Jason L. Williams

Interior point solution of fractional Bethe permanent

Statistical Signal Processing (SSP), 2014 IEEE Workshop on, 2014 / pp.213-216

Published version at: http://dx.doi.org/10.1109/SSP.2014.6884613

INTERIOR POINT SOLUTION OF FRACTIONAL BETHE PERMANENT

Jason L. Williams

National Security and ISR Division, Defence Science and Technology Organisation, Australia
and School of Electrical and Electronic Engineering, University of Adelaide, Australia

ABSTRACT

Many combinatorial problems in fields such as object tracking involve reasoning over correspondence, e.g., calculating the probability that a measurement belongs to a particular track. Recent studies have shown that loopy belief propagation (LBP) provides a highly desirable option in the trade-off between accuracy and computational complexity in this task. LBP can be understood as a particular method for optimising the Bethe free energy (BFE). In this paper, we directly optimise the BFE using an interior point Newton method. Exploiting the structure of the constraints, we arrive at an algorithm offers improvements in computation in cases in which LBP converges very slowly. The method also solves the recently-proposed fractional free energy (FFE); we use this to demonstrate that FFE can offer marginal estimates with improved accuracy.

1. INTRODUCTION

Problems involving inference over correspondence are common in areas such as tracking, e.g., in data association, where unknown measurement-object correspondence is addressed, and in fusion of tracks from different sensors [1]. These are also referred to as weighted bipartite matching problems, since they consider configurations in which each item in one group (e.g., each track) is paired with an item in the other group (e.g., a measurement). While the most likely association can be calculated efficiently using methods such as the auction algorithm, calculation of marginal correspondence probabilities is closely related to the #P-complete problem of calculating a matrix permanent.

An emerging method for estimation of the matrix permanent and marginal correspondence probabilities (required to implement standard tracking filters such as JPDA and its many extensions [1]) using probabilistic graphical models (PGM) is examined in [2–4]. The model studied is one in which the probability of a matching of n tracks to n measurements is:

$$p(a_1,\dots,a_n) \propto \begin{cases} \prod_{i=1}^n c_{ia_i}, & \text{matching feasible} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $a_i \in \{1,\dots,n\}$ is the index of the measurement with which the i-th track is matched, i.e., $a_i = j$ if track i is matched with measurement j. A matching is feasible if each measurement is matched with at most one track. In [2], it was shown that the following optimisation problem (with $\gamma = 1$) yields the optimal solution of the Bethe free energy (BFE) [5] for a particular PGM formulation of the problem:

$$\underset{q_{ij}}{\text{minimise}} \sum_{i,j} q_{ij} \log \frac{q_{ij}}{c_{ij}} - \gamma \sum_{i,j} (1-q_{ij}) \log(1-q_{ij}) \quad (2)$$
$$\text{subject to} \sum_j q_{ij} = 1 \;\forall\, i, \quad \sum_i q_{ij} = 1 \;\forall\, j, \quad q_{ij} \geq 0$$

where q_{ij} is the belief (i.e., approximate marginal probability) that track i is matched with measurement j. Furthermore, it was shown that the objective is convex on the affine subspace in which at least one of the two sets of equality constraints is satisfied; this is not obvious, since the second term in the objective is concave. The resulting beliefs were shown empirically in [4] to provide a remarkably accurate approximation of the marginal probabilities.

It is well-known that if loopy belief propagation (LBP) converges, the result is a stationary point of the BFE [5]. Convergence of LBP in the formulation of interest was proven simultaneously in conference papers preceding [2,4]. In most practical problems, convergence is sufficiently rapid; in [4] it is shown that convergence is at least linear, with a rate determined by the problem parameters. However, in a small proportion of practical problems, convergence is slow, and an alternative is needed.

