`involve-MI`: Informative Planning with High-Dimensional Non-Parametric Beliefs*

Supplementary Material

Gilad Rotman and Vadim Indelman

Technion – Israel Institute of Technology, Haifa, 3200003, Israel,
gilad.rotman@campus.technion.ac.il, vadim.indelman@technion.ac.il

This document provides supplementary material to [4]. Therefore, it should not be considered a self-contained document, but instead regarded as an appendix of [4]. Throughout this report, all notations and definitions are with compliance to the ones presented in [4].

1 Information-theoretic rewards

In this section we provide the definitions of the original Information Gain (IG) and Mutual Information (MI). IG quantifies the amount of information gained for a certain variable X (the state) by knowing the value of another variable Z (an observation). It is defined as the difference between the entropy prior to this additional knowledge and the entropy afterwards

$$IG\left[X; Z=z\right] \triangleq \mathcal{H}\left[X\right] - \mathcal{H}\left[X \mid Z=z\right]. \tag{1}$$

MI is IG in expectation, and it can also be defined as the difference between the entropy of the state and the expected entropy of the state given an observation

$$I\left[X;Z\right] \triangleq \mathop{\mathbb{E}}_{\mathcal{Z}}\Big[IG\left[X;Z=z\right]\Big] = \mathcal{H}\left[X\right] - \mathcal{H}\left[X \mid Z\right]. \tag{2}$$

where $\mathcal{H}\left[X \mid Z\right] = \mathop{\mathbb{E}}_{\mathcal{Z}}\big[\mathcal{H}\left[X \mid Z=z\right]\big]$. For the case where the state changes between time steps, as discussed in the paper, the original definitions of IG and MI are insufficient, since these do not account for the additional uncertainty obtained by changes in the state.

2 `involve-MI`

2.1 Proof of Lemma 1

We remind the augmented MI is

$$I_{aug}\left[X \boxplus X_{new}; Z\right] \triangleq \mathcal{H}\left[X\right] - \mathcal{H}\left[X, X_{new} \mid Z\right]. \tag{3}$$

*This work was partially supported by the Israel Science Foundation (ISF) and by US NSF/US-Israel BSF.

Using the following known identities

$$\mathcal{H}[X] \triangleq \mathcal{H}[X, X_{new}] - \mathcal{H}[X_{new} \mid X] \tag{4}$$

$$I[X, X_{new}; Z] \triangleq \mathcal{H}[X, X_{new}] - \mathcal{H}[X, X_{new} \mid Z], \tag{5}$$

we get the relation between MI and the augmented MI

$$I_{aug}[X \boxplus X_{new}; Z] = I[X, X_{new}; Z] - \mathcal{H}[X_{new} \mid X]. \tag{6}$$

2.2 Proof of Theorem 1

Using the chain rule for MI, where the state is arbitrarily partitioned as $X' = \{X^A, X^B\}$ yields

$$I[X'; Z] = I\left[X^A, X^B; Z\right] = I\left[X^A; Z\right] + I\left[X^B; Z \mid X^A\right]. \tag{7}$$

By definition, the conditional MI term is

$$\begin{aligned} I\left[X^B; Z \mid X^A\right] \triangleq \int_Z \int_{X^A} \int_{X^B} \mathbb{P}(Z, X^A, X^B) \cdot \\ \cdot \log\left[\frac{\mathbb{P}(X^B, Z \mid X^A)}{\mathbb{P}(X^B \mid X^A)\mathbb{P}(Z \mid X^A)}\right] dX^B dX^A dZ. \end{aligned} \tag{8}$$

Using the chain rule over the numerator inside the log term, we get

$$\mathbb{P}(X^B, Z \mid X^A) = \mathbb{P}(Z \mid X^A, X^B)\mathbb{P}(X^B \mid X^A). \tag{9}$$

Defining $X^A \triangleq \left\{X^{in+}, X_{new}\right\}$, meaning it contains all the variables involved in generating the observations Z (and all new states X_{new}), we can state that $\mathbb{P}(Z \mid X^A, X^B) = \mathbb{P}(Z \mid X^A)$, so eq. (9) becomes

$$\mathbb{P}(X^B, Z \mid X^A) = \mathbb{P}(Z \mid X^A)\mathbb{P}(X^B \mid X^A). \tag{10}$$

