Adaptive Belief Propagation – Supplementary Material

Georgios Papachristoudis GEOPAPA@MIT.EDU
John W. Fisher III FISHER@CSAIL.MIT.EDU
CSAIL, MIT, Cambridge, MA 02139, USA

Proofs

Proposition 1. *Messages in path $\mathcal{M}(w_{\ell-1} \to w_\ell)$ are correct.*

Proof. **Base case:** The messages in the path $w_1 \to w_2$ are correct. This is trivially true since all the incoming messages to w_1 and to the nodes in the path $\mathcal{M}(w_1 \to w_2)$ have been correctly evaluated during initialization. Therefore, after we absorb the measurement in the potential of X_{w_1}, propagating from $w_1 \to w_2$ will give us the correct messages.

Induction step: We will assume now that the messages in $\mathcal{M}(w_{j-1} \to w_j)$, $j \in \{2, \ldots, \ell-1\}$ are correct and we will show that the messages in $\mathcal{M}(w_{\ell-1} \to w_\ell)$ will be correct as well. W.l.o.g. assume the tree is rooted at $w_{\ell-1}$ as shown in Fig. 1 and i is one of $w_{\ell-1}$'s neighbors. We need to show that all the incoming messages to $w_{\ell-1}$ as well as the incoming messages to the other nodes in $\mathcal{M}(w_{\ell-1} \to w_\ell)$ are correct. Let's first show that the incoming messages to $w_{\ell-1}$ are correct. There are three cases for the subtree $\mathcal{T}_i$ rooted at i (if we ignore the branch containing the edge $(i, w_{\ell-1})$): **(a)** there are no previous measurements $\{w_1, \ldots, w_{\ell-2}\}$ from it, **(b)** the last measurement from it was taken at time $t_i < \ell-2$, or **(c)** at time $t_i = \ell-2$. In the first case (a), since there are no previous measurements, the incoming message $m_{i \to w_{\ell-1}}$ stayed intact since initialization and thus is correct. In the second case (b), since $t_i < \ell-2$, this means that at point t_i+1, we moved to a subtree of another neighbor of $w_{\ell-1}$ through $w_{\ell-1}$. Due to our assumption, that all messages from previous paths $\mathcal{M}(w_{j-1} \to w_j)$, $j < \ell$, are correct, this also implies that the messages in the path $\mathcal{M}(w_{t_i} \to w_{t_i+1})$ are correct and this includes message $m_{i \to w_{\ell-1}}$ as well. Lastly, if $t_i = \ell-2$, this means that the previous measurement, at time $\ell-2$, was taken from the subtree rooted at i (c). By assumption, all messages in $\mathcal{M}(w_{\ell-2} \to w_{\ell-1})$ are correct. So, in all cases, the incoming message from i to $w_{\ell-1}$ is correct. We follow similar logic for all neighbors of $w_{\ell-1}$. Lastly, we should demonstrate that the incoming messages to the other nodes in the path $\mathcal{M}(w_{\ell-1} \to w_\ell)$ are correct. The logic is similar as before. Let's refer to the subtrees that are attached to the path $\mathcal{M}(w_{\ell-1} \to w_\ell)$ as *tree branches*. Take a node (call it j) attached to $\mathcal{M}(w_{\ell-1} \to w_\ell)$ and consider the subtree $\mathcal{T}_j$ rooted at it. Let's denote by k the node in the path $\mathcal{M}(w_{\ell-1} \to w_\ell)$ that j links to, as show in Fig. 1. As before, we have three cases: **(d)** there are no previous measurements taken from $\mathcal{T}_j$, **(e)** the last measurement was taken at time $t_j < \ell-2$, or **(f)** at time $t_j = \ell-2$. If there are no previous measurements (d), this means that the message $m_{j \to k}$ stayed intact since initialization. If $t_j < \ell-2$ (e), then at point t_{j+1} we "exited" subtree $\mathcal{T}_j$ through node k and moved either to another branch of that path or to another subtree of $w_{\ell-1}$. In either case, due to our assumption, the messages in $\mathcal{M}(w_{t_j} \to w_{t_j+1})$ are correctly updated including message $m_{j \to k}$. Lastly, if $t_j = \ell-2$ (f), then due to our assumption, the messages $\mathcal{M}(w_{\ell-2} \to w_{\ell-1})$ are correct, including the message $m_{j \to k}$. We reason similarly for all nodes which are part of $\mathcal{M}(w_{\ell-1} \to w_\ell)$. Therefore, since all incoming messages to $w_{\ell-1}$ and nodes in $\mathcal{M}(w_{\ell-1} \to w_\ell)$ are correct, the messages in $\mathcal{M}(w_{\ell-1} \to w_\ell)$ would also be correct. □

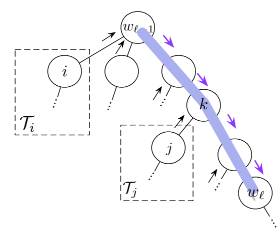


Figure 1. **Correctness of message updates.** Purple thick arrows represent the messages that will be propagated in the current iteration from $w_{\ell-1} \to w_\ell$, while solid black arrows the incoming messages to $\mathcal{M}(w_{\ell-1} \to w_\ell)$ which have been evaluated correctly from previous iterations.

Proposition 2. *The incoming messages of each node in $\mathcal{M}(w_{\ell-1} \to w_\ell)$ are correct.*

Proof. We denote by k a node in $\mathcal{M}(w_{\ell-1} \to w_\ell)$ and by j one of its neighbors, $j \in \mathcal{N}(k)$, as shown in Fig. 2. Denote further by $\mathcal{T}_j$ the tree that is rooted at j if we exclude the tree branch that contains the edge (j, k). We define as

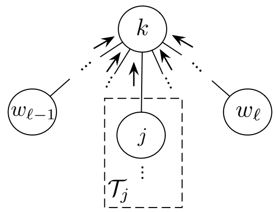
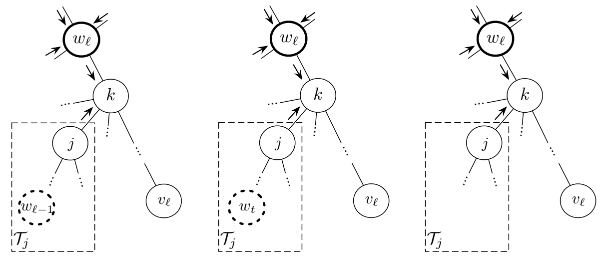


Figure 2. **Correctness of incoming messages of nodes in $\mathcal{M}(w_{\ell-1} \to w_\ell)$.** The incoming messages of every node in $\mathcal{M}(w_{\ell-1} \to w_\ell)$ are correct. Here, k is a node in $\mathcal{M}(w_{\ell-1} \to w_\ell)$ and $j \in \mathcal{N}(k)$, while $\mathcal{T}_j$ is the tree rooted at j, if exclude the tree branch that contains edge (j, k).

$t(k, j)$ the most recent time that a measurement has been obtained from tree $\mathcal{T}_j$. By default, if no measurement has been obtained from $\mathcal{T}_j$, we set $t(k, j) = \infty$. Also, if $\mathcal{T}_j$ includes node $w_{\ell-1}$, then obviously $t(k, j) = \ell - 1$. From the definition of $t(k, j)$, which indicates the time that the last measurement has been obtained from $\mathcal{T}_j$, we have that at time $t(k, j)+1$ we exited the tree $\mathcal{T}_j$ through edge (j, k). Due to Prop. 1, all messages in $\mathcal{M}(w_{t(k,j)} \to w_{t(k,j)+1})$, including message $m_{j\to k}$ are correct. We follow the same logic for all neighbors of k. $\square$

Proposition 3. *The incoming messages of each node in $\mathcal{M}(w_\ell \to v_\ell)$ are correct.*

Proof. We will first start by showing that the incoming messages of the neighbor of w_ℓ in path $\mathcal{M}(w_\ell \to v_\ell)$. Then, we can show with a similar logic that all the incoming messages of the remaining nodes in $\mathcal{M}(w_\ell \to v_\ell)$ are correct as well. From Prop. 2, we showed that the incoming messages of all nodes in $\mathcal{M}(w_{\ell-1} \to w_\ell)$ are correct. This includes node w_ℓ. Let's denote by k the neighbor of w_ℓ in $\mathcal{M}(w_\ell \to v_\ell)$. Since, the incoming messages of w_ℓ are correct and after the update of w_ℓ's potential, it follows from relation $m_{w_\ell\to k}(x_k) = \sum_{x_{w_\ell}} \varphi_{w_\ell}(x_{w_\ell})\psi_{w_\ell,k}(x_{w_\ell}, x_k) \prod_{s\in\mathcal{N}(w_\ell)\setminus k} m_{s\to w_\ell}(x_{w_\ell})$, that message $m_{w_\ell\to k}$ is correct as well. Now, let's denote by j a neighbor of k (other than w_ℓ), and by $\mathcal{T}_j$ the tree rooted at j that does not include the tree branch that contains edge (j, k), as shown in Fig. 3. Again, $t(k, j)$ denotes the most recent time a measurement has been obtained from tree $\mathcal{T}_j$. If $w_{\ell-1}$ is contained in tree $\mathcal{T}_j$, then $t(k, j) = \ell - 1$ (Fig. 3(a)), if no measurement has been obtained from $\mathcal{T}_j$, then $t(k, j) = \infty$ (Fig. 3(c)), and $t(k, j) = t < \ell - 1$, otherwise (Fig. 3(b)). If $t(k, j) = \ell - 1$, then during propagation $\mathcal{M}(w_{\ell-1} \to w_\ell)$, message $m_{j\to k}$ has been correctly updated (as part of the schedule $\mathcal{M}(w_{\ell-1} \to w_\ell)$). If $t(k, j) = t < \ell - 1$, this means that at time $t(k, j) + 1$, we exited tree $\mathcal{T}_j$ through edge (j, k). Hence, message $m_{j\to k}$ has been correctly updated during schedule $\mathcal{M}(w_{t(k,j)} \to w_{t(k,j)+1})$. Lastly,



(a) $t(k, j) = \ell - 1$. (b) $t(k, j) = t$. (c) $t(k, j) = \infty$.