The primary motivation of [2,3] was estimation of the matrix permanent, which is effectively the normalisation constant in (1). In [3], it is shown that for any problem, there is a value $\gamma \in [0,1]$ for which the solution of (2) yields the exact value of the matrix permanent, leading to FFE, which utilises a value $\gamma < 1$.[1] It not known whether this modification yields improved beliefs, or if LBP converges in this case.

With $\gamma = 0$, (2) reduces to the problem of matrix normalisation, i.e., multiplying each row of a non-negative matrix by a constant and each column by a constant in order to obtain a doubly stochastic matrix. The most common method for this is Sinkhorn algorithm [6], which is somewhat similar to LBP. Like LBP, Sinkhorn iteration converges rapidly in most cases,

[1] Note that γ is negated in comparison to [3].

but is problematic in a small but important subset. This has yielded further study such as [7], which proposed an interior point method for optimising (2) with $\gamma = 0$. It was reported that the complexity of the approach is $O(n^6)$; consequently, the method has not been applied widely.

1.1. Contributions

In this paper, we develop an interior point method for optimising (2) for any $\gamma \in (-\infty, 1]$. As in [4], the setting is generalised to examine non-square ($n \times m$) matrices in order to admit missed detection and false alarm events (although the probability of these can be set to zero to recover the original square case). The contributions in the development include:

- Whereas the standard formulation of interior point methods assumes that the objective is convex on $\mathbb{R}^n$, the objective (2) is only convex on the subspace in which the equality constraints are satisfied. Consequently, naïve application of equality constrained Newton optimisers would fail. We show that this can be addressed by modifying the problem solved in each Newton step, applying a projection in order to obtain a problem involving a positive semi-definite Hessian. While this may be viewed as sequential quadratic programming, the insight is that the convexity that is present in the particular problem of interest can be recovered, yielding quadratic sub-problems that can be solved analytically.
- Although the problem involves nm variables and $(nm + n + m)$ constraints and direct solution of the Newton step would have complexity $O(n^3m^3)$ (as reported in [7]), we show that the structure of the constraints can be exploited to obtain a solution with complexity $O(\min\{n^2m, m^2n\})$.

Empirically, we show that convergence is rapid, never extending beyond 70 iterations even in problems for which LBP requires $15{,}000$ iterations. Finally, using the newly developed method, we examine the accuracy of the marginal estimates obtained by setting $\gamma < 1$, thus demonstrating the improvement in marginal probability estimates that can be achieved using the FFE proposed in [3].

1.2. Notation

- Vectors are denoted by lower case letters with bold text, *e.g.*, $\mathbf{q}$, and matrices by upper case bold letters, *e.g.*, $\mathbf{A}$
- The notation $\mathbf{x} = [x_i]$ indicates that $\mathbf{x}$ is a column vector for which the i-th entry is the scalar x_i
- The notation ';' refers to vertical concatenation, e.g., $[\mathbf{x}_1;\ \mathbf{x}_2] = [\mathbf{x}_1^T\ \mathbf{x}_2^T]^T$
- The Kronecker product is denoted by $\otimes$
- The notation $\mathbf{1}_{N\times M}$ refers to the $N \times M$ matrix for which every element has the value one (similarly for $\mathbf{0}_{N\times M}$)
- The $N \times N$ identity matrix is denoted by $\mathbf{I}_N$

2. FORMULATION

This paper considers the following optimisation problem, of finding q_{ij} given input data c_{ij}, where $i \in \{0,\dots,n\}$ and $j \in \{0,\dots,m\}$:

$$\begin{aligned}\underset{q_{ij}}{\text{minimise}}\ &\sum_{i=1}^{n}\sum_{j=0}^{m} q_{ij}\log\frac{q_{ij}}{c_{ij}} + \gamma\sum_{j=1}^{m} q_{0j}\log\frac{q_{0j}}{c_{0j}} \\ &- \gamma\sum_{i=1}^{n}\sum_{j=1}^{m}(1-q_{ij})\log(1-q_{ij}) \qquad (3)\end{aligned}$$