Plugging this term back into eq. (8) yields

$$I\left[X^B; Z \mid X^A\right] \triangleq \int_Z \int_{X^A} \int_{X^B} \mathbb{P}(Z, X^A, X^B) \log(1)\, dX^B dX^A dZ = 0. \tag{11}$$

Using the above result, eq. (7) then transforms into

$$I[X'; Z] = I\left[X^{in+}, X_{new}; Z\right]. \tag{12}$$

With our definition of X^A, the prior state can be written as $X = \left\{X^{in+}, X^B\right\}$. Looking then at the conditional entropy term in the result of Lemma 1, we can rewrite it as $\mathcal{H}[X_{new} \mid X] = \mathcal{H}\left[X_{new} \mid X^{in+}, X^B\right]$. By our definition of X^{in+}, X_{new} is conditionally independent of X^B given X^{in+}, i.e. $\mathbb{P}(X_{new} \mid X^{in+}, X^B) = \mathbb{P}(X_{new} \mid X^{in+})$. Thus, one of the conditional entropy properties states that

$$\mathcal{H}[X_{new} \mid X] = \mathcal{H}\left[X_{new} \mid X^{in+}\right] \tag{13}$$

Plugging (12) and (13) back into the result of Lemma 1 (eq. (6)) we get that

$$I_{aug}\left[X \boxplus X_{new}; Z\right] = I\left[X^{in+}, X_{new}; Z\right] - \mathcal{H}\left[X_{new} \mid X^{in+}\right]. \tag{14}$$

We then observe that by using the result from eq. (6), the right hand side in eq. (14) is equal to $I_{aug}\left[X^{in+} \boxplus X_{new}; Z\right]$, and so we finally conclude that

$$I_{aug}\left[X \boxplus X_{new}; Z\right] = I_{aug}\left[X^{in+} \boxplus X_{new}; Z\right]. \tag{15}$$

3 MI-SMC

3.1 Proof of Theorem 2

We begin by using the definition of the augmented MI over the involved subset, which is

$$I_{aug}\left[X^{in} \boxplus X_{new}; Z\right] \triangleq \mathcal{H}\left[X^{in}\right] - \mathcal{H}\left[X^{in}, X_{new} \mid Z\right], \tag{16}$$

where we remind that we use X^{in} instead of X^{in+} for the readability of the paper, yet the analysis is true for the more general subset X^{in+}. Using the chain rule for conditional entropy over the second term on the right hand side of eq. (16) yields

$$\mathcal{H}\left[X^{in}, X_{new} \mid Z\right] = \mathcal{H}\left[X^{in}, X_{new}, Z\right] - \mathcal{H}\left[Z\right]. \tag{17}$$

Using the same principle twice again eventually yields

$$\mathcal{H}\left[X^{in}, X_{new} \mid Z\right] = \mathcal{H}\left[X^{in}\right] + \mathcal{H}\left[X_{new} \mid X^{in}\right] + \mathcal{H}\left[Z \mid X^{in}, X_{new}\right] - \mathcal{H}\left[Z\right]. \tag{18}$$

Plugging back into eq. (16), we observe that the term $\mathcal{H}\left[X^{in}\right]$ is canceled out. Then, by using the result of Theorem 1, given in eq. (15), the augmented MI term over the high-dimensional state finally becomes

$$I_{aug}\left[X \boxplus X_{new}; Z\right] = -\mathcal{H}\left[X_{new} \mid X^{in}\right] - \mathcal{H}\left[Z \mid X^{in}, X_{new}\right] + \mathcal{H}\left[Z\right]. \tag{19}$$

3.2 Developing the estimator

We remind that the augmented MI can be written as

$$\begin{aligned} I_{aug}\left[X \boxplus X_{new}; Z\right] = &\int_{\mathcal{X}^{in}} b\left[X^{in}\right] \left[\int_{\mathcal{X}_{new}} \mathcal{P}_T \log \mathcal{P}_T dX_{new}\right] dX^{in} \\ &+ \int_{\mathcal{X}^{in}} b\left[X^{in}\right] \left[\int_{\mathcal{X}_{new}} \mathcal{P}_T \left[\int_{\mathcal{Z}} \mathcal{P}_Z \log \mathcal{P}_Z dZ\right] dX_{new}\right] dX^{in} \\ &- \int_{\mathcal{X}^{in}} b\left[X^{in}\right] \left[\int_{\mathcal{X}_{new}} \mathcal{P}_T \left[\int_{\mathcal{Z}} \mathcal{P}_Z \log \eta^{-1} dZ\right] dX_{new}\right] dX^{in}, \end{aligned} \tag{20}$$

