On the Relationship Between Belief Propagation Decoding and Joint Maximum Likelihood Detection

John MacLaren Walsh, Member, IEEE, and Phillip Allan Regalia, Fellow, IEEE

Abstract—Belief propagation, via a novel reinterpretation of the Bethe free energy’s pseudo-dual, is shown to be related to a novel relaxation of maximum likelihood detection via a constrained optimization. The conventional maximum likelihood detection falls out for a zero constraint, and belief propagation’s fixed points are obtained for other constraint values.

Index Terms—belief propagation, factor graphs, probabilistic decoding, turbo decoding, iterative decoding

I. INTRODUCTION

The belief propagation method of statistical inference [1] is known to converge to the marginal a posteriori probabilities in cycle-free factor graphs. Additionally, the observed behavior in many applications involving factor graphs with cycles, such as to turbo decoding [2], [3] and soft low-density parity-check (LDPC) decoding [4], [5] exhibits detections with excellent performance. One important body of work aiming to explain this performance in factor graphs with cycles reveals that the fixed points of belief propagation minimize the Bethe variational free energy [6], [7], [8], a celebrated approximation in statistical physics. In this paper, extending our results for the turbo decoder [9], [10], [11] to the more general belief propagation family of algorithms, we use a modification of the pseudo-dual to the Bethe variational free energy to connect the fixed points of belief propagation to a different constrained optimization problem. This new problem has the benefit of providing a direct parameterizable connection between belief propagation detection and joint maximum likelihood detection. In particular, a constraint value of zero yields the joint maximum likelihood detection, while other constraint values dictated by a Lagrange multiplier yield the fixed points of belief propagation. Simulations using an LDPC decoder then show the constraint values are typically close to 0 when the belief propagation algorithm is yielding estimates with good performance, suggesting that the algorithm is performing well because its estimates are close to the joint maximum likelihood detection.

II. INFORMATION GEOMETRY PREREQUISITES

Consider random vectors $X := \{X_1, \ldots, X_M\}^T$, taking values in the Cartesian product space $\mathcal{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_M$, with cardinalities $|\mathcal{X}_i| = L_i$. This section details the relationship between the family of joint distributions $p_X(X)$, the marginal distributions $\{p_{X_i}(X_i)\}$, and the set of joint distributions which factor into the product of their marginals $p_X(X) = \prod_i p_{X_i}(X_i)$, in terms of their information geometric coordinates [12].

To begin, number the elements of the set $\mathcal{X}$ from 0 to $L - 1$, so that

$$\mathcal{X} := \{x(0), x(1), \ldots, x(L-1)\}, \quad L := \prod_i L_i$$

For convenience, we choose the order based on enumerations of the sets $\mathcal{X}_i := \{x_i(0), \ldots, x_i(L_i - 1)\}$, which advance like wheels on an odometer, so that

$$x(0) = (x_1(0), \ldots, x_{M-1}(0), x_M(0))$$

$$x(1) = (x_1(0), \ldots, x_{M-1}(0), x_M(1))$$

$$\vdots$$

$$x(L_{M-1}) = (x_1(0), \ldots, x_{M-1}(0), x_M(L_{M-1}))$$

$$x(L_M) = (x_1(0), \ldots, x_{M-1}(1), x_M(0))$$

$$\vdots$$

$$x(L - L_i) = (x_1(L_i - 1), \ldots, x_{M-1}(L_{M-1} - 1), x_M(0))$$

$$\vdots$$

$$x(L - 1) = (x_1(L_i - 1), \ldots, x_M(L_{M-1}))$$

Placing the corresponding probabilities $\Pr[X = x(i)]$ for all but the zeroth enumerated element into a vector, we may form expectation coordinates [12] for the set of probability distributions on $\mathcal{X}$:

$$q_X := [\Pr[X = x(1)], \ldots, \Pr[X = x(L - 1)]]^T$$

Now, the relationship between the joint distribution and its marginal distributions is linear, viz.

