(Z_f^{(\gamma=-0.5)}/Z)$

Figure 10: Scatter plots for the data shown in Figures (9). For better presentation the data is split into 6 sub-figures. Each scatter plot is specific to the behavior of the expression with respect to the matrix size and the color coding for different objects tested coincides with the one used in Figure 9.
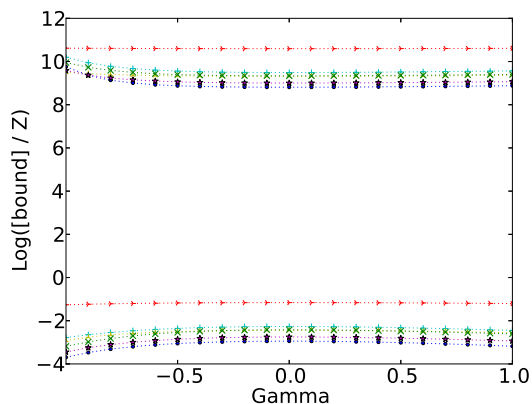
Figure 11: This plot shows γ-dependence of the gap between upper and lower bounds corresponding to Corollaries 24, 25. Here, the bounds were plotted for six different matrices, each generated with $\lambda_{\text{in}} = \lambda_{\text{out}} = (1,1,1,1)/2$, where the upper and lower bounds are color-coded and use matching symbols to indicate that they correspond to the same matrix.
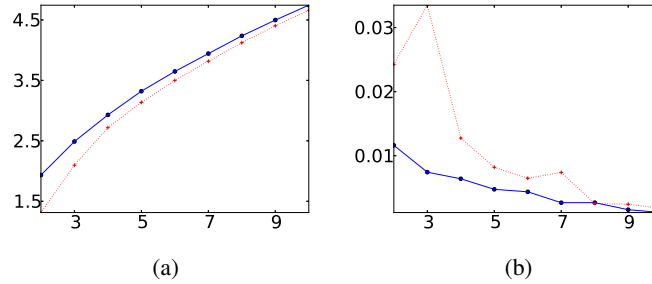
Figure 12: A plot of the average (Subfigure a) and standard deviation (Subfigure b) of $\mathrm{perm}(\beta)/\prod_{i,j}(1-\beta_{ij})^{1-\beta_{ij}}$ (in blue dots connected by a solid line) and $\mathrm{perm}(\beta)/Z_{o-BP}(\beta)$ (in red crosses connected by a dashed line), where β is a doubly stochastic matrix picked from 100 instances of the random ensemble described in the text, shown as a function of n.

- The construction of the FFE-based approach, parameterized by $\gamma \in [-1;1]$ and interpolating between BP ($\gamma = -1$) and MF ($\gamma = 1$) limits.

- The discovery of the exact relation between the permanent of a non-negative matrix, $\mathrm{perm}(p)$ and the respective FFE-based expression, $Z_f^{(\gamma)}(p)$, where the latter is computationally tractable.

- The proof of the continuity and monotonicity of $Z_f^{(\gamma)}(p)$ with γ, also suggesting that for some $\gamma_* \in [-1;0]$, $\mathrm{perm}(p) = Z_f^{(\gamma_*)}(p)$.

- The extension of the list of known BP-based upper and lower bounds for the permanent by their FFE-based counterparts.

- The experimental analysis of permanents of different ensembles of interest, including those expressing relations between consecutive images of stochastic flows visualized with particles.

- Our experimental tests include analysis of the gaps between exact expression for the permanent, evaluated within the ZDD technique adapted to permanents, and the aforementioned BP- and fractional-based lower/upper bounds.

- The experimental analysis of variations in the special γ for different ensembles of matrices suggests the following conclusions. First, the behavior of the special γ varies for different ensembles, but the general trend remains the same: as long as there is some element of randomness in the ensemble, the special γ decreases as matrix size increases. Second, for each ensemble the behavior of the special γ is highly distinctive. For some considered random matrix ensembles, the variance decreases quickly with increasing matrix size. All of the above suggest that the FFE-based approach offers a lot of potential for estimating matrix permanents.

We view these results as creating a foundation for further analysis of theoretical and computational problems associated with permanents of large matrices. Of the multitude of possible future problems, we consider the following ones listed below as the most interesting and important:

- Improving BP and FFE-based approaches and making the resulting lower and upper bounds tighter.

- Further analysis of the γ-dependence, making theoretical statements for statistics of log-permanents at large n and for different random ensembles.

- Using the new permanental estimations and bounds for learning flows in the setting of Chertkov et al. (2010). Combining within the newly introduced FFE-based approach the β-optimization with optimization over flow parameters (by analogy with what is done in Chertkov et al. 2010). Applying the improved technique to various particle image velocimetry (PIV) experiments of interest in fluid mechanics in general, and specifically to describe spatially smooth multi-pole flows in micro-fluidics, see, for example, discussion of the most recent relevant experiments in Drescher et al. (2010) and Guasto et al. (2010) and references therein.[10]

- Addressing other GM problems of the permanental type, for example, counting matchings (and not only perfect matchings) on arbitrary graphs, drawing inspiration from Sanghavi et al. (2011) generalizing Bayati and Nair (2006) and Bayati et al. (2008) in the ML setting, and higher dimensional matchings, in particular corresponding to matching of paths between multiple consecutive images within the "learning the flow" setting.