Figure 3. **Correctness of incoming messages of nodes in $\mathcal{M}(w_\ell \to v_\ell)$.** The incoming messages of every node in $\mathcal{M}(w_\ell \to v_\ell)$ are correct. Tree $\mathcal{T}_j$ represents the tree rooted at node j, if we exclude the brach that contains edge (j, k). **(a)** Node $w_{\ell-1}$ is included in $\mathcal{T}_j$. **(b)** The most recent measurement from $\mathcal{T}_j$ has been taken at time $t(k, j) < \ell - 1$. **(c)** No measurements have been received from tree $\mathcal{T}_j$.

if $t(k, j) = \infty$, this means that no measurement has been obtained from tree $\mathcal{T}_j$, and hence message $m_{j\to k}$ stayed intact since initialization. This obviously holds for every neighbor j of k. We have established that all incoming messages to k, with k being the direct neighbor of w_ℓ in $\mathcal{M}(w_\ell \to v_\ell)$ are correct. Therefore, the message from k to its other neighbor in $\mathcal{M}(w_\ell \to v_\ell)$ would also be correct, since all the incoming messages to k are correct. We argue that all the incoming messages to k's neighbor in $\mathcal{M}(w_\ell \to v_\ell)$ are correct in exactly the same fashion we argued for k. By following this logic, we show that the incoming messages of all nodes in $\mathcal{M}(w_\ell \to v_\ell)$ are correct. $\square$

Corollary 1. *The marginals of all nodes in $\mathcal{M}(w_\ell \to v_\ell)$ are correct.*

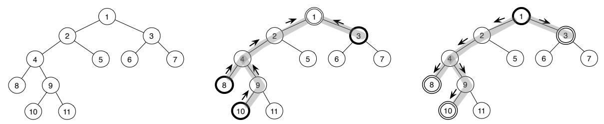
Proof. Since the marginal at a node i is given by

$$p_{\mathsf{X}_i}(x_i) \propto \varphi_i(x_i) \prod_{k\in\mathcal{N}(i)} m_{k\to i}(x_i),$$

and by Prop. 3 all incoming messages to a node in $\mathcal{M}(w_\ell \to v_\ell)$ are correct, then the marginal at node $i \in \mathcal{M}(w_\ell \to v_\ell)$ will also be correct. $\square$

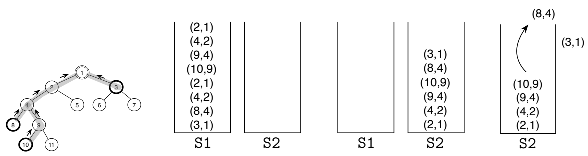
Extension to Multiple Measurements/Marginals

So far we have assumed that w_ℓ, v_ℓ are scalars. That is, we have assumed we obtain one measurement and are interested in just one marginal at a time. We can easily relax this assumption, by extending to multiple measurements or marginals at a time (cf. Fig. 4). Let's start with the case of multiple measurements and one marginal (Fig. 5(a)). That is, w_ℓ is a vector and v_ℓ a scalar. A naïve approach would be to propagate messages in $\mathcal{M}(u \to v_\ell)$, for each $u \in w_\ell$,



(a) Original graph. (b) Multiple w_ℓ, one v_ℓ. (c) One w_ℓ, multiple v_ℓ.

Figure 4. Extension to multiple measurements/marginals. **(a) Original graph. (b) Multiple w_ℓ, one v_ℓ.** Measurements at node 3, 8, 10 are obtained, while node 1's marginal is sought. **(c) One w_ℓ, multiple v_ℓ.** Measurement at node 1 is obtained, while marginals at nodes 3, 8, 10 are of interest.

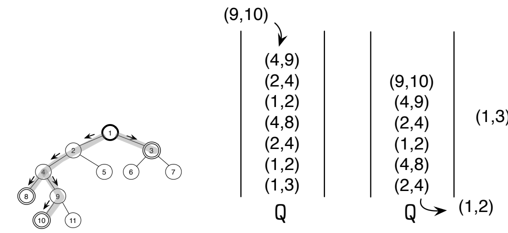


(a) Multiple w_ℓ, one v_ℓ. (b) (c) (d)

Figure 5. **(a) Multiple w_ℓ, one v_ℓ.** We need only propagate messages on the gray band from measurement nodes (in bold face) to the node of interest v_ℓ. Here, $w_\ell = \{3, 8, 10\}$, $v_\ell = 1$. **(b)** For each $u \in \mathcal{M}(u \to v_\ell)$, we push the messages that need to be evaluated on stack `S1`. This also contains duplicate messages due to the overlap of paths (e.g., messages $(2,1), (4,2)$). In this example, messages were pushed into the stack in the following order; first messages in $\mathcal{M}(3 \to 1)$, then in $\mathcal{M}(8 \to 1)$ and lastly messages in $\mathcal{M}(10 \to 1)$. **(c)** Pop each element from stack `S1` and push it to stack `S2`, while keeping a hash table to avoid duplicates. For instance, at the beginning message $(2,1)$ is pushed to stack `S2`, then $(4,2), (9,4), (10,9)$. When element $(2,1)$ is encountered again, it will be skipped since it already exists in the hash table. **(d)** After we pop all elements from stack `S1` and push them to stack `S2` (avoiding duplicates), we form the messaging schedule by popping messages from the top of stack `S2`.

but this would result in the re-evaluation of many messages that are found in overlapping paths $\mathcal{M}(u \to v_\ell)$, for all $u \in w_\ell$. Ideally, we would like to send messages on the gray band (Fig. 5(a)) just once in the right order. In that case, for each $u \in w_\ell$, we retrieve the messages in the path $\mathcal{M}(u \to v_\ell)$ that need to be evaluated and push them into a stack (`S1`) (see Fig. 5(b)). In order to place the messages in the right order of evaluation, we pop the messages and push them into a second stack, `S2`. To avoid duplicates, we keep a hash table with messages as a key. We evaluate messages by popping elements from stack `S2` one-by-one. See Fig. 5(c).

In the case of one measurement and multiple marginals (Fig. 6(a)), w_ℓ is a scalar while v_ℓ a vector. For each $u \in v_\ell$, we retrieve the messages in the path $\mathcal{M}(w_\ell \to u)$ that need to be evaluated and push them into a queue (`Q`) (see Fig. 6(b)). We poll messages from `Q` (retrieve and remove the head of the queue), while we avoid duplicates. To avoid



(a) One w_ℓ, multiple v_ℓ. (b) (c)

Figure 6. **(a) One w_ℓ, multiple v_ℓ.** We need only propagate messages on the gray band from measurement node (in bold face) to the nodes of interest v_ℓ. Here, $w_\ell = 1$, $v_\ell = \{3, 8, 10\}$. **(b)** For each $u \in \mathcal{M}(w_\ell \to u)$, we push the messages that need to be evaluated on queue `Q`. This also contains duplicate messages due to the overlap of paths (e.g., messages $(1,2), (2,4)$). In this example, messages were pushed into the queue in the following order; first messages in $\mathcal{M}(1 \to 3)$, then in $\mathcal{M}(1 \to 8)$ and lastly messages in $\mathcal{M}(1 \to 10)$. **(c)** We generate the messaging schedule by polling each element from the (head of) queue, while keeping a hash table to avoid duplicates. For instance, at the beginning message $(1,3)$ is polled from the queue and then message $(1,2)$. The second time that message $(1,2)$ will be encountered, it will be skipped since it already exists in the hash table.

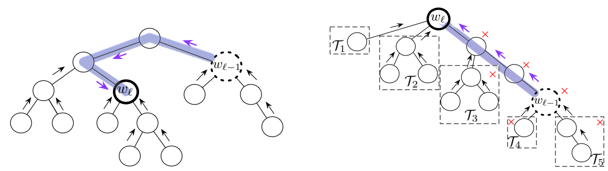
duplicates, we keep a hash table with messages as a key (Fig. 6(c)). If a message already exists in the hash table, it will not be considered in the messaging schedule.