$$\text{subject to } \textstyle\sum_{j=0}^{m} q_{ij} = 1\ \forall\, i \in \{1,\dots,n\} \qquad (4)$$

$$\textstyle\sum_{i=0}^{n} q_{ij} = 1\ \forall\, j \in \{1,\dots,m\} \qquad (5)$$

$$q_{ij} \geq 0\ \forall\, i,j, \quad q_{00} = c_{00} \qquad (6)$$

With $\gamma = 1$, this can be shown to be the extension of the formulation of [2] to non-square problems incorporating missed detections and false alarms (i.e., the formulation studied in [4]). The weighting of the second term in the objective by γ ensures that the correct solution is obtained in problems involving well-spaced tracks when $\gamma < 1$.

In vector form ($\mathbf{q} = [\mathbf{q}_0;\ \mathbf{q}_1;\ \dots;\ \mathbf{q}_m]$, where $\mathbf{q}_j = [q_{ij}]_{i=0}^{n}$), the equality constraints (4) and (5) can be respectively written as:

$$\mathbf{A}_1\mathbf{q} = \mathbf{1}_{n\times 1}, \quad \mathbf{A}_2\mathbf{q} = \mathbf{1}_{m\times 1} \qquad (7)$$

$$\begin{aligned}\mathbf{A}_1 &= \mathbf{1}_{1\times(m+1)} \otimes \begin{bmatrix} \mathbf{0}_{n\times 1} & \mathbf{I}_n \end{bmatrix} \qquad (8)\\ &= \begin{bmatrix} \mathbf{0}_{n\times 1} & \mathbf{I}_n & \dots & \mathbf{0}_{n\times 1} & \mathbf{I}_n \end{bmatrix}\end{aligned}$$

$$\begin{aligned}\mathbf{A}_2 &= \begin{bmatrix} \mathbf{0}_{m\times 1} & \mathbf{I}_m \end{bmatrix} \otimes \mathbf{1}_{1\times(n+1)} \qquad (9)\\ &= \begin{bmatrix} \mathbf{0}_{1\times(n+1)} & \mathbf{1}_{1\times(n+1)} & \dots & \mathbf{0}_{1\times(n+1)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1\times(n+1)} & \mathbf{0}_{1\times(n+1)} & \dots & \mathbf{1}_{1\times(n+1)} \end{bmatrix}\end{aligned}$$

These constraints can be combined into a single matrix constraint:

$$\mathbf{A}\mathbf{q} = \mathbf{1}_{(n+m)\times 1}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_1; & \mathbf{A}_2 \end{bmatrix} \qquad (10)$$

For numerical convenience, we reformulate the optimisation via a linear transformation, $\mathbf{q} = \mathbf{C}\mathbf{f}$ where $\mathbf{C}$ is the diagonal matrix such that $q_{ij} = c_{ij}f_{ij}$. In vector form, the equality constraints can be expressed as:

$$\tilde{\mathbf{A}}\mathbf{f} = \begin{bmatrix} \tilde{\mathbf{A}}_1 \\ \tilde{\mathbf{A}}_2 \end{bmatrix}\mathbf{f} = \begin{bmatrix} \mathbf{A}_1\mathbf{C} \\ \mathbf{A}_2\mathbf{C} \end{bmatrix}\mathbf{f} = \mathbf{1}_{(n+m)\times 1} \qquad (11)$$

3. NEWTON OPTIMISATION

Solving (3) using a Newton-based method is complicated by the fact that the objective is only convex on the linear subspace in which at least one of the two sets of equality constraints (4), (5) are satisfied. Accordingly, both sets of constraints cannot be relaxed (i.e., in order to address them via