Acknowledgments

Appendix A. Most Probable Perfect Matching

This short Appendix is introduced to guide the reader through material which is related, but only indirectly (through physics motivation and historical links), to the main subject of the manuscript.

A.1 Most Probable Perfect Matching over Bi-Partite Graphs

According to Equation (2), the permanent can be interpreted as the partition function of a GM. The partition function represents a weighted counting of the $n!$ perfect matchings. Using "physics terminology" one says that this perfect matching representation allows to interpret the permanent as the statistical mechanics of perfect matchings (called dimers in the physics literature) over the bi-partite graph. This is statistical mechanics at finite temperatures, as the partition function represents a (statistical) sum over the perfect matchings.

---

10. We are thankful to Eric Lauga for suggesting to us the micro-fluidics experiments as one possible application for the "learning the flow" BP-based approach.

However, it is still of interest to discuss (at least in the context of establishing historical links) the "zero temperature," or maximum likelihood (ML) version of Equation (2)

$$-\log(Z_{ML}(p)) = \min_{\sigma} \sum_{(i,j)\in\mathcal{E}} \sigma_{ij} \log(1/p_{ij}). \tag{24}$$

According to the logic of Yedidia et al. (2005), Equation (24) can also be stated in the probability space (i.e., in terms of $b(\sigma)$) as

$$-\log(Z_{ML}(p)) = \min_{b} \sum_{(i,j)\in\mathcal{E}} \sum_{\sigma} \sigma_{ij} \log(1/p_{ij}) b(\sigma). \tag{25}$$

Then, the function of $b(\sigma)$ which is the object of minimization over beliefs in Equation (25) is naturally the ML (zero temperature) version of the FE function (3).

By construction, $Z_{ML}(p) \leq Z(p)$ for any p. Note also that Equation (24) is a linear programming (LP) equation, but one which at first sight appears intractable, giving an optimization defined over a huge polytope and spanning all the perfect matchings with nonzero probability. For a general GM the LP-ML formulation is indeed intractable, but for the specific problem under consideration (finding the perfect matching over a bipartite graph) the ML-LP problem (25) becomes tractable, as discussed below in the next Subsection. Given classical results from optimization theory, related to the so-called Hungarian Algorithm, by Kuhn (1955), and the auction Algorithm, by Bertsekas (1992), this special solvability (reduced complexity) of the ML perfect matching problem is not surprising.

A.2 Linear Programming Relaxation of BP

The Bethe FE (4) can be split naturally into the self-energy term and the self-entropy terms (at unit temperature), $F_{BP} = E_{BP} - S_{BP}$:

$$E_{BP}(\beta|p) = -\sum_{(i,j)} \beta_{ij} \log(p_{ij}),$$

$$S_{BP}(\beta|p) = \sum_{(i,j)} \left(-\beta_{ij} \log(\beta_{ij}) + (1-\beta_{ij}) \log(1-\beta_{ij})\right).$$

If the entropy term is ignored in Equation (8) the problem turns into the linear programming (LP) formulation of BP

$$-\log(Z_{LP}(p)) = \min_{\beta} E_{BP}(\beta). \tag{26}$$

One can also arrive at the same LP formulation (26) by relaxing the original ML-LP setting (25). As shown in Bayati et al. (2008) and Chertkov (2008), the relaxation is provably tight for any p, that is, $Z_{LP}(p) = Z_{ML}(p)$, as the resulting matrix of constraints in the LP problem (26) describing the doubly stochasticity of β is totally uni-modular, so the corners of the respective polytope are in one-to-one correspondence with the perfect matching configurations/corners of the higher-dimensional polytope from Equation (25), also in accordance with the Birkhoff-von Neumann Theorem by (Kőnig, 1936; Birkhoff, 1946; von Neumann, 1953).

Appendix B. Bethe-Free Energy Approach

This Appendix provides some additional details on the Bethe-Free Energy Approach.

B.1 Exact BP-based Relations for Permanents

We present here a simple proof of Equation (7), essentially following a slightly modified version of what was the main statement of Watanabe and Chertkov (2010).