where the normalizer can be calculated with

$$\eta^{-1} = \int_{\mathcal{X}^{in}} b\left[X^{in}\right] \left[\int_{\mathcal{X}_{new}} \mathcal{P}_T \mathcal{P}_Z dX_{new}\right] dX^{in}. \tag{21}$$

We then approach to sampling, i.e.

$$\begin{aligned} \left(x^{in(i)}, w^{(i)}\right) &\sim b\left[X^{in}\right] \\ x_{new}^{(i,j)} &\sim \mathcal{P}_T\left(X_{new} \mid x^{in(i)}\right) \\ z^{(i,j,k)} &\sim \mathcal{P}_Z\left(Z \mid x^{in(i)}, x_{new}^{(i,j)}\right), \end{aligned} \tag{22}$$

which allows the augmented MI to be approximated as

$$\begin{aligned} I_{aug}\left[X \boxplus X_{new}; Z\right] \approx & \sum_{i=1}^{n_1} w^{(i)} \left[\frac{1}{n_2} \sum_{j=1}^{n_2} \log \mathcal{P}_T^{(i,j)}\right] \\ & + \sum_{i=1}^{n_1} w^{(i)} \left[\frac{1}{n_2} \sum_{j=1}^{n_2} \left[\frac{1}{n_3} \sum_{k=1}^{n_3} \log \mathcal{P}_Z^{(i,j,k)}\right]\right] \\ & - \sum_{i=1}^{n_1} w^{(i)} \left[\frac{1}{n_2} \sum_{j=1}^{n_2} \left[\frac{1}{n_3} \sum_{k=1}^{n_3} \log \eta^{-1(i,j,k)}\right]\right], \end{aligned} \tag{23}$$

where

$$\begin{aligned} \mathcal{P}_T^{(i,j)} &= \mathcal{P}_T\left(x^{in(i)}, x_{new}^{(i,j)}\right) \\ \mathcal{P}_Z^{(i,j,k)} &= \mathcal{P}_Z\left(x^{in(i)}, x_{new}^{(i,j)}, z^{(i,j,k)}\right) \\ \eta^{-1(i,j,k)} &= \eta^{-1}\left(x^{in(i)}, x_{new}^{(i,j)}, z^{(i,j,k)}\right), \end{aligned} \tag{24}$$

and the normalizer, for each sampled instance, is then also approximated as

$$\eta^{-1(i,j,k)} \approx \sum_{l=1}^{n_4} w^{(l)} \left[\frac{1}{n_5} \sum_{m=1}^{n_5} \mathcal{P}_Z^{(l,m,k)}\right]. \tag{25}$$

Remark: As in a particle filter, $\mathcal{P}_Z^{(l,m,k)}$ can be considered an update for the particle's weight. Thus, the approximation of $\eta^{-1(i,j,k)}$ can be viewed as an average of the updated weights.

3.3 Complexity

Complexity analysis

In terms of complexity, the most expensive step of this approach is the estimation of $\mathcal{H}\left[Z\right]$, thus its complexity is the complexity of the entire estimator.

Estimating each $\eta^{-1^{(i,j,k)}}$ has a complexity of $O\left(n_4 n_5 d\right)$. In turn, the complexity of estimating $\mathcal{H}\left[Z\right]$ is of $O\left(n_1 n_2 n_3 n_4 n_5 d\right)$. Considering that we have a total number of m observations instances, i.e. $n_1 n_2 n_3 = m$, and also that the total number of particles is n, i.e. $n_4 n_5 = n$, the complexity becomes $O\left(mnd\right)$.

In comparison to two more estimators

Many other entropy estimators exist in the literature, such as the nearest neighbor estimator, which can be found in [1], and the k-d partitioning estimator, presented in [5]. When estimating the MI value with these estimators, the complexity of both can get to $O\left(mn \log n\right)$, which is comparable to the complexity of `MI-SMC` when reminding that n should be exponential in the dimension d.