Lagrangian methods). Incorporating a log barrier to enforce the non-negativity constraints (6), we arrive at a barrier function:

$$B_\theta(\mathbf{f}) = \sum_{i=1}^{n}\sum_{j=0}^{m} c_{ij} f_{ij} \log f_{ij} + \gamma \sum_{j=1}^{m} c_{0j} f_{0j} \log f_{0j}$$
$$- \gamma \sum_{i=1}^{n}\sum_{j=1}^{m} (1 - c_{ij} f_{ij}) \log(1 - c_{ij} f_{ij}) - \theta \sum_{i=0}^{n}\sum_{j=0}^{m} \log f_{ij} \tag{12}$$

to which the equality constraints (4) and (5) must still be applied. Applying Newton's method, we minimise a second-order Taylor series approximation to this function about a nominal $\mathbf{f}$:

$$\underset{\delta\mathbf{f}}{\text{minimise}}\ B_\theta(\mathbf{f}) + \mathbf{g}^T\delta\mathbf{f} + \frac{1}{2}\delta\mathbf{f}^T\mathbf{H}\delta\mathbf{f} \tag{13}$$
$$\text{subject to } \tilde{\mathbf{A}}\delta\mathbf{f} = -\mathbf{r} \tag{14}$$

where $\mathbf{g} = \nabla B_\theta(\mathbf{f})$ is the vector gradient of B_θ at $\mathbf{f}$, and $\mathbf{H} = \nabla^2 B_\theta(\mathbf{f})$ is the Hessian; these can be evaluated as $\mathbf{H} = \operatorname{diag}\mathbf{h}$, where $\mathbf{g}$ (and $\mathbf{h}$) is comprised of elements g_{ij} similar to $\mathbf{q}$, and for $i > 0, j > 0$,

$$g_{ij} = c_{ij}\{\log f_{ij} + \gamma \log[1 - c_{ij} f_{ij}] + (1+\gamma)\} - \frac{\theta}{f_{ij}}$$
$$h_{ij} = \frac{c_{ij}}{f_{ij}} - \frac{\gamma c_{ij}^2}{1 - c_{ij} f_{ij}} + \frac{\theta}{f_{ij}^2}$$

and $g_{i0} = c_{i0}(\log f_{i0} + 1) - \theta/f_{i0}$, $h_{i0} = c_{i0}/f_{i0} + \theta/f_{i0}^2$, $g_{0j} = \gamma c_{0j}(\log f_{0j} + 1) - \theta/f_{0j}$, $h_{0j} = \gamma c_{0j}/f_{0j} + \theta/f_{0j}^2$. The residual is given by $\mathbf{r} = \tilde{\mathbf{A}}\mathbf{f} - \mathbf{1}_{(n+m)\times 1} = [\mathbf{r}_1;\ \mathbf{r}_2]$. We assume that the initial point is feasible with respect to at least one of the two sets of constraints (4), (5) (i.e., either $\mathbf{r}_1 = \mathbf{0}$ or $\mathbf{r}_2 = \mathbf{0}$) in order to ensure that the convexity result of [2] applies.

The difficulty of solving (13) is that the matrix $\mathbf{H}$ is not PSD, thus the regular solution handling the constraint (14) via Lagrangian methods is problematic. However, we will show that this difficulty can be overcome by projecting onto the feasible subspace. To proceed, let $\mathbf{P}$ be the matrix that projects onto the null space of $\tilde{\mathbf{A}}$:

$$\mathbf{P} = \mathbf{I} - \tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)^{-1}\tilde{\mathbf{A}} \tag{15}$$

Subsequently, the constraint $\tilde{\mathbf{A}}\delta\mathbf{f} = -\mathbf{r}$ will be satisfied if and only if $\delta\mathbf{f} = \mathbf{P}\delta\mathbf{f} - \tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)^{-1}\mathbf{r}$. Substituting this identity into the objective, dropping constant terms and simplifying, we find the equivalent problem:

$$\underset{\delta\mathbf{f}}{\text{minimise}}\ [\mathbf{g} - \mathbf{H}\tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)^{-1}\mathbf{r}]^T\mathbf{P}\delta\mathbf{f} + \frac{1}{2}\delta\mathbf{f}^T\mathbf{PHP}\delta\mathbf{f} \tag{16}$$
$$\text{subject to } \tilde{\mathbf{A}}\delta\mathbf{f} = -\mathbf{r} \tag{17}$$

This clearly does not change the location of the solution, since the objective is unchanged within the feasible set (other than an additive constant). The key difference is that whereas the matrix $\mathbf{H}$ is not PSD, the projection $\mathbf{PHP}$ is, as the following theorem shows.