Consider an interior minimum of the Bethe FE function (4) achieved with a strictly nonzero (for elements with positive p_{ij}) doubly stochastic matrix β. Then, the minimum satisfies Equations (6), where $\log(u)$ are respective Lagrangian multipliers. Weighting the logarithm of Equations (6) with β, summing up the result over all the edges, using Equation (4) and the double stochasticity of β, one derives

$$
\sum_{(i,j)\in\mathcal{E}} \beta_{ij}\log(u_iu_j) = \sum_{i\in\mathcal{V}_1}\log u_i + \sum_{j\in\mathcal{V}_2}\log u^j
$$
$$
= \sum_{(i,j)\in\mathcal{E}}(\beta_{ij}\log(p_{ij}/\beta_{ij}) - \beta_{ij}\log(1-\beta_{ij})) = \log Z_{BP} - \sum_{(i,j)\in\mathcal{E}}\log(1-\beta_{ij}). \tag{27}
$$

On the other hand, applying the permanent to both sides of Equation (6) one arrives at

$$
\text{perm}(p) = \text{perm}(\beta.*(1-\beta))\left(\prod_{i\in\mathcal{V}_1}u_i\right)\left(\prod_{j\in\mathcal{V}_2}u^j\right). \tag{28}
$$

Combining Equation (27) with Equation (28) results in Equation (7).

B.2 Iterative Algorithm(s) for Finding Solution of BP Equations

First of all, let us recall that according to Proposition 7, (4) is convex. However, as explained above the convexity is not trivial, as it is enforced by global constraints. This lack of convexity of individual edge-local terms in Equation (4) creates a technical obstacle to finding a valid fixed point of F_{BP}, suggesting that an iterative Algorithm converging to the fixed point of F_{BP} will be more elaborate than the one discussed below in the MF case.

To find a valid solution of BP in our numerical experiments we use the following practical iterative scheme (heuristics), previously described in Chertkov et al. (2008) (see Equations (7,8) as well as preceding and following explanations):

$$
\forall(i,j): \quad \beta_{ij}(n+1) = \lambda\beta_{ij}(n) + \frac{(1-\lambda)p_{ij}}{p_{ij}+(\sum_k\beta_{kj}(n)/2+\sum_k\beta_{ik}(n)/2-\beta_{ij}(n))^2(u_i(n)u^j(n))}, \tag{29}
$$

$$
\forall i: \quad u_i(n+1) = \frac{\sum_k p_{ik}/u^k(n)}{1-\sum_j(\beta_{ij}(n))^2}, \quad \forall j: \quad u^j(n+1) = \frac{\sum_k p_{kj}/u_k(n)}{1-\sum_i(\beta_{ij}(n))^2}, \tag{30}
$$

where the arguments of the β's indicate the order of the iterations. The damping parameter λ (typically chosen $0.4 \div 0.5$) helps with convergence. To ensure appropriate accuracy for solutions with β's close to zero or unity we also insert a normalization step after Equations (29) but prior to Equations (30), making the following two transformations subsequently, (a) $\forall(i,j): \beta_{ij} \to \beta_{ij}/\sum_k\beta_{ik}$, and (b) $\forall(i,j): \beta_{ij} \to \beta_{ij}/\sum_k\beta_{kj}$. (The two steps implement an elementary step of the Sinkhorn operation from Huang and Jebara 2009.) The Algorithm is sensitive to initial values for β and u. To ensure convergence, one initiates the Algorithm with the output of the MF scheme (which converges much better) as described in Appendix C.2, that is, $\beta(0) = \beta_{MF}$ and $u(0) = v_{MF}$. Numerical experiments show that this procedure always converges to an interior stationary point of the BFE

(4), when one exists and is not degenerate. In the special cases when the solution is on the boundary it seemed to converge there as well, but we did not study this systematically to make a definitive statement.

Note that the Algorithm presented above is certainly not the only option one can use to find a doubly stochastic matrix solution of BP Equations (6). In fact, the standard sum-product Algorithm (SPA) of Yedidia et al. (2005), stated for the problem of computing the permanent in Chertkov et al. (2010), is a serious competitor, which according to Theorem 32 of Vontobel (2013) always converges to the minimum of the Bethe FE. Future work is required to compare the convergence speed of the two Algorithms.

Appendix C. Mean-Field (Fermi) Approach

This Appendix provides some additional details on the Mean-Field Approach.

C.1 Exact MF-based Relations for Permanents

We present here a simple proof of Equation (12), essentially following the logic of what was described above for BP in Appendix B.1.

Weighting the logarithm of Equations (11) with that doubly stochastic matrix β which minimizes Equation (12), summing the result over all the edges, and making use of Equations (11,9), one derives

$$
\begin{aligned}
\sum_{(i,j)\in\mathcal{E}} \beta_{ij}\log(v_i v_j) &= \sum_{i\in\mathcal{V}_1}\log v_i + \sum_{j\in\mathcal{V}_2}\log v^j \\
&= \sum_{(i,j)\in\mathcal{E}} \left(\beta_{ij}\log(p_{ij}/\beta_{ij}) + \beta_{ij}\log(1-\beta_{ij})\right) = \log Z_{MF}(p) + \sum_{(i,j)\in\mathcal{E}}\log(1-\beta_{ij}).
\end{aligned}
\tag{31}
$$

On the other hand, applying the permanent to both sides of Equation (11) one arrives at

$$
\mathrm{perm}(p) = \mathrm{perm}(\beta./(1-\beta))\left(\prod_{i\in\mathcal{V}_1} v_i\right)\left(\prod_{j\in\mathcal{V}_2} v^j\right).
\tag{32}
$$

Combining Equation (31) with Equation (32) results in Equation (12).