Lastly, we treat the multiple measurements/multiple marginals case by applying the procedure of the multiple measurements/one marginal case to each different marginal.

Extension to Max-Product

In case of max-product, we just need to replace sum with max and introduce a new type of messages, called *delta messages*, that will be used for the recovery of the MAP sequence. A delta message $\delta_{i \to j}(x_j)$ indicates the value of the source node that corresponds to the MAP sequence of the subtree rooted at the source node (excl. the branch containing the target node) for a specific value of the target node. That is, if we denote by $\mathcal{T}_i$ the subtree rooted at i excluding the branch that contains j, then $[x^*_{\mathcal{T}_i}]_i = \delta_{i \to j}(x_j)$. That is, it provides the maximizing value at node i of the MAP subsequence $x^*_{\mathcal{T}_i}$ if node j had value x_j. In order to recover the MAP sequence, we need to propagate delta messages from $w_{\ell-1}$ to w_ℓ and then backtrack from w_ℓ down to the leaves (considering w_ℓ as the root).

In general, obtaining the MAP sequence is a linear operation in the number of nodes. However, local changes in node potentials (via the introduction of measurements) might induce only small changes in the MAP sequence.



(a) Message updates (max-product version).

(b) Savings in MAP computations.

Figure 7. **(a) Message updates (max-product version).** Purple thick arrows represent the messages that will be propagated in the current iteration, while solid black messages that have been evaluated correctly from previous iterations. **(b) Savings in MAP sequence computations.** During the ℓ-th step, the node potential at w_ℓ (bold-faced node) as well as delta messages $\delta_{i\to j}(x_j)$ in $\mathcal{M}(w_{\ell-1} \to w_\ell)$ (purple arrows) change. Let's assume the maximizing value at w_ℓ remained the same compared to the previous iteration, while the maximizing values of the remaining nodes in path $\mathcal{M}(w_{\ell-1} \to w_\ell)$ changed. We visualize this change with a red $\times$. Since the maximizing value at w_ℓ stayed intact, the maximizing values of all its subtrees (not including the one containing path $\mathcal{M}(w_{\ell-1} \to w_\ell)$) will remain the same (here, trees $\mathcal{T}_1, \mathcal{T}_2$). Therefore, there is no need to backtrack down to a node whose maximizing value did not change since the last iteration. On the other hand, since the maximizing values of the remaining nodes in $\mathcal{M}(w_{\ell-1} \to w_\ell)$ changed, their subtrees' maximizing values $(\mathcal{T}_3, \mathcal{T}_4, \mathcal{T}_5)$ would also potentially change and hence backtracking on these trees is necessary. Usually, a change in a node's maximizing value results only in local changes in the MAP sequence. Therefore, this scheme might practically lead to a lot of computational savings.

We should also note that the only delta messages pointing towards the root w_ℓ that have changed, are the ones in path $\mathcal{M}(w_{\ell-1} \to w_\ell)$, which were correctly updated (at iteration ℓ). See Fig. 7(a) for details. This observation can help us recover the MAP sequence in a more efficient way. In more detail, we can create an indicator sparse matrix, where rows would represent the source and columns the target of a delta message. We can assign the value 1 to any delta message that became "dirty" (changed) in the most recent iteration. That is, every message in path $\mathcal{M}(w_{\ell-1} \to w_\ell)$ (purple arrows in Fig. 7(a)). Therefore, when we backtrack from w_ℓ down to the leaves, we must consider the effect that these changed messages can have in the MAP sequence. Nevertheless, if the value of a node remains the same (with the previous iteration), then the subsequences of the subtrees rooted at the neighbors of this node will remain the same. Therefore, there is no need to backtrack further down to a subtree once a node's maximizing value remained the same and the subtree is linked to that node via a "clean" message. A visual explanation is provided in Fig. 7(b).

Extension to Gaussian Loopy MRFs

In the case of Gaussian loopy graphs, we should observe that in order to obtain a marginal at a node, we need to send two types of messages; first–round messages $J^{\mathcal{T}}_{i\to j}, h^{\mathcal{T}}_{i\to j}$ corresponding to the acyclic part of the graph $\mathcal{T}$ after the removal of FVS $\mathcal{F}$, "feedback" messages $h^{p}_{i\to j}$ for every feedback vertex p and second–round messages $\tilde{h}^{\mathcal{T}}_{i\to j}$, which are revised potential messages after the update of the potential vector h at the anchors (neighbors of feedback vertices). This change in potential vector requires the knowledge of the updated means $\mu_{\mathcal{F}}$ and covariance $\Sigma_{\mathcal{F}}$ of the FVS, which requires in turn the knowledge of all "partial" means $\hat{\mu}^{\mathcal{T}}_i$ and "feedback gains" g^p_i at the anchors.

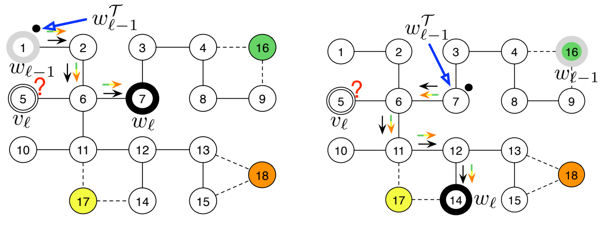
This observation leads to a natural extension of adaptive BP to Gaussian loopy graphs. First, let's denote the node from set $\mathcal{T}$, where a measurement has been obtained most recently as $w^{\mathcal{T}}_\ell$. It obviously holds

$$w^{\mathcal{T}}_\ell = \begin{cases} w_\ell & w_\ell \in \mathcal{T} \\ w^{\mathcal{T}}_{\ell-1} & \text{otherwise.} \end{cases}$$

If $w_{\ell-1}, w_\ell \in \mathcal{T}$, we send messages $J^{\mathcal{T}}_{i\to j}, h^{\mathcal{T}}_{i\to j}, h^{p}_{i\to j}$ from $w_{\ell-1} \to w_\ell$ exactly as we did in the acyclic case (see Fig. 8(a)). However, when $w_{\ell-1} \in \mathcal{F}$, we need to send messages from node $w^{\mathcal{T}}_{\ell-1}$, which is the node where a measurement has been obtained most recently, to propagate the effects of the past changes in the current node w_ℓ (see Fig. 8(b)). In summary, when $w_\ell \in \mathcal{T}$, we send messages from $w^{\mathcal{T}}_{\ell-1}$ to w_ℓ, while no action is necessary when $w_\ell \in \mathcal{F}$. By this procedure, we ensure that all incoming messages $J^{\mathcal{T}}_{i\to j}, h^{\mathcal{T}}_{i\to j}, h^{p}_{i\to j}$ to w_ℓ are correct.

Now let's assume we want to update $v_\ell \in \mathcal{F}$. This would require the knowledge of partial means $\hat{\mu}^{\mathcal{T}}_i$ and "feedback gains" $g^p_i, \forall p \in \mathcal{F}$ at the anchors, $i \in \mathcal{A}$. As a reminder, anchors are the neighbors of FVS nodes which belong in $\mathcal{T}$, $\mathcal{A} = \{i \mid i \in \mathcal{T}, i \in \mathcal{N}(p), \forall p \in \mathcal{F}\}$. The correct update of $\hat{\mu}^{\mathcal{T}}_i, g^p_i, \forall i \in \mathcal{A}$ leads to the correct evaluation of $\hat{h}_{\mathcal{F}}, \hat{J}_{\mathcal{F}}$ which in turn provides the correct mean and variance for v_ℓ since by our assumption belongs to the FVS. Partial means and "feedback gains" at the anchors would be correct if the change in the potential of the most recent measurement node $w^{\mathcal{T}}_\ell$ is propagated at the anchors after the update of node's w_ℓ potential. We achieve this by sending messages $J^{\mathcal{T}}_{i\to j}, h^{\mathcal{T}}_{i\to j}, h^{p}_{i\to j}$ from $w^{\mathcal{T}}_\ell$ to all anchors $\mathcal{A}, \forall i$. This guarantees that all incoming messages at the anchors are correct.

If $v_\ell \in \mathcal{T}$, we need to propagate a second–round of messages to account for the feedback provided by the updated parameters $\mu_{\mathcal{F}}, \Sigma_{\mathcal{F}}$ of the FVS nodes. In other words, we revise the potential vectors as $\tilde{h}_i = h_i + \sum_{j\in\mathcal{N}(i)\cap\mathcal{F}} J_{ij}[\mu_{\mathcal{F}}]_j$. From an inspection of the above relationship, we can easily see that the only potential vectors



(a) $w_{\ell-1}, w_\ell \in \mathcal{T}$. (b) $w_{\ell-1} \in \mathcal{F}, w_\ell \in \mathcal{T}$.

