Appendix to "An Efficient Message-Passing Algorithm for Optimizing Decentralized Detection Networks"

O. Patrick Kreidl and Alan S. Willsky

PROOF OF PROPOSITION 1

The proof follows the same key steps by which (3) is derived in the centralized case, but accounting for a composite measurement (Y_i, Z_i) and a cost function that also depends on non-local decision variables $(U_{-i}, \hat{X}_{-i})$. Assumption 1 is essential for the parameter values θ_i^* to be invariant to the local measurement Y_i.

Proof: The rule γ_i^* minimizes J in (1) over all $\Gamma_i^{\mathcal{G}}$, holding all other rules fixed at γ_{-i}^*, if and only if the process $(U_i, \hat{X}_i) = \gamma_i(Y_i, Z_i)$ minimizes

$$E[c(U_{-i}, u_i, \hat{X}_{-i}, \hat{x}_i, X)|Y_i, Z_i; \gamma_{-i}^*], \tag{21}$$

over all possible realizations $(u_i, \hat{x}_i) \in \mathcal{U}_i \times \mathcal{X}_i$, with probability one. Fix a realization $(u_i, \hat{x}_i)$ and consider the distribution $p(u_{-i}, \hat{x}_{-i}, x|y_i, z_i; \gamma_{-i}^*, u_i, \hat{x}_i)$ underlying (21), or equivalently

$$p(u_{-i}, \hat{x}_{-i}|x, y_i, z_i; \gamma_{-i}^*, u_i, \hat{x}_i) p(x|y_i, z_i; \gamma_{-i}^*, u_i, \hat{x}_i).$$

By virtue of Lemma 1, the first term simplifies to

$$p(u_{-i}, \hat{x}_{-i}|x, y_i, z_i; \gamma_{-i}^*, u_i) = \frac{p(u_{-i}, z_i, \hat{x}_{-i}|x; \gamma_{-i}^*, u_i)}{p(z_i|x; \gamma_{-i}^*)},$$

and, applying Bayes' rule, the second term simplifies to

$$p(x|y_i, z_i; \gamma_{-i}^*) = \frac{p(x)p(y_i|x)p(z_i|x; \gamma_{-i}^*)}{p(y_i, z_i; \gamma_{-i}^*)}$$

for every $z_i \in \mathcal{Z}_i$ such that $p(y_i, z_i; \gamma_{-i}^*) > 0$. Taking the product of the two fractions, the positive-valued denominator neither depends on x nor on $(u_i, \hat{x}_i)$ and, as such, has no bearing on the minimization of (21). Altogether, it suffices to require that $\gamma_i(Y_i, z_i)$ minimize

$$\sum_{x \in \mathcal{X}} \theta_i^*(u_i, \hat{x}_i, x; z_i) p(Y_i|x)$$

with probability one, where for each fixed value of $(u_i, \hat{x}_i)$,

$$\theta_i^*(u_i, \hat{x}_i, x; z_i) = \sum_{u_{-i}} \sum_{\hat{x}_{-i}} c(u, \hat{x}, x) p(u_{-i}, z_i, \hat{x}_{-i}, x; \gamma_{-i}^*, u_i) \tag{22}$$

This work supported by U.S. ARO under contract W911NF-05-1-0207 and by U.S. AFOSR under contract FA9550-04-1-0351.

The authors are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA {opk,willsky}@mit.edu

and, again by virtue of Lemma 1,

$$\frac{p(u_{-i}, z_i, \hat{x}_{-i}, x; \gamma^*_{-i}, u_i)}{p(z_i|x, u_{\pi(i)})} = p(x) \prod_{j \neq i} p(u_j, \hat{x}_j | x, u_{\pi(j)}; \gamma^*_j).$$

■

PROOF OF PROPOSITION 2

With Assumption 2 in effect, we may begin with the person-by-person optimality conditions expressed in Corollary 2. With Assumption 4 also in effect, we may substitute this additive cost into (1), obtaining for any fixed strategy $\gamma \in \Gamma^{\mathcal{G}}$ an additive global penalty function,

$$J(\gamma) \quad = \quad \sum_{i=1}^{n} G_i(\gamma)$$

with

$$G_i(\gamma) \quad = \quad \sum_{x_i} p(x_i) \sum_{u_i} \sum_{\hat{x}_i} c(u_i, \hat{x}_i, x_i) \sum_{z_i} p(z_i|x_i; \gamma) p(u_i, \hat{x}_i | x_i, z_i; \gamma_i)$$

for each i, where we have employed the identities

$$p(u_i, \hat{x}_i, x_i; \gamma) = p(x_i) \sum_{z_i} p(z_i, u_i, \hat{x}_i | x_i; \gamma) = p(x_i) \sum_{z_i} p(z_i|x_i; \gamma) p(u_i, \hat{x}_i | x_i, z_i; \gamma_i).$$