C.2 Iterative Scheme for Solving Mean-Field Equations

An efficient heuristic way to find a (unique) solution of the MF system of Equations (11) for doubly stochastic matrix β is to initialize with $v_i(0) = v^j(0) = 1$ and iterate according to

$$
\beta_{ij}(n+1) = \frac{p_{ij}}{p_{ij} + v_i(n)v^j(n)},
\tag{33}
$$

$$
v_i(n+1) = v_i(n)\sum_j \beta_{ij}(n), \quad v^j(n+1) = v^j(n)\sum_i \beta_{ij}(n),
\tag{34}
$$

until the tolerance $\delta > \max(\mathrm{abs}(\beta(n+1) - \beta(n)))$ is met.

Appendix D. FFE-based Approach

This Appendix provides some additional details on the FFE-based Approach.

D.1 Exact Relations for Permanents

We present here a simple proof of Equation (15), which is a direct generalization of what was discussed above in Appendices B.1, C.1.

Weighting the logarithm of Equations (14) with that doubly stochastic matrix β which minimizes Equation (15), summing the result over all the edges, and making use of Equations (14,13), one derives

$$
\begin{aligned}
\sum_{(i,j)\in\mathcal{E}} \beta_{ij}\log(w_i w_j) &= \sum_{i\in\mathcal{V}_1}\log w_i + \sum_{j\in\mathcal{V}_2}\log w^j \\
&= \sum_{(i,j)\in\mathcal{E}}\left(\beta_{ij}\log(p_{ij}/\beta_{ij}) + \gamma\beta_{ij}\log(1-\beta_{ij})\right) \\
&= \log Z_f^{(\gamma)}(\beta|p) + \gamma\sum_{(i,j)\in\mathcal{E}}\log(1-\beta_{ij}).
\end{aligned}
\tag{35}
$$

On the other hand, applying the permanent to both sides of Equation (14) one arrives at

$$
\mathrm{perm}(p) = \mathrm{perm}(\beta./(1-\beta).^{\gamma})\left(\prod_{i\in\mathcal{V}_1}w_i\right)\left(\prod_{j\in\mathcal{V}_2}w^j\right).
\tag{36}
$$

Combining Equation (35) with Equation (36) results in Equation (15).

D.2 Iterative Scheme for Solving FFE-based Equations

All edge-local terms in the FFE-based function (13) are convex in $\beta \in [0;1]$ for $\gamma > 0$, while for negative γ the edge-term convexity holds only when all elements of β are smaller than a threshold $\beta_c \geq 1/2$, which is a solution of $\beta_c\log(\beta_c) = -\gamma(1-\beta_c)\log(1-\beta_c)$. This suggests different iterative schemes for positive and negative γ.

When $\gamma > 0$ we use the following modification of the MF scheme (33,34):

$$
\begin{aligned}
\beta_{ij}(n+1) &= \frac{p_{ij}(1-\beta_{ij}(n))^{\gamma-1}}{p_{ij}(1-\beta_{ij}(n))^{\gamma-1} + w_i(n)w^j(n)}, \\
w_i(n+1) &= w_i(n)\sum_j\beta_{ij}(n), \quad w^j(n+1) = w^j(n)\sum_i\beta_{ij}(n),
\end{aligned}
$$

In the case of $\gamma \leq 0$ we use the following modification of the BP scheme

$$
\forall (i,j): \quad \beta_{ij}(n+1) = \lambda\beta_{ij}(n)
\tag{37}
$$
$$
+ \frac{(1-\lambda)p_{ij}(1+\beta_{ij}(n))^{1+\gamma}}{p_{ij}(1+\beta_{ij}(n))^{1+\gamma} + (\sum_k\beta_k^j(n)/2 + \sum_k\beta_i^k(n)/2 - \beta_{ij}(n))^2(w_i(n)w^j(n))},
$$
$$
\forall i: \quad w_i(n+1) = \frac{\sum_k p_{ik}(1+\beta_{ik}(n))^{1+\gamma}/w^k(n)}{1-\sum_j(\beta_{ij}(n))^2},
$$
$$
\forall j: \quad w^j(n+1) = \frac{\sum_k p_{kj}(1+\beta_{kj}(n))^{1+\gamma}/w_k(n)}{1-\sum_i(\beta_{ij}(n))^2},
\tag{38}
$$

where the arguments of the β's indicate the order of the iterations. The damping parameter λ (typically chosen $0.4 \div 0.5$) helps with convergence. To ensure appropriate accuracy for solutions with

β's close to zero or unity we also insert a normalization step after Equations (37) but prior to Equations (38), making the following two transformations consequently,

$$(a) \quad \forall (i,j): \quad \beta_{ij} \to \beta_{ij} / \sum_k \beta_{ik},$$

$$(b) \quad \forall (i,j): \quad \beta_{ij} \to \beta_{ij} / \sum_k \beta_{kj}.$$

The Algorithm is sensitive to initial values for β and w. To ensure convergence, we initiate the Algorithm with the output of the MF scheme (which converges much more easily) described in Appendix C.2, that is, $\beta(0) = \beta_{MF}$ and $w(0) = v_{MF}$. Numerical experiments show that this procedure converges to a stationary point of the fractional FE (13). We also verified that the iterative scheme designed for $\gamma < 0$ converges in the $\gamma > 0$ case, even thought the former scheme is obviously faster.