Theorem 1. *Suppose $J(\mathbf{f}) : \mathcal{F} \to \mathbb{R}$ is convex on an affine subset $\mathcal{A} = \{f \in \mathcal{F} | \mathbf{A}\mathbf{f} = \mathbf{b}\}$ of the convex set $\mathcal{F}$, and J is twice differentiable at a point $f \in \mathcal{A}$ in the interior of $\mathcal{F}$. Then $\mathbf{PHP} \succeq 0$, where $\mathbf{H}$ is the Hessian of J at f, and $\mathbf{P}$ is the matrix that projects onto the null space of $\mathbf{A}$.*

Proof. Given some fixed $\mathbf{f} \in \mathcal{A}$, any vector $\mathbf{y} \in \mathcal{A}$ can be written as $\mathbf{f} + \mathbf{P}\delta\mathbf{f}$. Let $\tilde{J}(\delta\mathbf{f}) = J(\mathbf{f} + \mathbf{P}\delta\mathbf{f})$. Since $\mathbf{f}$ is in the interior of $\mathcal{F}$, $\mathbf{f} + \mathbf{P}\delta\mathbf{f} \in \mathcal{A}$ for sufficiently small $|\delta\mathbf{f}|$. Since $\tilde{J}$ is convex on the neighbourhood around zero, $\nabla^2\tilde{J} \succeq 0$. By the vector chain rule, $\nabla^2\tilde{J} = \mathbf{P}(\nabla^2 J)\mathbf{P} = \mathbf{PHP}$. □

Consequently, we can solve (16) using a standard Lagrangian method (via convex duality) as: [8, p532]

$$\begin{bmatrix} \mathbf{PHP} & \tilde{\mathbf{A}}^T \\ \tilde{\mathbf{A}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \delta\mathbf{f} \\ \nu \end{bmatrix} = \begin{bmatrix} -\tilde{\mathbf{g}} \\ -\mathbf{r} \end{bmatrix} \tag{18}$$

where $\tilde{\mathbf{g}} = \mathbf{Pg} - \mathbf{PH}\tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)^{-1}\mathbf{r}$. Algebraic manipulations result in the solution $\nu = \mathbf{0}$, and

$$\delta\mathbf{f} = -\mathbf{H}^{-1}[\mathbf{I} - \tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\mathbf{H}^{-1}\tilde{\mathbf{A}}^T)^{-1}\tilde{\mathbf{A}}\mathbf{H}^{-1}]\mathbf{Pg}$$
$$- \mathbf{H}^{-1}\tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\mathbf{H}^{-1}\tilde{\mathbf{A}}^T)^{-1}\mathbf{r} \tag{19}$$

The solution obtained using the standard method (ignoring the fact that $\mathbf{H}$ is not PSD) simply replaces $\mathbf{Pg}$ with $\mathbf{g}$; the impact of correcting for the non-convexity is pre-projection of the gradient onto the feasible subspace. This is necessary for cases in which the objective is only convex on the affine subspace defined by the constraints (i.e., when $\gamma > 0$).