4 Applicability to belief trees

In this section we wish to relate the approaches in the paper to the informative planning optimization problem. We remind that although the following analysis considers an open-loop formulation, for which we seek for an optimal action sequence, $a_{0:T-1}$, it also applies for a close-loop formulation, for which we seek for a policy, $\pi_{0:T-1}$. The solution to the ρ-POMDP is obtained by maximization of the objective function, denoted shortly as $J_0 \triangleq J\left(b\left[X_0\right], a_{0:T-1}\right)$

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \mathop{\mathbb{E}}_{\mathcal{Z}_{1:T}} \left[\sum_{t=0}^{T-1} \rho_t + \rho_T \right] \right\}. \tag{26}$$

Formulating it recursively yields the Bellman optimality equation

$$J_t^\star = \max_{a_t} \left\{ \rho_t + \mathop{\mathbb{E}}_{\mathcal{Z}_{t+1}} \left[J_{t+1}^\star \right] \right\}. \tag{27}$$

where $J_t \triangleq J\left(b\left[X_t\right], a_{t:T-1}\right)$.

A common solver to this optimization problem is to construct a search over a tree. More specifically, for ρ-POMDP, which is the case of belief-dependent rewards, a belief tree is used. In a belief tree, the beliefs $b\left[X_t\right]$ are propagated using instances of future actions and observations, then the rewards ρ_t are calculated, and the action sequence providing the maximum value for the objective function is eventually chosen. Since, in general, the action and observation spaces can be large, in order to be able to solve this optimization problem in reasonable time, it is approximated with a belief tree which propagates only a few sampled instances of future actions and observations. Dealing with continuous such spaces, a belief tree is an approximation of the problem to begin with. The planning literature contains lots of tree-based solvers. However, since our analysis so far was done considering an expected reward, augmented MI, it is not trivial to prove that our approach, `involve-MI`, and our estimator, `MI-SMC`, are able to cope with such solvers. This is the purpose of this section.

We denote the augmented IG, the augmented MI and their involved counterparts shortly as

$$\begin{aligned} IG_0^t &\triangleq IG_{aug}\left[X_0 \boxplus x_{1:t}; Z_{1:t} = z_{1:t} \mid a_{0:t-1}\right] \\ I_0^t &\triangleq I_{aug}\left[X_0 \boxplus x_{1:t}; Z_{1:t} \mid a_{0:t-1}\right] \\ {IG_0^t}^{in} &\triangleq IG_{aug}\left[X_0^{in} \boxplus x_{1:t}; Z_{1:t} = z_{1:t} \mid a_{0:t-1}\right] \\ {I_0^t}^{in} &\triangleq I_{aug}\left[X_0^{in} \boxplus x_{1:t}; Z_{1:t} \mid a_{0:t-1}\right]. \end{aligned}$$

We will also from now omit the term "augmented" while still referring to the more general case of augmentation. For readability, our analysis is done for IG as the only term of the reward, meaning $\rho_t = IG_0^t, \forall\, t \in [1, T]$. Yet, the conclusions will also apply when there are additional terms for the reward, state-based terms for example. Eq. (26) then becomes

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \sum_{t=0}^{T} \mathop{\mathbb{E}}_{\mathcal{Z}_{1:T}} \left[IG_0^t\right] \right\} = \max_{a_{0:T-1}} \left\{ \sum_{t=0}^{T} I_0^t \right\}. \tag{28}$$

Using Theorem 1 over this equation yields

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \sum_{t=0}^{T} {I_0^t}^{in} \right\}. \tag{29}$$

Theorem 2 *Let us define a new reward, $\rho_t^{in} = {IG_0^t}^{in}$. Solving the ρ-POMDP optimization problem with this reward is equivalent to solving it with the original reward, $\rho_t = IG_0^t$, such that*

$$J_t^\star = \max_{a_t} \left\{ \rho_t^{in} + \mathop{\mathbb{E}}_{\mathcal{Z}_{t+1}} \left[J_{t+1}^\star\right] \right\}. \tag{30}$$

Proof

We remind eq. (29) is

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \sum_{t=0}^{T} {I_0^t}^{in} \right\}. \tag{31}$$

The involved MI is by definition an expectation over the involved IG

$${I_0^t}^{in} \triangleq \mathop{\mathbb{E}}_{\mathcal{Z}_{1:T}} \left[{IG_0^t}^{in} \right]. \tag{32}$$