Note that FFE-based version of the standard sum-product Algorithm (SPA) can be developed. It is also natural to expect, in view of the general convexity of the fractional FE discussed in the main body of the text, that there exists a provably convergent version of the SPA. It will be important to design such a convergent γ-SPA in the future and to compare its practical performance against one of the heuristics described above.

Appendix E. BP Gives Lower Bound on the Permanent

Here we give our version of the proof of the lower bound (16). First of all, in the case when the Bethe FE reaches its minimum in the interior of the domain, that is, at $\beta \in \mathcal{B}_p$, Equation (16) follows directly from the main result of Watanabe and Chertkov (2010), that is, Equation (7), and Schijver's inequality (17). Therefore, according to explanations of Section 3.1.1, we only need to analyze the case when the minimum of the Bethe FE is a partially resolved solution, with a β which can be split by appropriate permutations of rows and columns of the matrix into a perfect matching block (corresponding to a corner of the respective projected polytope), the block with all elements smaller than unity and nonzero unless the respective element of p is zero (thus lying in the interior of the respective subspace), and all cross elements of β (between the blocks) equal to zero. Then, Z_{o-BP} for such a partially resolved solution is split into the product of two contributions, $Z_{o-BP} = Z_{pm} \cdot Z_{int}$, where Z_{pm} corresponds to the perfect matching block, and Z_{int} corresponds to the interior block. In fact, Z_{pm} is equal to the weighted perfect matching block of p and $-\log(Z_{int})$ corresponds to the minimum of the Bethe FE computed for the interior block of p. On the other hand the full partition function, Z, can be bounded from below by the product $Z \geq Z_1 \cdot Z_2$, where Z_1 and Z_2 are permanents of the first and second blocks of the original matrix p. (Thus contributions of all the cross-terms of p into Z are ignored.) However, $Z_1 \geq Z_{pm}$, as counting only one perfect matching (and ignoring others), and $Z_2 \geq Z_{int}$ in accordance to what was already shown above for any minimum of Bethe FE achieved in the interior of the respective domain.

Appendix F. Pruning of the Matrix

Computing the permanent of sufficiently dense matrices exactly with the ZDD approach explained in Appendix G is infeasible for $n > 30$. To overcome this difficulty we choose to sparsify dense matrices generated in one of our experimental ensembles, removing their less significant entries in the following steps. First, we use LP, described in Appendix A.2, to find the permutation corresponding to the maximum perfect matching. To avoid getting a zero permanent in the result, we include

all components of the maximum perfect matching permutation in the pruned matrix. Second, we consider every other entry of the matrix (not contributing to the maximum perfect matching) and keep it in the matrix only if it is included in a perfect matching which is close to the maximum perfect matching, that is, the two permutations share all but a few of their entries and the ratio of their weighted contributions (in the permanent) is larger than a pre-defined value. Then, we act according to either of the two strategies, both of which are explored in this manuscript. One strategy is to include all permutations whose products are more than a given fraction of the main permutation. This method will tend to reduce the fluctuations in the error of the pruned matrix (i.e., will reduce the variation in $Z_{\text{pruned}}/Z_{\text{original}}$). The other method is to always prune a set fraction of entries from the matrix, and prune them in order of decreasing value as determined by the above criterion. This method will reduce the fluctuations in the runtime of the Algorithm.

Appendix G. Zero-suppressed Binary Decision Diagrams Method

Zero-suppressed binary decision diagrams, or ZDDs, are a tool useful for representing combinatorial problems. The concept was introduced by Shin-Ichi Minato in 1993 (Minato, 1993). The idea of ZDD is as follows: if one defines a combinatorial problem to be a function of many variables, each taking values in $\{0, 1\}$, with the value of the function itself being also in $\{0, 1\}$, then those sequences of inputs that lead to unity can be thought of as "solutions" to the problem. Furthermore, each solution can be described in terms of the input variables within it that are equal to unity. The problem, then, can be described as being a "family of sets," or set of sets, where the family is of all solutions to the problem and each set within the family is the set of input variables whose value is 1 in that solution.