$$p_{X_i}(x_i(j)) = \sum_{k=0}^{L-1} 1\{[x(k)]_i = x_i(j)\} p_X(x(k)),$$

for $j \in \{0, \ldots, L_i - 1\}$ where 1\{\cdot\} denotes the indicator function which equals one (resp., zero) when its argument is true (resp., false) and $[x]_i$ denotes the $i$th element of the vector x. We may thus write the expectation coordinates $q_{X_i}$ for $p_{X_i}$ as a linear map of the expectation coordinates for $p_X$, i.e., $q_{X_i} = C_i q_X$ where the matrix $C_i$ simply has rows made from the indicator function for $L_i - 1$ different values of $x_i$:

$$[C_i]_{j,k} := 1\{[x(k)]_i = x_i(j)\},$$
for \( j, k \in \{1, \ldots, L-1\} \), and \( i \in \{1, \ldots, M\} \). The same set of matrices \( \{C_i, i \in \{1, \ldots, M\}\} \) may be used to form the log coordinates \( \theta \) for a product density

\[
p_{X}(x) = \prod_{i=1}^{M} p_{X_i}(x_i)
\]

from the log coordinates \( \theta_i \) of its marginal densities [12]:

\[
\theta = \sum_{i=1}^{M} C_i \theta_i, \quad [\theta]_k := \log \left( \frac{\Pr[X = x(k)]}{\Pr[X = x(0)]} \right),
\]

\[
[\theta]_j := \log \left( \frac{\Pr[X_i = x_i(j)]}{\Pr[X_i = x_i(0)]} \right)
\]

for \( k \in \{1, \ldots, L-1\}, \ j \in \{1, \ldots, L_i-1\}, \ i \in \{1, \ldots, M\} \).

Finally, the two ways of parameterizing a probability distribution via its expectation coordinates \( q_X \) and its log coordinates \( \theta \) arise from the Legendre transforms between two potential functions, the log partition function and the (negative) of the Shannon entropy [12]:

\[
\psi(\theta) := \log (1 + \| \exp(\theta) \|_1),
\]

\[
H(q) := -\langle q, \log(q) \rangle - (1 - 1^T q) \log(1 - 1^T q).
\]

Here \( \exp \) and \( \log \) are understood to operate element-wise on vector arguments. Let \( r \) and \( q \) be two probability distributions with expectation coordinates and log coordinates \( r, q \) and \( \rho, \theta \) respectively. The entropy \( H(r) \) and log partition function \( \psi(\theta) \) are related via

\[
\psi(\theta) - H(r) = \langle r, \theta \rangle + D(r\|q),
\]

where \( D \) is the relative entropy (Kullback-Leibler divergence), implying that

\[
\psi(\theta) - H(r) \geq \langle r, \theta \rangle
\]

with equality if and only if \( r = q \).

Due to the Legendre transform relationship, the gradients of these two potential functions form inverse maps to one another and translate from one coordinate system to another [12]:

\[
\nabla_{\theta} \psi(\theta) = q, \quad -\nabla_{q} H(q) = \theta.
\]

III. BP AND JOINT ML DETECTION

Let \( r \) be an observed vector related to the hidden variables in \( x \) through a joint likelihood function \( p_{r,x}(r,x) \) that admits a factorization of the form

\[
p_{r,x}(r,x) = \prod_{a=1}^{A} f_a(x_a)
\]

Here each \( x_a \) contains a subset of the variables in \( x \); the indices comprising this subset are collected in the set \( \mathcal{N}(a) \).

Similarly, each variable node \( x_i \) may belong to many factors, whose indices are collected into the “membership” set \( \mathcal{M}(i) \); see Figure 1. As is typically the case when belief propagation is used to decode an error correcting code, we will assume throughout that \( |\mathcal{M}(i)| \geq 2 \) for all \( i \in \{1, \ldots, M\} \), so that each variable node has degree at least two.

Maximizing the likelihood function via an exhaustive search among all configurations of \( x \) is usually computationally prohibitive. Optimizing an approximation to the likelihood function, on the other hand, can be accomplished efficiently.