Figure 8. **First phase of AdaBP: Messages from $w_{\ell-1}$ to w_ℓ.** Nodes filled with color represent the nodes in FVS $\mathcal{F}$. Removal of FV nodes breaks the graph in an acyclic graph $\mathcal{T}$. Bold node in black represents the current measurement node, while in gray the previous measurement node $w_{\ell-1}$. Small black dot represents the most recent measurement node from $\mathcal{T}$ (before step ℓ), while double stroke node next to a question mark, the node of interest v_ℓ. **(a)** $w_{\ell-1}, w_\ell \in \mathcal{T}$. In this case, because $w_{\ell-1} \in \mathcal{T}$, we have $w_{\ell-1}^{\mathcal{T}} = w_{\ell-1}$. We send messages $J_{i\to j}^{\mathcal{T}}, h_{i\to j}^{\mathcal{T}}$ (black color) and messages $h_{i\to j}^{p}, \forall p \in \mathcal{F}$ (multi color) from $w_{\ell-1}$ to w_ℓ. **(b)** $w_{\ell-1} \in \mathcal{F}, w_\ell \in \mathcal{T}$. In this case, $w_{\ell-1} \in \mathcal{F}$. Therefore, $w_{\ell-1}^{\mathcal{T}} \neq w_{\ell-1}$. We send messages $J_{i\to j}^{\mathcal{T}}, h_{i\to j}^{\mathcal{T}}$ (black color) and messages $h_{i\to j}^{p}, \forall p \in \mathcal{F}$ (multi color) from $w_{\ell-1}^{\mathcal{T}}$ to w_ℓ. This phase corresponds to l. 8–10 of Alg. 1.

which would change are the ones at the anchors, since the sum involves the intersection of the FVS nodes $\mathcal{F}$ and the neighbors of a node in $i \in \mathcal{T}$. This means we need to propagate messages $\tilde{h}_{i\to j}^{\mathcal{T}}$ from the anchors $\mathcal{A}$ to node v_ℓ. We obtain the right mean at v_ℓ from $(\tilde{h}_{\mathcal{T}}, J_{\mathcal{T}\mathcal{T}}, \tilde{h}_{i\to j}^{\mathcal{T}}, J_{i\to j}^{\mathcal{T}})$. Lastly, we correct the variance from $\sigma_{v_\ell}^2 = (\hat{J}_{v_\ell,v_\ell}^{\mathcal{T}})^{-1} + \sum_{p,q\in\mathcal{F}} g_i^p [\Sigma_{\mathcal{F}}]_{pq} g_i^q$, where $\hat{J}_{v_\ell,v_\ell}^{\mathcal{T}}$ is obtained from the previous run of BP. As we observe, the "feedback gains" at node v_ℓ are essential for the correct evaluation of variance at v_ℓ. As a last step, we need to propagate messages $h_{i\to j}^{p}, \forall p \in \mathcal{F}$ from $w_\ell^{\mathcal{T}}$ to v_ℓ to transfer the effect of $w^{\mathcal{T}}$ to node v_ℓ. This concludes the algorithm. A full description of the algorithm is provided in Alg. 1, while a depiction of the algorithmic flow is shown in Figs. 8, 9, 10.

Corollary 2. *If $w_\ell \in \mathcal{T}$, messages $h_{i\to j}^{\mathcal{T}}, J_{i\to j}^{\mathcal{T}}, h_{i\to j}^{p}, \forall p \in \mathcal{F}$ in path $\mathcal{M}(w_{\ell-1}^{\mathcal{T}} \to w_\ell)$ are correct.*

Proof. The proof follows the same logic with that of Prop. 1. The only difference here is that $w_{\ell-1}$ is substituted by $w_{\ell-1}^{\mathcal{T}}$, which is defined as:

$$w_{\ell-1}^{\mathcal{T}} = \begin{cases} w_{\ell-1} & \text{, if } w_{\ell-1} \in \mathcal{T} \\ w_{\ell-2}^{\mathcal{T}} & \text{, otherwise.} \end{cases}$$

In other words, $w_{\ell-1}^{\mathcal{T}}$ represents the most recent measurement that has been obtained from $\mathcal{T}$. The reason for propagating from $w_{\ell-1}^{\mathcal{T}}$ to w_ℓ is that we need to propagate the effect of the most recent measurement in $\mathcal{T}$ to w_ℓ. Obviously, when $w_\ell \in \mathcal{F}$, schedule $\mathcal{M}(w_{\ell-1}^{\mathcal{T}} \to w_\ell) = \emptyset$. □

Algorithm 1 ADAPTIVE BP FOR GAUSSIAN LOOPY GRAPHS

1: **Preprocessing**
2: Find FVS $\mathcal{F}$ using one of known algorithms (e.g., (Bafna et al., 1999)).
3: Build the RMQ structure on tree $\mathcal{T} = V \setminus \mathcal{F}$ as described in Sec. 3 of main paper.
4: **Initialization**
5: Before incorporating any measurements, run BP on tree $\mathcal{T}$ using parameters $(h_{\mathcal{T}}, J_{\mathcal{T}\mathcal{T}}), (J_{\mathcal{T}p}, J_{\mathcal{T}\mathcal{T}}), \forall p \in \mathcal{F}$. This will generate first–round $J_{i\to j}^{\mathcal{T}}, h_{i\to j}^{\mathcal{T}}, h_{i\to j}^{p}, \forall p \in \mathcal{F}$, and second–round messages $\tilde{h}_{i\to j}^{\mathcal{T}} \equiv h_{i\to j}^{\mathcal{T}}$. Also, initialize $w_0^{\mathcal{T}} = 0$.
6: **Iteration**
7: **for** $\ell = 1, \ldots, M$ **do**
8: **if** $\ell > 1 \wedge w_{\ell-1}^{\mathcal{T}} \neq 0 \wedge w_\ell \in \mathcal{T}$ **then**
9: Send $J_{i\to j}^{\mathcal{T}}, h_{i\to j}^{\mathcal{T}}, h_{i\to j}^{p}, \forall p\in\mathcal{F}$ in $\mathcal{M}(w_{\ell-1}^{\mathcal{T}} \to w_\ell)$.
10: **end if**
11: Update the node potential at X_{w_ℓ}: this changes $h_{w_\ell}, J_{w_\ell,w_\ell}$.
12: Send messages $J_{i\to j}^{\mathcal{T}}, h_{i\to j}^{\mathcal{T}}, h_{i\to j}^{p}, \forall p \in \mathcal{F}$ in $\mathcal{M}(w_\ell^{\mathcal{T}} \to \mathcal{A})$.
13: Evaluate partial means $\hat{\mu}_i^{\mathcal{T}}$ from $(h_{\mathcal{T}}, J_{\mathcal{T}\mathcal{T}}, h_{i\to j}^{\mathcal{T}}, J_{i\to j}^{\mathcal{T}})$ and "feedback gains" g_i^p from $(J_{\mathcal{T}p}, J_{\mathcal{T}\mathcal{T}}, h_{i\to j}^{p}, J_{i\to j}^{\mathcal{T}})$, for all $i \in \mathcal{A}, p \in \mathcal{F}$.
14: Obtain the K–sized FVS graph with updated parameters $\hat{h}_{\mathcal{F}}, \hat{J}_{\mathcal{F}}$ as

$$[\hat{J}_{\mathcal{F}}]_{pq} = J_{pq} - \sum_{i\in\mathcal{N}(p)\cap\mathcal{T}} J_{pi} g_i^q, \ \forall p, q \in \mathcal{F} \quad (1)$$

$$[\hat{h}_{\mathcal{F}}]_p = h_p - \sum_{i\in\mathcal{N}(p)\cap\mathcal{T}} J_{pi} \hat{\mu}_i^{\mathcal{T}}, \ \forall p \in \mathcal{F} \quad (2)$$

and solve for $\Sigma_{\mathcal{F}} = \hat{J}_{\mathcal{F}}^{-1}$ and $\mu_{\mathcal{F}} = \Sigma_{\mathcal{F}} \hat{h}_{\mathcal{F}}$.
15: **if** $v_\ell \in \mathcal{F}$ **then**
16: $\mu_{v_\ell} = [\mu_{\mathcal{F}}]_{v_\ell}, \sigma_{v_\ell}^2 = [\Sigma_{\mathcal{F}}]_{v_\ell,v_\ell}$.
17: **else**
18: Revise potential vectors as $\tilde{h}_i = h_i + \sum_{j\in\mathcal{N}(i)\cap\mathcal{F}} J_{ij} [\mu_{\mathcal{F}}]_j$.
19: Send messages $\tilde{h}_{i\to j}^{\mathcal{T}}$ in $\mathcal{M}(\mathcal{A} \to v_\ell)$.
20: Send messages $J_{i\to j}^{\mathcal{T}}, h_{i\to j}^{p}, \forall p \in \mathcal{F}$ in $\mathcal{M}(w_\ell^{\mathcal{T}} \to v_\ell)$.
21: Evaluate $\mu_{v_\ell} = (\hat{J}_{v_\ell,v_\ell}^{\mathcal{T}})^{-1} \hat{h}_{v_\ell}^{\mathcal{T}}$, where

$$\hat{h}_{v_\ell}^{\mathcal{T}} = \tilde{h}_{v_\ell} + \sum_{k\in\mathcal{N}(v_\ell)} \tilde{h}_{k\to v_\ell}^{\mathcal{T}} \quad (3)$$

$$\hat{J}_{v_\ell,v_\ell}^{\mathcal{T}} = J_{v_\ell,v_\ell} + \sum_{k\in\mathcal{N}(v_\ell)} J_{k\to v_\ell}^{\mathcal{T}}. \quad (4)$$

$$\text{and } \sigma_{v_\ell}^2 = (\hat{J}_{v_\ell,v_\ell}^{\mathcal{T}})^{-1} + \sum_{p,q\in\mathcal{F}} g_{v_\ell}^p [\Sigma_{\mathcal{F}}]_{pq} g_{v_\ell}^q. \quad (5)$$