When $\mathbf{A}$ is as given in (10), the expression $(\tilde{\mathbf{A}}\mathbf{H}^{-1}\tilde{\mathbf{A}}^T)^{-1}\mathbf{b}$ (where $\mathbf{b} = \tilde{\mathbf{A}}\mathbf{H}^{-1}\mathbf{Pg} - \mathbf{r}$)) can be found as the solution of

$$\begin{bmatrix} \tilde{\mathbf{A}}_1\mathbf{H}^{-1}\tilde{\mathbf{A}}_1^T & \tilde{\mathbf{A}}_1\mathbf{H}^{-1}\tilde{\mathbf{A}}_2^T \\ \tilde{\mathbf{A}}_2\mathbf{H}^{-1}\tilde{\mathbf{A}}_1^T & \tilde{\mathbf{A}}_2\mathbf{H}^{-1}\tilde{\mathbf{A}}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \tag{20}$$

Algebraic manipulations reveal the solution as

$$\mathbf{x}_2 = \mathbf{G}^{-1}[\mathbf{b}_2 - \tilde{\mathbf{A}}_2\mathbf{H}^{-1}\tilde{\mathbf{A}}_1^T(\tilde{\mathbf{A}}_1\mathbf{H}^{-1}\tilde{\mathbf{A}}_1^T)^{-1}\mathbf{b}_1] \tag{21}$$
$$\mathbf{G} = \tilde{\mathbf{A}}_2\mathbf{H}^{-1}\tilde{\mathbf{A}}_2^T - \tilde{\mathbf{A}}_2\mathbf{H}^{-1}\tilde{\mathbf{A}}_1^T(\tilde{\mathbf{A}}_1\mathbf{H}^{-1}\tilde{\mathbf{A}}_1^T)^{-1}\tilde{\mathbf{A}}_1\mathbf{H}^{-1}\tilde{\mathbf{A}}_2^T \tag{22}$$
$$\mathbf{x}_1 = (\tilde{\mathbf{A}}_1\mathbf{H}^{-1}\tilde{\mathbf{A}}_1^T)^{-1}[\mathbf{b}_1 - \tilde{\mathbf{A}}_1\mathbf{H}^{-1}\tilde{\mathbf{A}}_2^T\mathbf{x}_2] \tag{23}$$

The complexity of calculating $\mathbf{G}$ is $O(m^2n)$, and the complexity of inverting $\mathbf{G}$ is $O(m^3)$ (it can easily be shown that $\tilde{\mathbf{A}}_1\mathbf{H}^{-1}\tilde{\mathbf{A}}_1^T$ and $\tilde{\mathbf{A}}_2\mathbf{H}^{-1}\tilde{\mathbf{A}}_2^T$ are diagonal). If $m > n$, the rows of (20) can be reversed so that the overall complexity is $O(\min\{m^2n, n^2m\})$. A similar approach can be used to calculate the projection $\mathbf{Pg}$.

4. RESULTS

We evaluate the proposed method using the experiments described in [4], examining a 2×3 grid of regularly spaced objects, where the spacing is varied from 0 to 10 units to observe the impact of target interaction. Track estimates are initialised by simulating the observation process (with known association) for 30 time steps. The probability of detection is set to $P_\text{d} = 0.98$, and false alarm densities of $\lambda_\text{fa} \in \{10^{-2}, 10^{-4}, 10^{-6}\}$ are considered; remaining problem parameters follow the baseline case described in [4] (as discussed in [4], the most challenging cases for convergence and accuracy of LBP are those with high P_d and/or low λ_fa). In each condition, 1000 Monte Carlo (MC) trials are executed, yielding a total over all conditions of 153,000 MC trials for each algorithm. The exact marginal association distribution for each target is calculated using a junction tree method as described in [4].

Results are shown in Figure 1. As expected, LBP obtains an identical result to the Newton-based method with $\gamma = 1$. The computation time of the Newton-based method is essentially unaffected by the problem conditions, whereas the complexity of LBP increases by two orders of magnitude in the lowest λ_fa case (which suffers from slow convergence), resulting in a computation time $10\times$ that of the Newton method. In the low λ_fa cases, choosing $\gamma < 1$ also improves the accuracy of the marginal estimates. In the lowest λ_fa case, the marginal estimates obtained with $\gamma = 1$ have KL divergences of over 0.1, whereas with $\gamma = 0.7$, they are less than 0.01.