One such approximation is the Bethe variational free energy, based on region approximations [7]. Let \( t_a(x_a) \) be a probability mass function that is to approximate \( f_a(x_a) \) as described shortly, and \( s_i(x_i) \) a candidate marginal density for the variable \( x_i \). The Bethe variational free energy is the difference between the region average energies

\[
\mathcal{U}_a(t_a) = -\sum_{x_a} t_a(x_a) \log f_a(x_a)
\]

and the region entropies

\[
\mathcal{S}_i(s_i) := -\sum_{x_i} s_i(x_i) \log s_i(x_i),
\]

weighted with counting numbers to ensure that each variable is counted only once [7]:

\[
\mathcal{F}_B = \sum_a [\mathcal{U}_a(t_a) - \mathcal{S}_i(s_i)] - \sum_i (1 - |\mathcal{M}(i)|) \mathcal{S}_i(s_i) \quad (2)
\]

The fixed points of belief propagation have been shown [7] to be related in a certain sense to the minimization of (2), subject to the constraint that the region densities generate the same marginals, i.e., for each variable \( x_i \), we have

\[
\sum_{x_a \setminus x_i} t_a(x_a) = s_i(x_i) \quad \text{for all } a \in \mathcal{M}(i), \ x_i \in X_i \quad (3)
\]
and \( i \in \{1, \ldots, M\} \), and are valid probability mass functions:
\[
t_a(x_a) \geq 0, \quad s_i(x_i) \geq 0, \quad \forall x_a, x_i
\]
\[
\sum_{x_a} t_a(x_a) = 1, \quad \sum_{x_i} s_i(x_i) = 1 \quad \forall i, a
\]  
(4)

By introducing the log and expectation coordinates of the region densities and factors, viz.
\[
\begin{align*}
\phi_a &:= \left[\log f_a(x_a(1)), \log f_a(x_a(2)), \ldots\right] , \\
t_a &:= [t_a(x_a(1)), t_a(x_a(2)), \ldots], \\
s_i &:= [s_i(x_i(1)), s_i(x_i(2)), \ldots]
\end{align*}
\]

we can rewrite (2) as
\[
\mathcal{F}_B(\{t_a\}, \{s_i\}) = \sum_a [-\langle \phi_a, t_a \rangle + \psi(\phi_a) - H(t_a)]
\]
\[
- \sum_i (1 - |M(i)|) H(s_i)
\]  
(5)

where \( \{t_a\} \) and \( \{s_i\} \) are to be understood as the collection of \( t_a \) over \( a \in \{1, \ldots, A\} \) and \( s_i \) over \( i \in \{1, \ldots, M\} \), respectively. Likewise, (3) may be rewritten as
\[
\begin{align*}
s_i - C_{a,i} t_a &= 0, \quad \forall a \in M(i), \ i \in \{1, \ldots, M\}
\end{align*}
\]  
(6)

where \( C_{a,i} \) maps the joint distribution on \( x_a \) to its marginal distribution on \( x_i \), and is defined in a manner analogous to (1). Note that the domain \( D \) of \( \mathcal{F}_B \) is given by
\[
D := \{ \{t_a\}, \{s_i\} \mid t_a \geq 0, \ s_i \geq 0, \ 1^T t_a \leq 1, \ 1^T s_i \leq 1 \}
\]

Because the objective function \( \mathcal{F}_B \) is undefined outside of this domain, we incorporate the inequality constraints into the domain set \( D \) rather than including them with Lagrange multipliers. We may thus form the Lagrangian for minimizing (2) subject to (3) as
\[
\mathcal{L}_B = \sum_a [-\langle \phi_a, t_a \rangle + \psi(\phi_a) - H(t_a)]
\]
\[
- \sum_i (|M(i)| - 1) H(s_i) + \sum_a \sum_{i \in N(a)} \lambda_{a,i} (s_i - C_{a,i} t_a)
\]  
(7)

Our first observation is that the dual to the constrained minimization of (2) with respect to the constraints (3) yields \( \{s_i\} \) which lie on the boundary of the domain \( D \) of the Bethe variational free energy, and thus are not defined by zero gradients. As such, zeroing the gradient of the interior Lagrangian by solving for \( \{s_i\}, \{t_a\} \) in terms of the Lagrange multipliers \( \lambda_{a,i} \) (as the fixed points of belief propagation do) has nothing to do with the Lagrangian dual to the constrained Bethe variational free energy minimization.

**Theorem 1 (Dual Lies at Boundary):** The Lagrangian dual function for the optimization problem of minimizing (2) subject to (3), defined as
\[
\inf_{\{s_i\}, \{t_a\} \in D} \mathcal{L}(\{\lambda_{a,i}\}, \{s_i\}, \{t_a\})
\]

must yield \( \{s_i\} \) lying on the boundary of the region of probability densities, and, as such, is not characterized by nulling the gradient of the associated interior Lagrangian (7).