22: Reset messages $\tilde{h}_{i\to j}^{\mathcal{T}}$ in $\mathcal{M}(\mathcal{A} \to v_\ell)$.
23: **end if**
24: **end for**

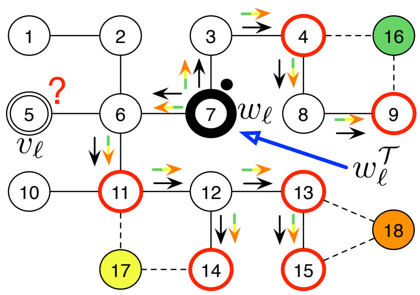


Figure 9. **First phase of AdaBP: Messages from $w_\ell^{\mathcal{T}}$ to $\mathcal{A}$.** Nodes filled with color represent the nodes in FVS $\mathcal{F}$. Bold node in black represents the current measurement node, while nodes in red represent the anchors $\mathcal{A}$. Small black dot represents the most recent measurement node from $\mathcal{T}$ (incl. step ℓ), while double stroke node next to a question mark, the node of interest v_ℓ. In this example, because $w_\ell \in \mathcal{T}$, we have that $w_\ell^{\mathcal{T}} = w_\ell$. After we update the potential at X_{w_ℓ}, we propagate messages $J_{i\to j}^{\mathcal{T}}, h_{i\to j}^{\mathcal{T}}$ (black color) and messages $h_{i\to j}^{p}, \forall p \in \mathcal{F}$ (multi color) from $w_\ell^{\mathcal{T}}$ to the set of nodes $\mathcal{A}$ (l. 12 of Alg. 1). After the end of this step, we are guaranteed that the partial means $\hat{\mu}_i^{\mathcal{T}}$ and "feedback gains" g_i^p, for all $i \in \mathcal{A}, p \in \mathcal{F}$ are correct.

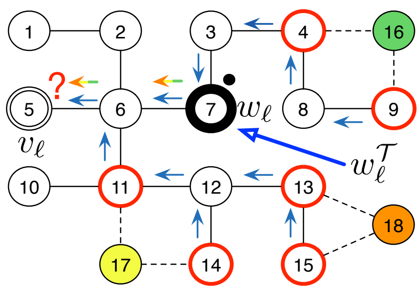
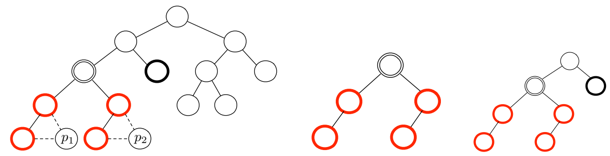


Figure 10. **Second phase of AdaBP: Messages from $\mathcal{A}$ to v_ℓ.** This phase would only take place if $v_\ell \in \mathcal{T}$. Nodes filled with color represent the nodes in FVS $\mathcal{F}$. Bold node in black represents the current measurement node, while nodes in red represent the anchors $\mathcal{A}$. Small black dot represents the most recent measurement node from $\mathcal{T}$ (incl. step ℓ), while double stroke node next to a question mark, the node of interest v_ℓ. In this example, because $w_\ell \in \mathcal{T}$, we have that $w_\ell^{\mathcal{T}} = w_\ell$. After we compute the mean $\mu_{\mathcal{F}}$ and covariance $\Sigma_{\mathcal{F}}$ at the FVS, the marginal means and variances at the FVS $\mathcal{F}$ are correct. If $v_\ell \in \mathcal{F}$, we simply retrieve its mean and variance by $\mu_{v_\ell} = [\mu_{\mathcal{F}}]_{v_\ell}, \sigma_{v_\ell}^2 = [\Sigma_{\mathcal{F}}]_{v_\ell, v_\ell}$. Otherwise, after revising the potential vectors $\tilde{h}_i$ at the anchors $\mathcal{A}$, we send messages $\tilde{h}_{i\to j}^{\mathcal{T}}$ from anchors $\mathcal{A}$ to v_ℓ (blue color) and messages $J_{i\to j}^{\mathcal{T}}, h_{i\to j}^{p}, \forall p \in \mathcal{F}$ from $w_\ell^{\mathcal{T}}$ to v_ℓ (multi color). This phase corresponds to l. 15–20 of Alg. 1.



(a) Original graph. (b) $\mathcal{T}_\ell^v$ (c) $\mathcal{T}_\ell^w$

Figure 11. **(a) Original loopy graph.** The graph $G = (V, \mathcal{E})$ is divided in FVS nodes $\mathcal{F}$ (nodes p_1, p_2) and the acyclic part $\mathcal{T} = V \setminus \mathcal{F}$. The black bold-faced node indicates the node from $\mathcal{T}$ where a measurement has been taken most recently, $w_\ell^{\mathcal{T}}$, the double-stroke node represents the node of interest, v_ℓ, and the red bold-faced nodes represent the anchor nodes $\mathcal{A}$, that is, nodes in $\mathcal{T}$ that are neighbors to FVS nodes. **(b) $\mathcal{T}_\ell^v$ tree.** Tree $\mathcal{T}_\ell^v$ is the subtree of $\mathcal{T}$ that has node v_ℓ as a root and passes through all anchor nodes $\mathcal{A}$. **(c) $\mathcal{T}_\ell^w$ tree.** Tree $\mathcal{T}_\ell^v$ is the subtree of $\mathcal{T}$ that that has node $w_\ell^{\mathcal{T}}$ as a root and passes through all anchor nodes $\mathcal{A}$.

Corollary 3. *If $w_\ell \in \mathcal{T}$, the incoming messages $h_{i\to j}^{\mathcal{T}}, J_{i\to j}^{\mathcal{T}}, h_{i\to j}^{p}, \forall p \in \mathcal{F}$ of each node in $\mathcal{M}(w_{\ell-1}^{\mathcal{T}} \to w_\ell)$ are correct.*

Proof. The proof follows the same logic with that of Prop. 2. For every node k in the path $\mathcal{M}(w_{\ell-1}^{\mathcal{T}} \to w_\ell)$, we consider one of its neighbors in tree $\mathcal{T}$. Let's denote it by j. We are interested in showing that the messages $h_{j\to k}^{\mathcal{T}}, J_{j\to k}^{\mathcal{T}}, h_{j\to k}^{p}, \forall p \in \mathcal{F}$ are correct. Again, we denote by $\mathcal{T}_j$ the tree rooted at j, if we exclude the branch that contains edge (j, k) and by $t(j, k)$ the most recent time that a measurement has been obtained from tree $\mathcal{T}_j$. Then, at time $t(j, k) + 1$, we exited tree $\mathcal{T}_j$ through the edge (j, k), and by Cor. 2 messages in $\mathcal{M}(w_{t(j,k)}^{\mathcal{T}} \to w_{t(j,k)+1})$ are correct, which includes messages $h_{j\to k}^{\mathcal{T}}, J_{j\to k}^{\mathcal{T}}, h_{j\to k}^{p}, \forall p \in \mathcal{F}$. This holds for every neighbor of k in $\mathcal{T}$. □

Let's denote the (minimal) subtree of $\mathcal{T}$ rooted at $w_\ell^{\mathcal{T}}$ that passes through all the anchor nodes $\mathcal{A}$ by $\mathcal{T}_\ell^w$ and by $\mathcal{T}_\ell^v$ the (minimal) subtree that is rooted at $v_\ell \in \mathcal{T}$ and passes through all the anchor nodes $\mathcal{A}$. See Fig. 11 for a visualization of trees $\mathcal{T}_\ell^w, \mathcal{T}_\ell^v$. Messages from $w_\ell^{\mathcal{T}}$ to $\mathcal{A}$ in the tree $\mathcal{T}_\ell^w$ represent the messages in $\mathcal{M}(w_\ell^{\mathcal{T}} \to \mathcal{A})$. Equivalently, messages from all $u \in \mathcal{A}$ to $v_\ell \in \mathcal{T}$ in the tree $\mathcal{T}_\ell^v$ represent the messages in $\mathcal{M}(\mathcal{A} \to v_\ell)$.