To give an example of the "family of sets" concept, consider the $\overline{\text{XOR}}$ function, which returns 1 if and only if the inputs are equal. This function can also be represented as the family of sets $\{\emptyset, \{1, 2\}\}$, where 1 and 2 correspond to inputs 1 and 2 to the function, because if the function is to have value 1 then either both inputs must be equal to 1 or neither must be. Once this has been understood, it is best to see the ZDD as nothing more than a concise representation of this family of sets, since the family can get quite cumbersome for problems with many solutions and many variables. Note that this system of representing problems provides the greatest improvement when there are few solutions, and when the solutions themselves are sparse, since the family of sets is then small. Correspondingly, ZDDs are most efficient under these conditions.

The actual format of a ZDD is that of a directed tree of nodes, with each node having a directed edge to two other nodes. Each edge emanating from a node has an identity, in that it is either a high "HI" branch or a low "LO" branch, and of the two edges emanating from each node, there must be exactly one "HI" branch and one "LO" branch. Each node also has an identity, a number from 1 to n if there are n inputs to the combinatorial problem. The tree must contain one or two special nodes, or "sinks", one of which is the "true" sink, and optionally the "false" sink. We also introduce the conventions that nodes can only point to nodes of higher identity than themselves and that no two nodes can be identical in both their identity and their LO and HI pointers.

Each node in a ZDD represents a choice about the variable the node identifies. If one begins at the top node of a ZDD, taking the HI branch represents including the variable represented by the node's identity in a prospective solution, and taking the LO branch represents not including that variable. If a LO or HI branch points to the true sink, that implies that a solution is reached if and only if all variables with identity greater than the current node identity are not included. If a
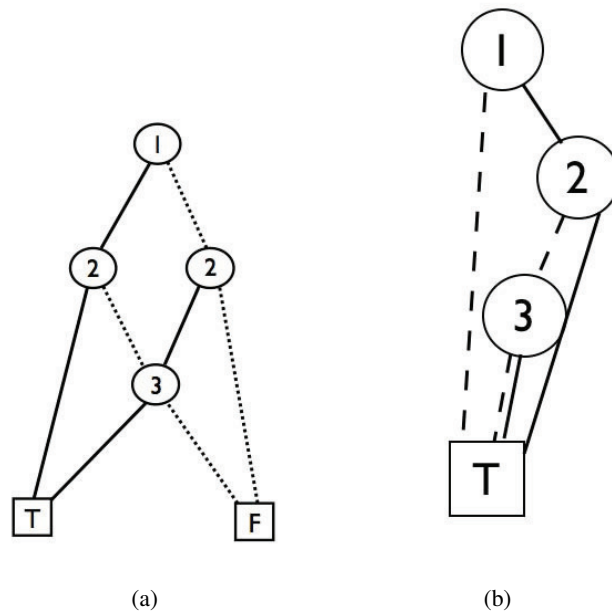
(a)   (b)

Figure 13: Two simple ZDD diagrams discussed in the text.

LO branch points to the false sink, that implies that no solution is possible given the choices made previously. Interestingly, the constraints introduced in the paragraph previous to this one imply that a HI branch can never point to the False sink.

ZDDs are best understood with examples. The first example, also illustrated in Figure 13(a), is of the ZDD for the exactly-two function of three variables, in other words, the function that returns 1 if exactly two of its three inputs have value 1 and 0 otherwise. It can also be described as the family of sets $\{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$. Here, a dotted line denotes a LO branch and a normal line denotes a HI branch. Furthermore, the T and F symbols denote the true and false sinks, respectively, and the numbers inside each node refer to the nodes' identities (the variables that they represent). Our second example, shown in Figure 13(b), represents the family of sets $\{\emptyset, \{1\}, \{1,2\}, \{1,3\}\}$. Note the absence of a False sink. Note, also, the fact that a node's HI and LO branches need not point to the different locations. A more in-depth exploration of the ZDD concept can be found in Knuth (2009).

Once the basic concept of ZDDs is introduced, one can use it for solving various combinatorial problems, for example, to represent a permanent as a ZDD in order to use the method. When we apply ZDDs to the computations of permanent, we classify each entry of the matrix as either zero or nonzero. Then, we define a variable for each nonzero entry in the matrix. Each solution of our resulting ZDD will represent a possible permutation, meaning a set of entries in the matrix such that exactly one entry in each row and column is included in the set. There is a recursive Algorithm, suggested in Knuth (2009), that allows for efficient counting of the solutions of the ZDD. The Algorithm is simple: the number of solutions of a ZDD rooted at a node is equal to the sum of the numbers of solutions of the ZDD rooted at the HI and LO children of that node. The True sink is defined as having 1 solution, and the False sink as having 0. Note that the number of solutions of
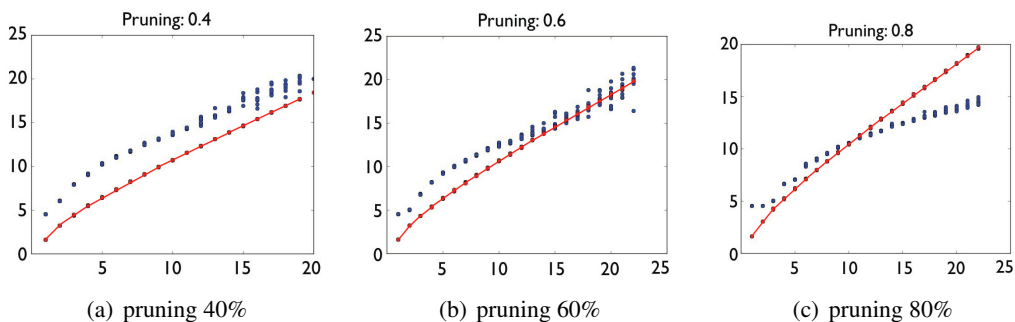
(a) pruning 40%   (b) pruning 60%   (c) pruning 80%