**Proof:** By \( \cap \) convexity of the Shannon entropy, the Lagrangian is convex \( \cup \) in \( \{t_a\} \) and is convex \( \cap \) in \( \{s_i\} \). Thus, minimizing with respect to \( \{s_i\} \) must push these variables to the boundary of the associated domain \( D \), i.e. to the boundary of \( \{s_i \geq 0, 1^T s_i \leq 1\} \), and the minimum so obtained is not associated with a zero gradient of the interior Lagrangian (7) with respect to \( s_i \).

From the convexity properties noted in the proof, however, it is direct that setting the gradient of (2) with respect to \( \{s_i\}, \{t_a\} \) equal to zero yields a unique solution for \( \{s_i\}, \{t_a\} \) in terms of \( \{\lambda_{a,i}\} \); substituting these back into (2), we obtain the pseudo-dual (see [13] Def. 1.2 and Lem. 2.2), as distinguished from the conventional dual [14] to the minimization of (2) discussed in Theorem 1. The pseudo-dual is thus
\[
\mathcal{F}_B^*(\{\lambda_{a,i}\}) := \min_{\{t_a\} \{s_i\}} \mathcal{L}(\{\lambda_{a,i}\}, \{s_i\}, \{t_a\})
\]
\[
= \sum_a \left( \psi(\phi_a) - \psi(\phi_a + \sum_{i \in N(a)} C_{a,i}^T \lambda_{a,i}) \right)
\]
\[
+ \sum_i (|M(i)| - 1) \psi \left( \frac{1}{|M(i)| - 1} \sum_{a \in M(i)} \lambda_{a,i} \right)
\]  
(8)

This pseudo-dual may also be recognized as the Legendre transformation ([15] pp. 256) of the unconstrained Bethe variational free energy (5) evaluated on a parametrization of the orthogonal complement of the constraint space (6).

When understanding the relationship between the Bethe variational free energy and belief propagation, one interprets the Lagrange multipliers \( \{\lambda_{a,i}\} \) in (8) as the log coordinates of “right-going” (variable to factor) messages. For convenience, we additionally introduce log coordinates for the “left-going” (factor to variable) messages \( \{\gamma_{a,i}\} \), which relate to the “right-going” (variable to factor) log coordinates through the variable node update relationship
\[
\lambda_{a,i} = \sum_{b \in M(i) \setminus a} \gamma_{b,i}
\]  
(9)

This allows us to rewrite the pseudo-dual (8) as
\[
\mathcal{F}_B^* := \sum_a \left( \psi(\phi_a) - \psi(\phi_a + \sum_{i \in N(a)} C_{a,i}^T \lambda_{a,i}) \right)
\]
\[
- \sum_i \psi \left( \sum_{a \in M(i)} \gamma_{a,i} \right) + \sum_i \psi \left( \gamma_{a,i} + \lambda_{a,i} \right)
\]  
(10)

The problem of finding stationary points of (10) subject to the constraints (9) is equivalent to the problem of finding stationary points of (8), and is referred to as an augmentation of (8) [13].

This form of the pseudo-dual to the Bethe variational free energy is related not only to the fixed points of belief propagation, but also to the dynamics of the belief propagation algorithm. In particular, belief propagation is a nonlinear block Gauss Seidel method [9] on the system of equations
\[
\nabla \mathcal{F}_B^*(\{\lambda_{a,i}\}, \{\gamma_{a,i}\}) = 0.
\]

To see this, simply note that the log coordinates \( \lambda_{a,i} \) for the right-going (variable to factor) messages are updated by
Lemma 1: Although unconventional at first sight, proves quite amenable to a single deterministic value. Because $P$ selects only those $q^*$ which have $z$ and $y$ independent, and $C(0)$ then forces these two independent random variables to be equal, the set of distributions in $C(0)$ must then be the set of

$$q^*(z, y) = \delta[z - z_0] \delta[y - y_0]$$

for some $z_0 \in Y_1 \times \cdots \times Y_A$.