Proposition 4. *The incoming messages $h_{i\to j}^{\mathcal{T}}, J_{i\to j}^{\mathcal{T}}, h_{i\to j}^{p}, \forall p \in \mathcal{F}$ of each node in $\mathcal{T}_\ell^w$ are correct.*

Proof. We should show that the messages from the "root" $w_\ell^{\mathcal{T}}$ towards the leaves of the minimal subtree $\mathcal{T}_\ell^w$ that contains all the nodes in $\mathcal{A}$ are correct. If $w_\ell \in \mathcal{T}$, then $w_\ell^{\mathcal{T}} = w_\ell$ and we showed in Cor. 3 that the incoming messages of each node in $\mathcal{M}(w_{\ell-1}^{\mathcal{T}} \to w_\ell)$ are correct. This includes the incoming messages to node w_ℓ. If $w_\ell \in \mathcal{F}$, and τ was the last time a measurement was obtained from $\mathcal{T}$, then $w_\ell^\mathcal{T} = w_\tau$ and by Cor. 3 all incoming messages to every node in $\mathcal{M}(w_{\tau-1}^\mathcal{T} \to w_\tau)$ are correct, which includes those of node w_τ. Since, by assumption all remaining measurements (from $\tau + 1$ to ℓ have been taken from $\mathcal{F}$), the incoming messages $h_{i\to j}^\mathcal{T}, J_{i\to j}^\mathcal{T}, h_{i\to j}^p, \forall p \in \mathcal{F}$ to w_τ would reflect the correct value up to iteration ℓ. Therefore, we established that whether $w_\ell \in \mathcal{T}$ or $w_\ell \in \mathcal{F}$, the incoming messages to $w_\ell^\mathcal{T}$ are correct. Consequently, messages to its children would be correct. We show that the incoming messages of the remaining nodes in $\mathcal{T}_\ell^w$ are correct in exactly the same fashion as in Prop. 3. That is, if we denote by k a child of $w_\ell^\mathcal{T}$ and by j one of k's neighbors, we show that messages $h_{j\to k}^\mathcal{T}, J_{j\to k}^\mathcal{T}, h_{j\to k}^p, \forall p \in \mathcal{F}$ are correct by claiming that they have been part of a past message schedule $\mathcal{M}(w_{t(k,j)} \to w_{t(k,j)+1})$, where $t(k,j)$ is the most recent time a measurement has been obtained from subtree $\mathcal{T}_j$.[1] We continue this reasoning in a top-down approach, from the "root" $w_\ell^\mathcal{T}$ to the nodes in $\mathcal{A}$. □

Corollary 4. *The "partial" means $\hat{\mu}_i^\mathcal{T}$ of all nodes in $\mathcal{T}_\ell^w$ are correct.*

Proof. This follows trivially from Prop. 4, since all incoming messages $h_{i\to j}^\mathcal{T}, J_{i\to j}^\mathcal{T}$ to every node in $\mathcal{T}_\ell^w$ are correct. □

Corollary 5. *The "feedback gains" g_i^p of all nodes in $\mathcal{T}_\ell^w$ are correct.*

Proof. This follows trivially from Prop. 4, since all incoming messages $h_{i\to j}^p, J_{i\to j}^\mathcal{T}, \forall p \in \mathcal{F}$ to every node in $\mathcal{T}_\ell^w$ are correct. □

Corollary 6. *The mean $\mu_\mathcal{F}$ and covariance $\Sigma_\mathcal{F}$ are correct.*

Proof. From Eqs. (1), (2), we see that $\hat{h}, \hat{J}$ are correct, since by Cor. 4, 5 "partial" means $\hat{\mu}_i^\mathcal{T}$ and "feedback gains" $g_i^p, \forall p \in \mathcal{F}$ at the anchors are correct and the node potential at X_{w_ℓ} has been already updated (l. 11, Alg. 1). □

Corollary 7. *The revised potentials $\tilde{h}_i$ for every node $i \in \mathcal{T}$ are correct.*

Proof. From l. 18 of Alg. 1, the revised potential is defined as

$$\tilde{h}_i = h_i + \sum_{j \in \mathcal{N}(i) \cup \mathcal{F}} J_{ij} [\mu_\mathcal{F}]_j.$$

Since by Cor. 6, we showed that $\mu_\mathcal{F}$ is correct, then the revised potentials would be correct as well. From the summation, it is clear that the only potential vectors that are revised are the ones at the anchors. □

[1] As a reminder, subtree $\mathcal{T}_j$ is defined as the subtree rooted at j that excludes the branch which contains edge (j, k).

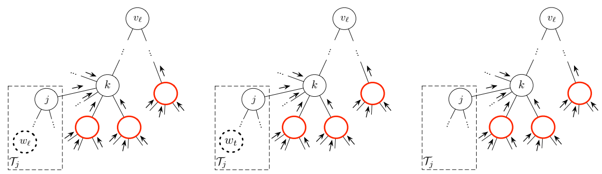
Proposition 5. *The incoming messages $\tilde{h}_{i\to j}^\mathcal{T}$ of each node in $\mathcal{T}_\ell^v$ are correct.*

Proof. Let's start with the first iteration, $\ell = 1$. Messages $\tilde{h}_{i\to j}^\mathcal{T}$ to nodes in $\mathcal{A}$ are identical to $h_{i\to j}^\mathcal{T}$, since no other node potential has been revised yet. Initially, the incoming messages $h_{i\to j}^\mathcal{T}, J_{i\to j}^\mathcal{T}$ to the anchors $\mathcal{A}$ are correct by Prop. 4. The only potentials that are revised after we learn $\mu_\mathcal{F}, \Sigma_\mathcal{F}$ are the ones at the anchors, which by Cor. 7 are correct. This implies, that revised messages $\tilde{h}_{i\to j}^\mathcal{T}$ from the anchors to their parent nodes would also be correct. Let's denote by k a parent of an anchor node and by j one of its neighbors, as shown in Fig. 12. If j is one of the anchors, since we assumed that k is parent node to an anchor node, we just argued above that the message $\tilde{h}_{j\to k}^\mathcal{T}$ is correct. Now, if j does not belong to the tree $\mathcal{T}_\ell^v$, we denote by $\mathcal{T}_j$ the tree rooted at j excluding the tree branch that contains edge (j, k), as shown in Fig. 12. There are three scenarios, the last measurement that has been received from tree $\mathcal{T}_j$ is in time $t(j,k) = \ell$ (Fig. 12(a)), $t(j,k) = t < \ell$ (Fig. 12(b)) or $t(j,k) = \infty$ (Fig. 12(c)), which means that no measurement has been obtained from that tree yet. For the first two cases, message $\tilde{h}_{j\to k}^\mathcal{T}$, which is identical to $h_{j\to k}^\mathcal{T}$, is correct as it is part of the schedule $\mathcal{M}(w_{t(k,j)\to t(k,j)+1})$. By Prop. 4, all incoming messages $h_{j\to k}^\mathcal{T}, J_{j\to k}^\mathcal{T}, h_{j\to k}^p$, of each node in $\mathcal{M}(w_{t(k,j)\to t(k,j)+1})$ are correct. For the third case, when there is no measurement from subtree $\mathcal{T}_j$, message $h_{j\to k}^\mathcal{T}$ has stayed intact since initialization. So, in all three cases message $\tilde{h}_{j\to k}^\mathcal{T}$ is correct. Hence, when node k sends a message to its own parent it will also be correct. Obviously, here because we start with the first iteration $\ell = 1$, $t(k,j)$ can only be $t(k,j) = 1$ or $t(j,k) = \infty$. Continuing in this logic, we show that all incoming messages to node v_ℓ, $\tilde{h}_{j\to v_\ell}^\mathcal{T}$, are correct as well. As a last step of the algorithm, after we evaluate the marginal at the node of interest, we reset all messages $\tilde{h}_{i\to j}^\mathcal{T}$ in T_ℓ^v to their previous values as the revised potentials $\tilde{h}_i$ at the anchors reflect "imaginary" changes produced by the feedback of FVS nodes, rather than real changes that would be an outcome of obtaining a new measurement. By doing so, we guarantee that messages $\tilde{h}_{i\to j}^\mathcal{T}$ coincide with messages $h_{i\to j}^\mathcal{T}$ at the end of the first iteration. Therefore, when we move to the second iteration, we follow an identical logic to show that messages $\tilde{h}_{i\to j}^\mathcal{T}$ in T_ℓ^v would be correct. Similarly, messages $\tilde{h}_{i\to j}^\mathcal{T}$ in T_ℓ^v for every iteration ℓ would be correct. □

Corollary 8. *The incoming messages $J_{i\to j}^\mathcal{T}, h_{i\to j}^p, \forall p \in \mathcal{F}$ of each node in $\mathcal{M}(w_\ell^\mathcal{T} \to v_\ell)$ are correct.*

Proof. The proof follows exactly the same logic as that of Prop. 4. □

Corollary 9. *The marginal at v_ℓ $(\mu_{v_\ell}, \sigma_{v_\ell}^2)$ is correct.*



(a) $t(k, j) = \ell$. (b) $t(k, j) = t$. (c) $t(k, j) = \infty$.