Plugging this back into eq. (31) yields

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \sum_{t=0}^{T} \left[\mathop{\mathbb{E}}_{\mathcal{Z}_{1:t}} \left[{IG_0^t}^{in} \right] \right] \right\}. \tag{33}$$

Due to commutativity, we can again switch between the order of expectation and summation, which yields

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \mathop{\mathbb{E}}_{\mathcal{Z}_{1:T}} \left[\sum_{t=0}^{T} \left[{IG_0^t}^{in} \right] \right] \right\}. \tag{34}$$

We then separate the first action a_0 from the rest of the actions $a_{1:T-1}$. We also observe that Z_1 is not a function of $a_{1:T-1}$ and that ${IG_0^0}^{in}$ is not a function of both $a_{1:T-1}$ and Z_1. This yields

$$J_0^\star = \max_{a_0} \left\{ {IG_0^0}^{in} + \mathop{\mathbb{E}}_{\mathcal{Z}_1} \left[\max_{a_{1:T-1}} \left\{ \mathop{\mathbb{E}}_{\mathcal{Z}_{2:T}} \left[\sum_{t=1}^{T} {IG_0^t}^{in} \right] \right\} \right] \right\}. \tag{35}$$

We then observe that the term inside the expectation over Z_1 is equal to $J_1^\star$, which yields the following recursive form

$$J_0^\star = \max_{a_0} \left\{ {IG_0^0}^{in} + \mathop{\mathbb{E}}_{\mathcal{Z}_1} \left[J_1^\star \right] \right\}, \tag{36}$$

and, in general, for each $t \in [1, T-1]$

$$J_t^\star = \max_{a_t} \left\{ {IG_0^t}^{in} + \mathop{\mathbb{E}}_{\mathcal{Z}_{t+1}} \left[J_{t+1}^\star \right] \right\}. \tag{37}$$

We observe that this is the Bellman optimality equation with a new reward, $\rho_t^{in} \triangleq {IG_0^t}^{in}$. This eventually means that Solving the ρ-POMDP optimization problem with this reward is equivalent to solving it with the original reward we have started with, $\rho_t = IG_0^t$.

This, in turn, means that any optimization solver suitable for the original problem, with the reward ρ_t, is also suitable when changing it to ρ_t^{in}. This is a key result, since in general these rewards are not equal, however it is much more efficient to calculate ρ_t^{in}, as already discussed for the MI case. The belief tree which resembles this new equivalent optimization problem is shown in Figure 1a. We can then use the IG definition and calculate it through the entropy terms. This result is again general, but the non-parametric setting then necessitate the usage of entropy estimators, plenty of which exist in the literature, as already discussed.

Proposition 1 *Naively calculating the values I_0^t yields a degenerate belief tree, in which there are only action nodes, without observation nodes.*

Proof

We remind eq. (29) is

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \sum_{t=0}^{T} {I_0^t}^{in} \right\}. \tag{38}$$