The second term in (13) represents the probability that all of the edge random variables in $Y$ neighboring a common variable node in the factor graph are equal. Because $C(0)$ constrains the probability mass functions to be of the form (14) this second term in (13) is in fact either $-\infty$ or 0, since the $z_0$ that all of the probability mass is placed at with either satisfies the test $\exists x \in X$ such that $z_0^a = x_a \forall a$ or not. This then means that we can focus only on those cases for which the $z_0$ that all of the probability mass is placed satisfies $\exists x \in X$ such that $z_0^a = x_a \forall a$. Among these $z_0$ the one which places all of its probability mass on $z_0^a$ drawn from the vector $x_{a}(0)$ maximizes the first term in (13).

Now, relax the constraint $q \in C(0)$ in (13) to $q \in C(c)$ with $c < 0$, corresponding to $\log(\Pr_q[Z = Y]) = c$. This amounts to reducing the probability that the replicas $Z$ and $Y$ on either side of the cut factor graph are the same from 1 to $e^c$. The relaxed maximization then occurs over distributions $q$ in $C(c)$, giving the optimization problem

$$\arg\max_{q \in C(c)} \log \left( \sum_{\{z^a\}} \frac{\prod_a \prod_{i \in N(a)} q_{a,i,z}(z_i^a) q_{a,i,y}(y_i^a)}{\prod_{i \in N(a)} q_{a,i,z}(z_i^a) q_{a,i,y}(y_i^a)} \right)$$

Switching to information geometric notation under which $\lambda_{a,i}$, $\gamma_{a,i}$, and $\phi_a$ are the log probability coordinates for $q_{a,i,z}(z_i^a)$, $q_{a,i,y}(y_i^a)$, and $f_a(z^a)$ respectively, this relaxed optimization problem then becomes

$$\arg\max_{q \in C(0)} \sum_{i} \log \left( \frac{\exp(\sum_{a \in M(i)} \gamma_{a,i})}{\prod_{a \in M(i)} (1 + \|\exp(\gamma_{a,i})\|_1)} + \exp(\phi_a)^T \exp(\sum_{i \in N(a)} C_{a,i}^T \lambda_{a,i}) + 1 \right)$$

The constraint set (11) when transformed into the information
The geometric notation becomes the set
\[ \mathcal{C}(c) := \left\{ \{\lambda_{a,i}\}, \{\gamma_{a,i}\} \right\} \]
\[ \sum_a \sum_{i \in N(a)} \psi(\lambda_{a,i} + \gamma_{a,i}) - \psi(\gamma_{a,i}) - \psi(\lambda_{a,i}) = c \] (17)

Because the original optimization problem (13) yielding the joint maximum likelihood detection, is obtained as the special case of the relaxed optimization problem (15) for a constraint \( c = 0 \), it is reasonable to expect that the global optimum of the relaxed optimization problem (15) should approach the joint maximum likelihood detection as \( c \to 0 \). The following Lemma shows this.

**Lemma 2**: The global optimum of the relaxed maximization (15) is continuous in \( c \) near 0. Thus the global optimum of the relaxed maximization (15) approaches the joint maximum likelihood detection, as discussed in Lemma 1, as \( c \to 0 \).

**Proof**: A detailed proof is provided in [16], only the intuition is given here. The proof proceeds by showing that for \( c \) small enough, the constraint set \( \mathcal{C}(c) \) consists of a series of sets lying in small balls surrounding consistent point mass distributions, i.e. ones for which \( q(z, y) = \delta(z - z_0)\delta(y - z_0) \).

The size of the balls shrinks to 0 as \( c \to 0 \). Next, the objective function is bounded above and below over each of the balls. As \( c \) approaches 0 the portion of the constraint set lying in the ball surrounding the maximum likelihood detection has a lower bound that is above the upper bound for the portion of the constraint set around all of the other point mass distributions. These intuitions are made precise in [16].

This relaxed constrained maximization problem is deeply related to the fixed points of belief propagation, and also related to the pseudo-dual of the Bethe variational free energy. To see this, set up the Lagrangian for (16) as
\[ \mathcal{L}_{RJML} = \sum_a \psi(\phi_a) + \sum_{i \in N(a)} C_{a,i} \psi(\lambda_{a,i}) - \sum_a \psi(\phi_a) - \sum_a \sum_{i \in N(a)} \psi(\lambda_{a,i}) - \sum_i \sum_{a \in M(i)} \psi(\gamma_{a,i}) + \mu \left( \sum_a \sum_{i \in N(a)} \psi(\lambda_{a,i} + \gamma_{a,i}) - \psi(\gamma_{a,i}) - \psi(\lambda_{a,i}) - c \right) \] (18)