Figure 12. **Correctness of incoming messages of nodes in T_ℓ^v.** The incoming messages of every node in T_ℓ^v are correct. Red bold faced nodes represent the anchors and the fact that their potential vectors have been revised (changed). Tree $\mathcal{T}_j$ represents the tree rooted at node j, if we exclude the brach that contains edge (j, k). **(a)** Node w_ℓ is included in $\mathcal{T}_j$. **(b)** The most recent measurement from $\mathcal{T}_j$ has been taken at time $t(k, j) < \ell$. **(c)** No measurements have been received from tree $\mathcal{T}_j$.

Proof. If $v_\ell \in \mathcal{F}$, the marginal (mean and variance) have been correctly estimated in Eqs. (1), (2) as shown in Cor. 6. If $v_\ell \in \mathcal{T}$, by Cor. 7, Prop. 5, and Cor. 8, the revised potential at v_ℓ, $\tilde{h}_{v_\ell}$, and the incoming messages to node v_ℓ, $\tilde{h}^{\mathcal{T}}_{k\to v_\ell}$, $J^{\mathcal{T}}_{k\to v_\ell}$ are correct. Therefore, by Eqs. (3), (4), $\hat{h}^{\mathcal{T}}_{v_\ell}, \hat{J}^{\mathcal{T}}_{v_\ell,v_\ell}$ are correct, which makes the mean at v_ℓ, μ_{v_ℓ} correct. Lastly, since by Cor. 8, messages $h^p_{i\to j}, J^{\mathcal{T}}_{i\to j}, \forall p \in \mathcal{F}$ in $\mathcal{M}(w^{\mathcal{T}}_\ell \to v_\ell)$ are correct, it follows that the "feedback gain" at v_ℓ, g_{v_ℓ} is also correct and hence variance at v_ℓ, $\sigma^2_{v_\ell}$ as estimated by Eq. (5) will also be estimated correctly. □

Complexity. In standard FMP, $\mathcal{O}(|\mathcal{T}|K)$ messages are propagated in the first round. Then updated potential vector $\hat{h}_{\mathcal{F}}$ and information matrix $\hat{J}_{\mathcal{F}}$ are filled in $O(K^2 N_{\mathcal{F}})$ time, where $N_{\mathcal{F}} = \max_{p\in\mathcal{F}} |\mathcal{N}(p) \setminus \mathcal{F}|$. In other words, $N_{\mathcal{F}}$ is the size of the largest neighbor set of an FV (excluding any nodes belonging to FVS $\mathcal{F}$). Information matrix $\hat{J}_{\mathcal{F}}$ is inverted in $\mathcal{O}(K^3)$ time and produces revised potential vector in $\mathcal{O}(|\mathcal{A}|)$ time. Lastly, $\mathcal{O}(|\mathcal{T}|)$ computations are required for the second–round messages and $\mathcal{O}(K^2)$ for the correction of the variance at a single node. If $K \leq \min\{|\mathcal{T}|/N_{\mathcal{F}}, \sqrt{|\mathcal{T}|}\}$, the dominant term is $\mathcal{O}(K|\mathcal{T}|)$. If $|\mathcal{T}|/N_{\mathcal{F}} \leq K \leq N_{\mathcal{F}}$, the dominant term is $\mathcal{O}(K^2 N_{\mathcal{F}})$. If $K \geq \max\{N_{\mathcal{F}}, \sqrt{|\mathcal{T}|}\}$, the dominant term is $\mathcal{O}(K^3)$. In case the FVS is moderate in size (e.g., $K \leq \min\{|\mathcal{T}|/N_{\mathcal{F}}, \sqrt{|\mathcal{T}|}\}$), the dominant term is $O(|\mathcal{T}|K)$ which comes from the evaluation of first–round messages. By applying adaptive BP to the first– and second–round messages, we save a great deal of computations, since instead of sending a fixed number of $(K+3)|\mathcal{T}|$ messages per iteration, we send only the absolutely necessary messages.

We should remind that the subtrees of $\mathcal{T}$ rooted at $w^{\mathcal{T}}_\ell, v_\ell \in \mathcal{T}$ that pass through all the anchor nodes $\mathcal{A}$ are denoted by $\mathcal{T}^w_\ell$ and $\mathcal{T}^v_\ell$, respectively. See Fig. 11 for a visualization of trees $\mathcal{T}^w_\ell, \mathcal{T}^v_\ell$. Depending on the size of the anchor set, $\mathcal{A}$, and the allocation of its nodes inside $\mathcal{T}$, subtrees $\mathcal{T}^w_\ell, \mathcal{T}^v_\ell$ can be much smaller than $\mathcal{T}$, $|\mathcal{T}^w_\ell|, |\mathcal{T}^v_\ell| \ll |\mathcal{T}|$. Going back to the complexity analysis, we send $(K + 2)\mathrm{dist}(w^{\mathcal{T}}_{\ell-1}, w_\ell)$ messages, if $w_\ell \in \mathcal{T}$ and zero, otherwise. We additionally send $(K + 2)(|\mathcal{T}^w_\ell| - 1)$ between node $w^{\mathcal{T}}_\ell$ and the anchors $\mathcal{A}$. If, in addition, $v_\ell \in \mathcal{T}$, the propagation of $(|\mathcal{T}^v_\ell| - 1)$ $\tilde{h}^{\mathcal{T}}_{i\to j}$ messages between the anchors $\mathcal{A}$ and v_ℓ is necessary, plus $K\mathrm{dist}(w^{\mathcal{T}}_\ell, v_\ell)$ $h^p_{i\to j}$ messages from $w^{\mathcal{T}}_\ell$ to v_ℓ. Therefore, we send $\mathcal{O}(K(\mathrm{dist}(w^{\mathcal{T}}_{\ell-1}, w_\ell) + \mathrm{dist}(w^{\mathcal{T}}_\ell, v_\ell) + |\mathcal{T}^w_\ell|) + |\mathcal{T}^v_\ell|)$ messages per iteration.

Compare this to the $\mathcal{O}(K|\mathcal{T}|)$ messages per iteration of standard FMP. To understand the difference in complexity, let's assume for the shake of exposition that $|\mathcal{T}^w_\ell| \geq \mathrm{dist}(w^{\mathcal{T}}_{\ell-1}, w_\ell), \mathrm{dist}(w^{\mathcal{T}}_\ell, v_\ell), |\mathcal{T}^v_\ell|$. This means that the complexity of adaptive BP is $\mathcal{O}(K|\mathcal{T}^w_\ell|)$, which results in a speedup on the order of $\mathcal{O}(|\mathcal{T}|/|\mathcal{T}^w_\ell|)$, since it always holds that $|\mathcal{T}^w_\ell| \leq |\mathcal{T}|$. Throughout the analysis, we have made the assumption that FVS is moderate in size. For FVS comparable to the full graph size N, the proposed method (as well as standard FMP) will perform poorly, since the dominant computation per iteration would be $\mathcal{O}(K^3)$ and is comparable to the complexity of the inversion of the full information matrix J.

Determining a nearly optimal measurement schedule

We have made the assumption that the measurement order is not known to us in advance. An equally interesting problem arises when we are given constraints on the number of measurements we can draw from each latent node and the task is to construct an optimal schedule of obtaining them. More formally, suppose we can draw k_t measurements from X_t and we draw measurements from S distinct latent nodes.[2] Obviously, the schedule should be designed in such a way that it would result in the minimum number of propagated messages. Since there is no propagation of messages when measurements are taken consecutively from the same node, we can reduce this problem to one where there is one measurement vector (of size k_t) for each of the S nodes. In other words, once we reach a node X_t (dictated by the measurement schedule), we will draw k_t measurements from that node. Even though we can find the optimal solution to the above problem for small S, the exhaustive search becomes intractable as S grows, since there are $S!$ possible solutions. The problem of determining an optimal schedule of measurements that visits each of the S nodes exactly once, which corresponds to finding a schedule with the minimum number of computations, can be reduced to the shortest Hamiltonian path problem. As a reminder, a *Hamiltonian path* is a path that visits each node

[2] As a reminder, there are N latent nodes in total.

exactly once. A *Hamiltonian cycle* is a cycle that visits each node exactly once except for the starting node, which is visited twice. A graph that contains a Hamiltonian cycle is called a Hamiltonian graph. A graph that has a Hamiltonian cycle has trivially a Hamiltonian path as well, since the edge between the last node in the visitation order and the starting node can be removed. The shortest Hamiltonian path problem has shown to be NP-complete (Arora & Barak, 2009).