Figure 14: Scatter plot of the number of memory accesses required to exactly compute the perma-
nent of a sparse matrix for instances with different degree of pruning. Red and blue dots
mark results of the Ryser formula and of the ZDD-based method, respectively. The num-
bers of memory accesses required to compute Ryser's formula at each trial are closely
clumped, and so we link their results by a red line. The ZDD method's performance is
less predictable, so we do not draw a line through its scatter plot.

a ZDD representing a matrix is equal to the permanent of the corresponding to $0 - 1$ matrix, with
each 1 corresponding a nonzero entry.

In order to find the permanent of matrices that are not 0-1 matrices, only a small modification
is necessary. Instead of purely counting solutions of the ZDD, we do a weighted count, where the
weighted number of solutions of a ZDD rooted at a node is equal to the value of the corresponding
matrix entry times the weighted number of solutions at the HI child added to the weighted number
of solutions at the LO child. In other words, if we are considering a node n with children HI and
LO whose identity corresponds to a matrix entry of nonzero value v, then

$$\text{WeightedCount}(n) = v \cdot \text{WeightedCount}(\text{HI}) + \text{WeightedCount}(\text{LO}).$$

The WeightedCount of the root node of the ZDD will be equal to the permanent of the corresponding
matrix.

This leaves the question of how to build the ZDD from the matrix. This is done using Knuth's
"melding" Algorithm. The Algorithm is somewhat complex and will not be described here, but it is
described in detail in Knuth (2009). The melding Algorithm is an efficient and systematic method
for constructing larger ZDD out of the logical combination of smaller ones. The smallest ZDD
being melded together using Knuth's Algorithm are ZDD representing the "exactly-one" constraint
for each row and column of the matrix; in other words, they are constraints requiring exactly one
matrix entry in every row and column to be included in a permutation which will be a "solution" to
our problem.

Appendix H. Comparison of Ryser's Formula with the ZDD-based Method

As part of our experiments we compared the speed of Ryser's formula with the speed of the ZDD-
based method by counting memory accesses in each of the two Algorithms in order to fairly compare
them. We found that the values we got for memory accesses were strongly correlated with the
actual speed of the Algorithm. We found that for very dense matrices, Ryser's formula is faster,

but for sparser matrices the ZDD-based method is faster. We performed experiments with matrices that were 20%, 40%, and 60% sparse in order to get a good idea of the point where the ZDD-based method starts outperforming Ryser's formula. (Naturally, with no pruning, Ryser's formula outperforms the ZDD-based method significantly.)

As can be seen from Figure 14, the ZDD-based method begins outperforming Ryser's formula when matrices are around 60% sparse.

References

M. Bayati and C. Nair. A rigorous proof of the cavity method for counting matchings. In *Proc. 44th Allerton Conf. on Communications, Control, and Computing*, 2006.

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.

D.P. Bertsekas. Auction algorithms for network flow problems: a tutorial introduction. *Comput. Optimiz. Applic.*, 1:7–66, 1992.

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

G. Birkhoff. Three observations on linear algebra. *Univ. Nac. Tacuman Rev. Ser. A*, 5:147–151, 1946.

M. Chertkov. Exactness of belief propagation for some graphical models with loops. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10016 (12pp), 2008. URL http://stacks.iop.org/1742-5468/2008/P10016.

M. Chertkov and V. Y. Chernyak. Loop calculus in statistical physics and information science. *Phys. Rev. E*, 73(6):065102, Jun 2006a. doi: 10.1103/PhysRevE.73.065102.

M. Chertkov and V.Y. Chernyak. Loop series for discrete statistical models on graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(06):P06009, 2006b. URL http://stacks.iop.org/1742-5468/2006/i=06/a=P06009.

M. Chertkov, L. Kroc, and M. Vergassola. Belief propagation and beyond for particle tracking. *arxiv*, abs/0806.1199, 2008.

M. Chertkov, L. Kroc, F. Krzakala, M. Vergassola, and L. Zdeborova. Inference in particle tracking experiments by passing messages between images. *Proceedings of the National Academy of Science*, 107:7663–7668, April 2010. doi: 10.1073/pnas.0910994107.

B. Cseke and T. Heskes. Properties of Bethe Free Energies and message passing in Gaussian models. *J. Artif. Intell. Res. (JAIR)*, 41:1–24, 2011.