Lemma 2: Let Assumption 2 and Assumption 4 hold, and let $\delta(i)$ denote the *descendants* of node i (i.e., the children $\chi(i)$, each such child's children, and so on). Then Corollary 2 applies with (15) specialized to

$$\phi^*_i(u_i, \hat{x}_i, x_i; z_i) \propto p(x_i) P^*_i(z_i|x_i) \left[c(u_i, \hat{x}_i, x_i) + C^*_i(u_i, x_i; z_i) \right]$$

with likelihood function

$$P^*_i(z_i|x_i) = p(z_i|x_i; \gamma^*_{-i})$$

and cost-to-go function

$$C^*_i(u_i, x_i; z_i) = \sum_{m \in \delta(i)} \sum_{x_m} \sum_{u_m} \sum_{\hat{x}_m} p(x_m, u_m, \hat{x}_m | z_i, u_i, x_i; \gamma^*_{-i}) c(u_m, \hat{x}_m, x_m).$$

Proof: Substitute the cost in Assumption 4 into (22) and rearrange summations to obtain

$$\theta^*_i(u_i, \hat{x}_i, x; z_i) = p(x, z_i; \gamma^*_{-i}) \Bigg[c(u_i, \hat{x}_i, x_i) + \sum_{m \neq i} \sum_{u_m} \sum_{\hat{x}_m} p(u_m, \hat{x}_m | x, z_i, u_i; \gamma^*_{-i}) c(u_m, \hat{x}_m, x_m) \Bigg].$$

Conditioned on $Z_i = z_i$, the penalty term for m other than the local node i or any one of its descendants $\delta(i)$ is invariant to the candidate decision $(u_i, \hat{x}_i)$, so each such term has no bearing on the minimization in (11). That is, in Proposition 1 it now suffices to satisfy

$$\theta^*_i(u_i, \hat{x}_i, x; z_i) \propto p(x, z_i; \gamma^*_{-i}) \Bigg[c(u_i, \hat{x}_i, x_i) + \sum_{m \in \delta(i)} \sum_{u_m} \sum_{\hat{x}_m} p(u_m, \hat{x}_m | x, z_i, u_i; \gamma^*_{-i}) c(u_m, \hat{x}_m, x_m) \Bigg]$$

and, in turn, in Corollary 2 it now suffices to satisfy

$$\begin{aligned}
\phi_i^*(u_i, \hat{x}_i, x_i; z_i) &\propto \sum_{x_{-i}} p(x, z_i; \gamma_{-i}^*) \Bigg[c(u_i, \hat{x}_i, x_i) + \\
&\qquad \sum_{m \in \delta(i)} \sum_{u_m} \sum_{\hat{x}_m} p(u_m, \hat{x}_m | x, z_i, u_i; \gamma_{-i}^*) c(u_m, \hat{x}_m, x_m) \Bigg] \\
&= p(x_i, z_i | \gamma_{-i}^*) \Bigg[c(u_i, \hat{x}_i, x_i) + \\
&\qquad \sum_{x_{-i}} p(x_{-i} | x_i, z_i; \gamma_{-i}^*) \sum_{m \in \delta(i)} \sum_{u_m} \sum_{\hat{x}_m} p(u_m, \hat{x}_m | x, z_i, u_i; \gamma_{-i}^*) c(u_m, \hat{x}_m, x_m) \Bigg] \\
&= p(x_i) P_i^*(z_i | x_i) \left[c(u_i, \hat{x}_i, x_i) + C_i^*(u_i, x_i; z_i) \right].
\end{aligned}$$

■

Lemma 3: Let Assumption 2 and Assumption 3 hold, and let $\alpha(i)$ denote the *ancestors* of node i (i.e., the parents $\pi(i)$, each such parent's parents, and so on). Then, under any fixed strategy $\gamma \in \Gamma^{\mathcal{G}}$, the *local* likelihood function for received information Z_i at each node i (with at least one ancestor) satisfies

$$p(z_i | x_i; \gamma) \propto \sum_{u_{\pi(i)}} p(z_i | x_i, u_{\pi(i)}) \sum_{x_{\pi(i)}} p(x_{\pi(i)} | x_i) \prod_{j \in \pi(i)} p(u_j | x_j; \gamma)$$

with

$$p(u_j | x_j; \gamma) = \sum_{z_j} p(z_j | x_j; \gamma) \sum_{\hat{x}_j} p(u_j, \hat{x}_j | x_j, z_j; \gamma_j)$$

for every parent $j \in \pi(i)$.