5. CONCLUSION

This paper has demonstrated how an interior point method can be used to estimate association probabilities using the BFE formulation of [2–4]. The proposed method complements LBP, providing an alternative that can be utilised in cases in which convergence of LBP is problematic. We have also demonstrated that the fractional free energy proposed by [3] can considerably improve the accuracy of the resulting estimates in high SNR cases. Future work includes devising an approach to automatically select the value of γ.

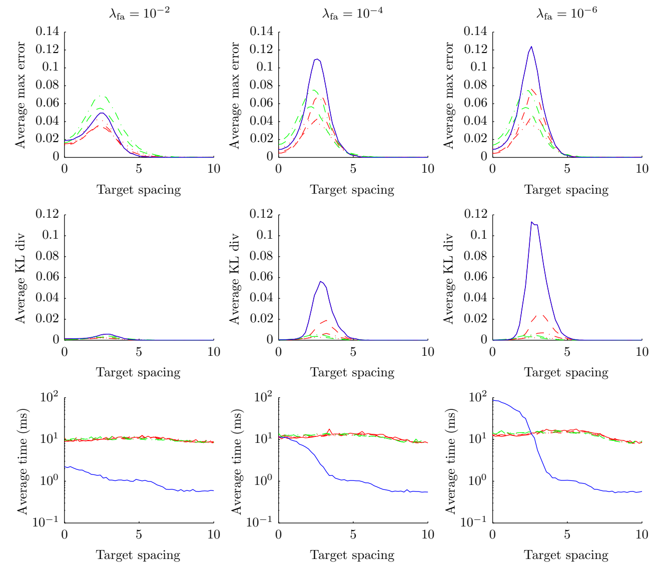

Fig. 1. Results of experiments. Each column of figures corresponds to a different value of λ_fa. Top row shows average maximum error in the marginal association distribution for each target (averaged over targets and MC trials). Middle row shows average KL divergence between the exact marginal distribution for the target and the estimate (again, averaged over targets and MC trials). Bottom row shows average computation time. The LBP approach of [4] is shown in blue. Other lines show the proposed method with $\gamma = 1.0$ (solid red, coincident with the blue line in the first two rows), $\gamma = 0.9$ (dashed red), $\gamma = 0.8$ (dot-dashed red), $\gamma = 0.7$ (dotted red), $\gamma = 0.6$ (dashed green) and $\gamma = 0.5$ (dot-dashed green).

6. REFERENCES

[1] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. Storrs, CT: YBS Publishing, 2011.

[2] P. Vontobel, "The Bethe permanent of a nonnegative matrix," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1866–1901, 2013.

[3] M. Chertkov and A. B. Yedidia, "Approximating the permanent with fractional belief propagation," *Journal of Machine Learning Research*, vol. 14, pp. 2029–2066, 2013.

[4] J. L. Williams and R. A. Lau, "Approximate evaluation of marginal association probabilities with belief propagation," *Accepted for publication, IEEE Trans. Aerosp. Electron. Syst.*, 2014. [Online]. Available: http://arxiv.org/abs/1209.6299

[5] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," *Exploring artificial intelligence in the new millennium*, pp. 239–269, 2003.

[6] R. Sinkhorn, "A relationship between arbitrary positive matrices and doubly stochastic matrices," *The Annals of Mathematical Statistics*, vol. 35, no. 2, pp. 876–879, June 1964.

[7] H. Balakrishnan, I. Hwang, and C. Tomlin, "Polynomial approximation algorithms for belief matrix maintenance in identity management," in *Proc. 43rd IEEE Conference on Decision and Control*, vol. 5, 2004, pp. 4874–4879.

[8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.