If we set \( \mu = -1 \), this expression reduces to the negative of (10), the pseudo-dual to the Bethe variational free energy, plus the constant \( c \):
\[ \mathcal{L}_{RJML}|_{\mu=-1} = -\mathcal{F}_B + c. \]

The pseudo-dual of the Bethe variational free energy may thus be interpreted as the (negative of the) Lagrangian (18) of the optimization problem (15), evaluated at a Lagrange multiplier of \( \mu = -1 \) (minus the constant \( c \)). In this way, we have proven a relation between the fixed points of belief propagation and a relaxation of the joint maximum likelihood detection, as summarized in the following theorem.

**Theorem 2** (Belief Propagation as Constrained Joint Maximum Likelihood): The fixed points of belief propagation are critical points, with Lagrange multiplier \( \mu = -1 \), of the Lagrangian (18) for the relaxation (15,16,17) of the joint maximum likelihood detector. This optimization problem yields the joint maximum likelihood detection as its global optimum when \( c = 0 \), and the belief propagation fixed points null the gradient of its Lagrangian for Lagrange multiplier \( \mu = -1 \). Furthermore, its Lagrangian for Lagrange multiplier \( \mu = -1 \) equals, up to an additive constant, the negative of the pseudo-dual (10) of the Bethe variational free energy.

As Lemma 2 showed, as \( c \to 0 \), the global optimum of (15) will approach the joint maximum likelihood detection solution discussed in Lemma 1. The practical implication of Theorem 2 is thus that when the constraint value \( c \) obtained by setting the Lagrange multiplier \( \mu = -1 \) is close to 0, the global optimum of (15) will be close to the joint maximum likelihood detection. Thus, when belief propagation converges, it finds a critical point of the optimization problem (15). When this solution is a global rather than local optimum, its belief will be close to the joint maximum likelihood detection if the constraint value \( c \) associated with this critical point is close to 0.

**IV. Simulations**

We have seen in the preceding section that the belief propagation decoder, when convergent, finds critical points of the Lagrangian for maximizing an approximation to the true joint likelihood function within a constraint set \( \mathcal{C}(c) \) which fixes the log probability \( \log(\Pr[Z = Y]) \) to a constant \( c \). Furthermore, we saw that as this constant \( c \to 0 \) the global optimum for this optimization problem converged to the joint maximum likelihood detection. Here we give some simulated examples which show that the value of this constraint \( c \) usually converges very near to 0 with successive iterations for SNRs which provide a low probability of bit error at the output of the decoder. Figure 2 depicts the results for the optimized LDPC code degree distributions from [17] for block lengths \( 10^3 \) and \( 10^4 \), with the exception that we did not avoid four-cycles and larger when picking the random factor graph. The dotted lines in these graphs correspond to the log of the average across Monte Carlo runs of the constraint probability \( \Pr_q[Z = Y] \) after \( \ell \) iterations of the LDPC decoder. In terms of the value of \( c_\ell \) after \( \ell \) iterations of the LDPC decoder, the dotted lines represent Monte Carlo estimates of \( \log(\mathbb{E}[\exp(c_\ell)]) \). One can easily discern that for low bit error rates the average constraint probability converges to one (i.e. its log converges to zero) with successive iterations. This then suggests that in situations where the belief propagation decoder is providing a low probability of bit error in this loopy factor graph, the belief propagation decoder is providing decisions which are
close to the joint maximum likelihood detection in the sense of the relaxation of the joint maximum likelihood that we have discussed.

V. CONCLUSIONS

We showed that the pseudo-dual (characterized by a min-max rather than a minimum) of the Bethe variational free energy provides for a direct connection between belief propagation decoding and joint maximum likelihood decoding. In particular, it may be interpreted as the negative of the Lagrangian of a constrained optimization problem which yields the maximum likelihood sequence detections for a constraint value of 0, while yielding the fixed points of the belief propagation decoder for constraint values chosen by selecting a Lagrange multiplier of \( \mu = -1 \). Simulations then showed that in the moderate to low bit error rate regime in which LDPC codes are used, the constraint value is typically close to 0, suggesting that belief propagation works in these cases because it is performing a relaxation of joint maximum likelihood detection.

REFERENCES