Proof: Starting from Corollary 2, for every node i without ancestors (and hence without information Z_i), we have $p(z_i | x; \gamma) = 1$ and $p(u_i, \hat{x}_i | x, z_i; \gamma) = p(u_i, \hat{x}_i | x_i; \gamma_i)$. For every node i with ancestors, the forward partial order of network topology $\mathcal{G}$ implies the recursive definition

$$\begin{aligned}
p(z_i | x; \gamma) &= \sum_{z_{\pi(i)}} \sum_{u_{\pi(i)}} \sum_{\hat{x}_{\pi(i)}} p(z_{\pi(i)}, u_{\pi(i)}, \hat{x}_{\pi(i)}, z_i | x; \gamma) \\
&= \sum_{u_{\pi(i)}} p(z_i | x_i, u_{\pi(i)}) \sum_{z_{\pi(i)}} \sum_{\hat{x}_{\pi(i)}} p(z_{\pi(i)}, u_{\pi(i)}, \hat{x}_{\pi(i)} | x; \gamma) \\
&= \sum_{u_{\pi(i)}} p(z_i | x_i, u_{\pi(i)}) \sum_{z_{\pi(i)}} p(z_{\pi(i)} | x; \gamma) \sum_{\hat{x}_{\pi(i)}} p(u_{\pi(i)}, \hat{x}_{\pi(i)} | x, z_{\pi(i)}; \gamma) \\
&= \sum_{u_{\pi(i)}} p(z_i | x_i, u_{\pi(i)}) \sum_{z_{\pi(i)}} p(z_{\pi(i)} | x_{\alpha(i)}; \gamma_{\alpha(i) - \pi(i)}) \prod_{j \in \pi(i)} \sum_{\hat{x}_j} p(u_j, \hat{x}_j | x_j, z_j; \gamma_j) \\
&\equiv p(z_i | x_{\alpha(i)}, x_i; \gamma_{\alpha(i)}).
\end{aligned} \tag{23}$$

We see that the global likelihood function for information Z_i received by each node i from its parents $\pi(i)$ (if any) depends at most on the rules $\gamma_{\alpha(i)}$ local to all ancestors and the states $(X_{\alpha(i)}, X_i)$ local to itself and its ancestors. In turn, the global likelihood function for information U_i transmitted by each node i to its children $\chi(i)$ (if any) is

$$\begin{aligned} p(u_i|x;\gamma) &= \sum_{z_i} p(z_i|x;\gamma) \sum_{\hat{x}_i} p(u_i, \hat{x}_i | x_i, z_i; \gamma_i) \\ &\equiv p(u_i | x_{\alpha(i)}, x_i; \gamma_{\alpha(i)}, \gamma_i). \end{aligned} \tag{24}$$

Now, Assumption 3 ensures that no two nodes have a common ancestor, or equivalently that the collection of index sets $\{\alpha(j); j \in \pi(i)\}$ partition the index set $\alpha(i) - \pi(i)$. Because individual measurements are assumed to be mutually independent (conditioned on X), information derived from mutually-exclusive subsets of measurements will be similarly independent i.e.,

$$p(z_{\pi(i)}|x;\gamma) = \prod_{j \in \pi(i)} p(z_j|x;\gamma). \tag{25}$$

Combining (23)-(25) yields

$$\begin{aligned} p(z_i|x;\gamma) &= \sum_{u_{\pi(i)}} p(z_i|x_i, u_{\pi(i)}) \prod_{j\in\pi(i)} \left(\sum_{z_j} p(z_j|x_{\alpha(j)}, x_j; \gamma_{\alpha(j)}) \sum_{\hat{x}_j} p(u_j, \hat{x}_j | x_j, z_j; \gamma_j) \right) \\ &= \sum_{u_{\pi(i)}} p(z_i|x_i, u_{\pi(i)}) \prod_{j\in\pi(i)} p(u_j|x;\gamma), \end{aligned}$$

so that

$$\begin{aligned} p(z_i|x_i;\gamma) &= \sum_{x_{-i}} p(x_{-i}|x_i) p(z_i|x;\gamma) \\ &= \sum_{u_{\pi(i)}} p(z_i|x_i, u_{\pi(i)}) \sum_{x_{\alpha(i)}} p(x_{\alpha(i)}|x_i) \prod_{j\in\pi(i)} p(u_j|x;\gamma) \\ &= \sum_{u_{\pi(i)}} p(z_i|x_i, u_{\pi(i)}) \sum_{x_{\pi(i)}} p(x_{\pi(i)}|x_i) \sum_{x_{\alpha(i)-\pi(i)}} p(x_{\alpha(i)-\pi(i)}|x_{\pi(i)}, x_i) \prod_{j\in\pi(i)} p(u_j|x;\gamma). \end{aligned} \tag{26}$$