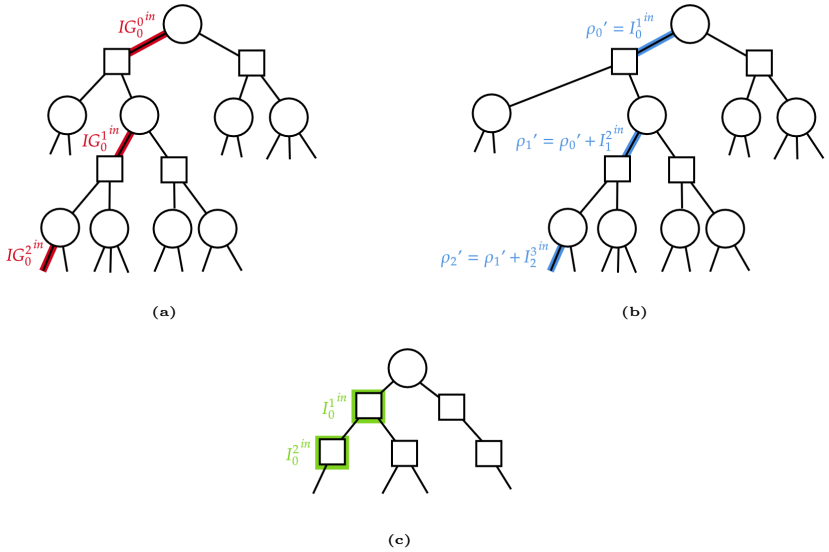

Fig. 1: Belief trees over the involved variables only, which resemble optimization problems equivalent to the original optimization problem. Circles are observation nodes, squares are action nodes. (a) shows a belief tree where the rewards are sequential involved IGs; (b) shows a belief tree where the rewards are updated incrementally with consecutive involved MIs; (c) shows the resultant degenerate belief tree when trying to directly go through the calculation of the sequential involved MIs. It is degenerate in the sense that there are only action nodes, without observation nodes.

Since by definition $I_0^{0\,in} = 0$, we can start the summation from $t = 1$

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \sum_{t=1}^{T} I_0^{t\,in} \right\}. \tag{39}$$

We then separate the first action a_0 from the rest of the actions $a_{1:T-1}$. We also observe that $I_0^{1\,in}$ is not a function of $a_{1:T-1}$. This yields

$$J_0^\star = \max_{a_0} \left\{ I_0^{1\,in} + \max_{a_{1:T-1}} \left\{ \sum_{t=2}^{T-1} I_0^{t\,in} \right\} \right\}. \tag{40}$$

We then observe that the term $\max_{a_{1:T-1}} \left\{ \sum_{t=2}^{T-1} I_0^{t\,in} \right\}$ is equal to $J_1^\star$, which yields the following recursive form

$$J_0^\star = \max_{a_0} \left\{ I_0^{1\,in} + J_1^\star \right\}, \tag{41}$$

and, in general, $\forall t \in [1, T-1]$

$$J_t^\star = \max_{a_t} \left\{ I_0^{t+1\,in} + J_{t+1}^\star \right\}. \tag{42}$$

We observe that this recursive form is slightly different than the Bellman optimality equation. The Bellman optimality equation, as can be seen in eq. (27), includes also expectation over future observations, while in this formulation it is omitted (more specifically, it is considered at the level of calculating the values I_0^t). This, in turn, means that a corresponding tree will lack observation nodes, thus it will be a degenerate belief-tree. We note again that this is the result of naively going through direct calculations of the values I_0^t.

This type of a tree can be seen in Figure 1c. We do not here analyze whether this formulation is good or bad compared to the standard formulation nor whether it would even suit a policy formulation or not. We leave it for future research. We cling to the fact that none of the state-of-the-art tree-based solvers work this way, and suggest another approach.

Lemma 2 *Let I_0^t denote a sequential MI between times 0 and t, and $I_{i-1}^i = I_{aug}\left[X_{i-1} \boxplus x_i; Z_i \mid h_i^-\right]$ denote a consecutive MI between times $i-1$ and i, where $h_i^- = \{z_{1:i-1}, a_{0:i-1}\}$ is the history up to time i, without the last observation z_i. The sequential MI can be decomposed into multiple consecutive MI values, such that*

$$I_0^t \triangleq \sum_{i=1}^{t} \left[\mathop{\mathbb{E}}_{\mathcal{Z}_{1:i-1}} \left[I_{i-1}^i \right] \right]. \tag{43}$$

Proof

We remind that the sequential augmented MI is defined as

$$I_0^t \triangleq I_{aug}\left[X_0 \boxplus x_{1:t}; Z_{1:t} \mid a_{0:t-1}\right] \triangleq \mathop{\mathbb{E}}_{\mathcal{Z}_{1:t}} \left[IG_{aug}\left[X_0 \boxplus x_{1:t}; Z_{1:t} = z_{1:t} \mid a_{0:t-1}\right] \right]. \tag{44}$$

Detaching the observations $Z_{l+1:t}$, where $0 < l < t$, and expressing the augmented IG with entropies, we get

$$I_{aug}\left[X_0 \boxplus x_{1:t}; Z_{1:t} \mid a_{0:t-1}\right] = \mathop{\mathbb{E}}_{\mathcal{Z}_{1:l}} \left[\mathop{\mathbb{E}}_{\mathcal{Z}_{l+1:t}} \left[\mathcal{H}\left[X_0\right] - \mathcal{H}\left[X_t \mid h_t\right] \right] \right]. \tag{45}$$