K. Drescher, R. E. Goldstein, N. Michel, M. Polin, and I. Tuval. Direct measurement of the flow field around swimming microorganisms. *Phys. Rev. Lett.*, 105(16):168101, Oct 2010. doi: 10.1103/PhysRevLett.105.168101.

P.V. Egorychev. Proof of the van der Waerden conjecture for permanents. *Siberian Mathematical Journal*, 22(6):854–859, 1981. URL http://www.springerlink.com/content/k692377516k1x778/.

G. M. Engel and H. Schneider. Inequalities for determinants and permanents. *Linear and Multilinear Algebra*, 1:187–201, 1973.

D.I. Falikman. Proof of the van der Waerden conjecture regarding the permanent of a doubly stochastic matrix. *Mathematical Notes*, 29(6):475–479, 1981. URL http://www.springerlink.com/content/h41162g677317110/.

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

J. S. Guasto, K. A. Johnson, and J. P. Gollub. Oscillatory flows induced by microorganisms swimming in two dimensions. *Phys. Rev. Lett.*, 105(16):168102, Oct 2010. doi: 10.1103/PhysRevLett.105.168102.

L. Gurvits. Van der Waerden/Schrijver-Valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all. *Electronic Journal of Combinatorics*, 15:R66, 2008. URL http://www.emis.ams.org/journals/EJC/Volume_15/PDF/v15i1r66.pdf.

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

B. Huang and T. Jebara. Approximating the permanent with belief propagation, arxiv:0908.1769, 2009. URL http://arxiv.org/abs/0908.1769.

M. Huber and J. Law. Fast approximation of the permanent for very dense problems. *in: SODA 08: Proc. 19th ACM-SIAM Sympos. on Discrete Algorithms*, page 681689, 2008.

M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004. ISSN 0004-5411. doi: http://doi.acm.org/10.1145/1008731.1008738.

D. Kőnig. *Theorie der Endlichen und Unendlichen Graphen*. Akademische Verlags Gesellschaft, Leipzig, 1936.

P. Knopp and R. Sinkhorn. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific J. Math.*, 21(2):343–348, 1967.

D. E. Knuth. *The Art of Computer Programming, Volume 4, Fascicle 1: Bitwise Tricks & Techniques; Binary Decision Diagrams*. Addison-Wesley Professional, 12th edition, 2009. ISBN 0321580508, 9780321580504.

H. W. Kuhn. The Hungarian Method for the assignment problem. *Naval Research Logistics Quarterly*, 2:8397, 1955.

M. Laurent and A. Schrijver. On Leonid Gurvits's proof for permanents. *American Mathematical Monthly*, 117(10), 2010. URL http://homepages.cwi.nl/~lex/files/perma5.pdf.

N. Linial, A. Samorodnitsky, and A. Wigderson. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. *Combinatorica*, 20(4):545–568, 2000.

S. Minato. Zero-suppressed BDDs for set manipulation in combinatorial problems. In *Design Automation, 1993. 30th Conference on*, pages 272 – 277, 1993. doi: 10.1109/DAC.1993.203958.

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

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

H. J. Ryser. Combinatorial mathematics. *The Carus Mathematical Monographs*, 14, 1963.

S. Sanghavi, D. Malioutov, and A. Willsky. Belief propagation and lp relaxation for weighted matching in general graphs. *Information Theory, IEEE Transactions on*, 57(4):2203 –2212, april 2011. ISSN 0018-9448. doi: 10.1109/TIT.2011.2110170.

A. Schrijver. Counting 1-factors in regular bipartite graphs. *Journal of Combinatorial Theory*, 72: 122–135, 1998. doi: 10.1006/jctb.1997.1798.

TheCodeProject. `http://www.codeproject.com/KB/applications/ RyserPermanent.aspx`. Computes permanent of a matrix with Ryser's algorithm.

L.G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8: 189201, 1979.

B. L. van der Waerden. *[Aufgabe] 45, Jahresbericht der Deutschen Mathematiker-Vereinigung*, 35: 117, 1926.

J. von Neumann. A certain zero-sum two-person game equivalent to an optimal assignment problem. *Ann. Math. Studies*, 28:5–12, 1953.

P.O. Vontobel. The Bethe permanent of a non-negative matrix. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pages 341 –346, 29 2010-oct. 1 2010. doi: 10.1109/ALLERTON.2010.5706926.

P.O. Vontobel. The bethe permanent of a non-negative matrix. *IEEE Trans. Inf. Theory*, 59(3): 1866–1901, 2013.

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. URL `http://stacks.iop.org/1751-8121/43/i=24/a=242002`.

W. Wiegerinck and T. Heskes. Fractional belief propagation. *Advances in Neural Information Processing Systems 15*, 12:438–445, 2003.

A.B. Yedidia. Counting independent sets and kernels of regular graphs. *arxiv*, abs/0910.4664, 2009.

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, July 2005. ISSN 0018-9448. doi: 10.1109/TIT.2005.850085.