The last step is to recognize that we may write

$$p(x_{\alpha(i)-\pi(i)}|x_{\pi(i)}, x_i) = p(x_{\alpha(i)-\pi(i)-\alpha(m)}|x_{\pi(i)}, x_i) p(x_{\alpha(m)}|x_{\alpha(i)-\alpha(m)}, x_i)$$

for any particular $m \in \pi(i)$, in which case the inner sum in (26) is equivalent to

$$p(u_m|x_{\alpha(i)-\alpha(m)}, x_i;\gamma) \sum_{x_{\alpha(i)-\pi(i)-\alpha(m)}} p(x_{\alpha(i)-\pi(i)-\alpha(m)}|x_{\pi(i)}, x_i) \prod_{j\in\pi(i)-m} p(u_j|x;\gamma)$$

with

$$\begin{aligned}
p(u_m|x_{\alpha(i)-\alpha(m)}, x_i;\gamma) &= \sum_{x_{\alpha(m)}} p(x_{\alpha(m)}|x_{\alpha(i)-\alpha(m)}, x_i) p(u_m|x;\gamma) \\
&= \sum_{x_{\alpha(m)}} \left(\frac{p(x_{\alpha(i)}, x_i|x_m)}{p(x_{\alpha(i)-\alpha(m)}, x_i|x_m)} \right) p(u_m|x;\gamma) \\
&= \frac{\displaystyle\sum_{x_{\alpha(m)}} p(x_{\alpha(i)}, x_i|x_m) p(u_m|x;\gamma)}{p(x_{\alpha(i)-\alpha(m)}, x_i|x_m)} \\
&\propto \sum_{x_{\alpha(m)}} p(x_{\alpha(m)}|x_m) p(u_m|x;\gamma).
\end{aligned}$$

For any other parent $\ell \in \pi(i) - m$, where we let $\alpha(m,\ell)$ denote the union $\alpha(m) \cup \alpha(\ell)$, we may similarly write

$$p(x_{\alpha(i)-\pi(i)-\alpha(m)}|x_{\pi(i)}, x_i) = p(x_{\alpha(i)-\pi(i)-\alpha(m,\ell)}|x_{\pi(i)}, x_i) p(x_{\alpha(\ell)}|x_{\alpha(i)-\alpha(m,\ell)}, x_i)$$

and conclude that the inner sum in (26) is equivalent to

$$p(u_m|x_{\alpha(i)-\alpha(m)}, x_i;\gamma) p(u_\ell|x_{\alpha(i)-\alpha(\ell)}, x_i;\gamma) \sum_{x_{\alpha(i)-\pi(i)-\alpha(m,\ell)}} p(x_{\alpha(i)-\pi(i)-\alpha(m,\ell)}|x_{\pi(i)}, x_i) \prod_{j\in\pi(i)-\{m,\ell\}} p(u_j|x;\gamma)$$

with

$$p(u_\ell|x_{\alpha(i)-\alpha(\ell)}, x_i;\gamma) \propto \sum_{x_{\alpha(\ell)}} p(x_{\alpha(\ell)}|x_\ell) p(u_\ell|x;\gamma).$$

Continuing this procedure on a parent-by-parent basis, we conclude that the inner sum in (26) is proportional to

$$\prod_{j\in\pi(i)} \sum_{x_{\alpha(j)}} p(x_{\alpha(j)}|x_j) p(u_j|x;\gamma) = \prod_{j\in\pi(i)} p(u_j|x_j;\gamma),$$

where each jth factor is seen to be equal to $p(u_j|x_j;\gamma)$ by virtue of (24). ■

Taken together, Lemma 2 and Lemma 3 lead directly to the forward likelihood recursions in Proposition 2. The backward cost-to-go recursions also result from Lemma 2 and Lemma 3, taken alongside a couple of additional arguments. Firstly, by virtue of Assumption 3, the one path from any ancestor of node i to any descendant of node i includes node i. So, when conditioning on received information $Z_i = z_i$ and holding local decision $(u_i, \hat{x}_i)$ fixed, the information already received and transmitted by all ancestors is independent (conditioned on X) of the information to be received and transmitted by all descendents; mathematically, for each descendant $m \in \delta(i)$ in Lemma 2, we have

$$p(u_m, \hat{x}_m|x, z_i, u_i;\gamma^*_{-i}) = p(u_m, \hat{x}_m|x, u_i;\gamma^*_{\alpha(m)-i-\alpha(i)}, \gamma^*_m) \qquad \Rightarrow \qquad C^*_i(u_i, x_i; z_i) = C^*_i(u_i, x_i)$$

and, in turn, the pbp-optimal parameter values ϕ^*_i specialize to the form in Proposition 2. Secondly, Assumption 3 also guarantees no two children have a common descendant, implying that downstream costs decompose additively across child nodes i.e., for each i,

$$\sum_{j\in\delta(i)} G_j(\gamma) = \sum_{j\in\chi(i)} \left[G_j(\gamma) + \sum_{m\in\delta(j)} G_m(\gamma) \right].$$