To formulate the shortest Hamiltonian path problem, we are given a set of nodes $X_1, \dots, X_S$ that form an edge set $\mathcal{E}$. For every edge $(i, j) \in \mathcal{E}$ linking two nodes, there is a non-negative distance (cost) $\text{dist}(i, j)$ associated with them. The goal is to find an ordering $\boldsymbol{w}$, where each node is visited exactly once, that minimizes the total distance traveled

$$\max_{\boldsymbol{w}} \sum_{j=1}^{S-1} \text{dist}(w_j, w_{j+1}).$$

When the triangle inequality holds, that is, for every triplet $(i, j), (i, k), (j, k) \in \mathcal{E}$, $\text{dist}(i, j) \le \text{dist}(i, k) + \text{dist}(j, k)$, there are approximate techniques with nice theoretical guarantees that provide nearly optimal solutions. One algorithm that runs in polynomial time $\mathcal{O}(S^3)$ is a variant of Christofides' algorithm, which was initially designed for the Traveling Salesman Problem (TSP) (Christofides, 1976). The TSP is very related to the shortest Hamiltonian path, since the objective is the same with the additional constraint that at the end of the visitation order, we return to the starting point. In other words, it is a shortest Hamiltonian cycle problem. The variant of Christofides' algorithm that gives an approximate solution for the shortest Hamiltonian path problem is proposed in (Hoogeveen, 1991). This algorithm serves as a 3/2 approximation in the worst case.

We convert the problem of finding a schedule of minimum computations to a shortest Hamiltonian path as follows. We concatenate all k_t measurements of variable X_t into one vector of measurements. Since we draw measurements from S latent nodes, we compute the distance between every pair of latent nodes as

$$\text{dist}(i, j) = D_i + D_j - 2D_{\text{lca}(i,j)}, \quad (6)$$

where D is the depth of a node and $\text{lca}(i, j)$ is the lowest common ancestor of i, j, which is recovered in constant time through the reduction to the RMQ problem. With this approach, we form a full undirected graph of S nodes, where each edge is weighted by the distance between the incident nodes. This graph is guaranteed to have a Hamiltonian path, since Dirac (1952) showed that a simple graph with S vertices with $S \ge 3$ is Hamiltonian if every node has degree $S/2$ or greater, which applies to full graphs. You can see a visualization of the measurement plan designation in Fig. 13.

If we denote the length of the nearly shortest Hamiltonian path by ℓ_H, then in the Gaussian case, the overall complexity of message passing would be $\mathcal{O}(\ell_H d^3)$, where d is the dimension of latent variables. If, in addition, the dimension d is comparable to the number of latent variables N, the complexity of finding a shortest Hamiltonian path $\mathcal{O}(S^3)$ does not affect the overall complexity (in asymptotic terms), since $S \le N$.

Computing messages in path $\mathcal{M}(w_\ell \to w_{\ell+1})$

When we move on computing messages from $\mathcal{M}(w_\ell \to w_{\ell+1})$, we do not have to recompute the messages that are common with those in the path $\mathcal{M}(w_\ell \to v_\ell)$, that have already been evaluated. More specifically,

$$\begin{aligned} &\mathcal{M}(w_\ell \to w_{\ell+1}) \\ &= \begin{cases} \{\mathcal{M}(w_\ell \to \text{lca}(w_\ell, v_\ell)), \quad \mathcal{M}(\text{lca}(w_\ell, v_\ell) \to w_{\ell+1})\} \\ \qquad\qquad , \text{if } \text{lca}(w_\ell, w_{\ell+1}) = \text{lca}(v_\ell, w_{\ell+1}) \\ \{\mathcal{M}(w_\ell \to \text{lca}(w_\ell, w_{\ell+1})), \mathcal{M}(\text{lca}(w_\ell, w_{\ell+1}) \to w_{\ell+1})\} \\ \qquad\qquad , \text{if } \text{lca}(w_\ell, w_{\ell+1}) > \text{lca}(v_\ell, w_{\ell+1}) \\ \{\mathcal{M}(w_\ell \to \text{lca}(v_\ell, w_{\ell+1})), \ \mathcal{M}(\text{lca}(v_\ell, w_{\ell+1}) \to w_{\ell+1})\} \\ \qquad\qquad , \text{if } \text{lca}(w_\ell, w_{\ell+1}) < \text{lca}(v_\ell, w_{\ell+1}) \end{cases} \end{aligned}$$

The first part of path $\mathcal{M}(w_\ell \to w_{\ell+1})$ is already evaluated during the computation of messages in the path $\mathcal{M}(w_\ell \to v_\ell)$. We can omit computing the messages in the common path by recovering the lca of nodes $(v_\ell, w_{\ell+1})$, which is accomplished in constant time.

References

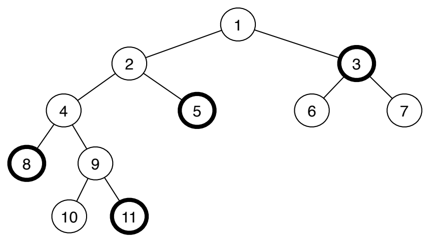
Arora, S. and Barak, B. *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.

Bafna, V., Berman, P., and Fujito, T. A 2-Approximation Algorithm for the Undirected Feedback Vertex Set Problem. *SIAM Journal on Discrete Mathematics*, 12(3):289–297, Sep 1999. ISSN 0895-4801.

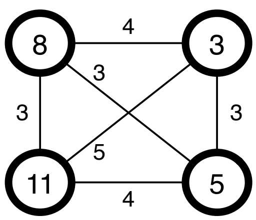
Christofides, N. Worst-case analysis of a new heuristic for the traveling salesman problem. Technical Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, 1976.

Dirac, G. A. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, 3(1):69–81, 1952.

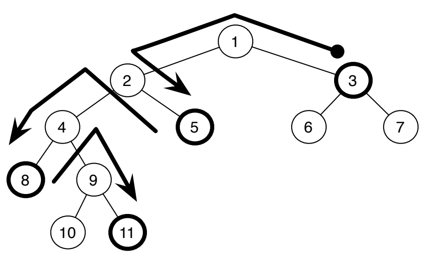
Hoogeveen, J. A. Analysis of Christofides' heuristic: Some paths are more difficult than cycles. *Operations Research Letters*, 10 (5):291–295, 1991.



(a) Original graph.

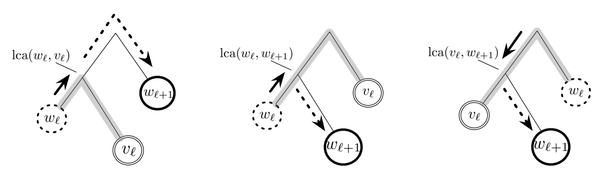


(b) Full graph of measurement nodes.



(c) Minimum computational schedule.

Figure 13. **Reduction of finding optimal schedules to shortest Hamiltonian path. (a)** The nodes where we would obtain measurements from are $3, 5, 8, 11$ (depicted as boldface). Our task is to design a measurement plan with the minimum number of messages for inference purposes. **(b)** We form a full undirected graph comprised of the measurement nodes. The weight of each edge would be the distance between these two nodes in the original graph, calculated by Eq. (6). **(c)** The path shown is one possible optimal solution. The arrow with a circle in one end indicates the starting node of sequence $\boldsymbol{w}$.



(a) (b) (c)

Figure 14. **(a)** $\mathcal{M}(w_\ell \rightarrow w_{\ell+1}) = \{\mathcal{M}(w_\ell \rightarrow \mathrm{lca}(w_\ell, v_\ell)), \mathcal{M}(\mathrm{lca}(w_\ell, v_\ell) \rightarrow w_{\ell+1})\}$. In this case, next measurement node, $w_{\ell+1}$, is outside the subtree of w_ℓ, v_ℓ. Bold arrow indicates the already computed messages in $\mathcal{M}(w_\ell \rightarrow \mathrm{lca}(w_\ell, v_\ell))$, while the dotted the ones to be computed. **(b)** $\mathcal{M}(w_\ell \rightarrow w_{\ell+1}) = \{\mathcal{M}(w_\ell \rightarrow \mathrm{lca}(w_\ell, w_{\ell+1})), \mathcal{M}(\mathrm{lca}(w_\ell, w_{\ell+1}) \rightarrow w_{\ell+1})$. In this case, next measurement node, $w_{\ell+1}$, is inside the subtree of w_ℓ, v_ℓ and closer to w_ℓ. Bold arrow indicates the already computed messages in $\mathcal{M}(w_\ell \rightarrow \mathrm{lca}(w_\ell, w_{\ell+1}))$, while the dotted the ones to be computed. **(c)** $\mathcal{M}(w_\ell \rightarrow w_{\ell+1}) = \{\mathcal{M}(w_\ell \rightarrow \mathrm{lca}(v_\ell, w_{\ell+1})), \mathcal{M}(\mathrm{lca}(v_\ell, w_{\ell+1}) \rightarrow w_{\ell+1})\}$. In this case, next measurement node, $w_{\ell+1}$, is inside the subtree of w_ℓ, v_ℓ and closer to v_ℓ. Bold arrow indicates the already computed messages in $\mathcal{M}(w_\ell \rightarrow \mathrm{lca}(v_\ell, w_{\ell+1}))$, while the dotted the ones to be computed. Gray bands encompass all the messages in $\mathcal{M}(w_\ell \rightarrow v_\ell)$ that have been correctly updated from the previous iteration.