Adding and subtracting the term $\mathcal{H}\left[X_l \mid h_l\right]$, it becomes

$$\begin{aligned} I_{aug}\left[X_0 \boxplus x_{1:t}; Z_{1:t} \mid a_{0:t-1}\right] = \mathop{\mathbb{E}}_{\mathcal{Z}_{1:l}} \Bigg[\mathop{\mathbb{E}}_{\mathcal{Z}_{l+1:t}} \Big[\left\{ \mathcal{H}\left[X_0\right] - \mathcal{H}\left[X_l \mid h_l\right] \right\} + \\ + \left\{ \mathcal{H}\left[X_l \mid h_l\right] - \mathcal{H}\left[X_t \mid h_t\right] \right\} \Big] \Bigg]. \end{aligned} \tag{46}$$

Observing that both new differences are augmented IGs as well, and that the first difference is not a function of the last observation, we get

$$\begin{aligned} I_{aug}\left[X_0 \boxplus x_{1:t}; Z_{1:t} \mid a_{0:t-1}\right] = \mathop{\mathbb{E}}_{\mathcal{Z}_{1:l}} \Bigg[IG_{aug}\left[X_0 \boxplus x_{1:l}; z_{1:l} \mid a_{0:l-1}\right] + \\ + \mathop{\mathbb{E}}_{\mathcal{Z}_{l+1:t}} \left[IG_{aug}\left[X_l \boxplus x_{l+1:t}; z_{l+1:t} \mid a_{0:t-1}, z_{1:l}\right] \right] \Bigg], \end{aligned} \tag{47}$$

The expectation over the augmented IG is the augmented MI, and so we get the following recursive form

$$I_{aug}\left[X_0 \boxplus x_{1:t}; Z_{1:t} \mid a_{0:t-1}\right] = I_{aug}\left[X_0 \boxplus x_{1:l}; Z_{1:l} \mid a_{0:l-1}\right] + \\ + \mathop{\mathbb{E}}_{\mathcal{Z}_{1:l}}\Big[I_{aug}\left[X_l \boxplus x_{l+1:t}; Z_{l+1:t} \mid a_{0:t-1}, z_{1:l}\right]\Big]. \tag{48}$$

The specific case of choosing $l = t-1$ yields

$$I_{aug}\left[X_0 \boxplus x_{1:t}; Z_{1:t} \mid a_{0:t-1}\right] = I_{aug}\left[X_0 \boxplus x_{1:t-1}; Z_{1:t-1} \mid a_{0:t-2}\right] + \\ + \mathop{\mathbb{E}}_{\mathcal{Z}_{1:t-1}}\Big[I_{aug}\left[X_{t-1} \boxplus x_t; Z_t \mid h_t^-\right]\Big], \tag{49}$$

where $h_t^- = \{z_{1:t-1}, a_{0:t-1}\}$ is the history up to time t, without the last observation z_t. Opening the recursive form of the sequential augmented MI in eq. (49) yields

$$I_{aug}\left[X_0 \boxplus x_{1:t}; Z_{1:t} \mid a_{0:t-1}\right] = I_{aug}\left[X_0 \boxplus x_1; Z_1 \mid h_1^-\right] + \\ + \mathop{\mathbb{E}}_{\mathcal{Z}_1}\Big[I_{aug}\left[X_1 \boxplus x_2; Z_2 \mid h_2^-\right]\Big] + \cdots + \mathop{\mathbb{E}}_{\mathcal{Z}_{1:t-1}}\Big[I_{aug}\left[X_{t-1} \boxplus x_t; Z_t \mid h_t^-\right]\Big], \tag{50}$$

which can more compactly be written as

$$I_{aug}\left[X_0 \boxplus x_{1:t}; Z_{1:t} \mid a_{0:t-1}\right] = \sum_{i=1}^{t}\left[\mathop{\mathbb{E}}_{\mathcal{Z}_{1:i-1}}\Big[I_{aug}\left[X_{i-1} \boxplus x_i; Z_i \mid h_i^-\right]\Big]\right]. \tag{51}$$

Returning to the short notations, we finally get

$$I_0^t = \sum_{i=1}^{t}\left[\mathop{\mathbb{E}}_{\mathcal{Z}_{1:i-1}}\Big[I_{i-1}^i\Big]\right]. \tag{52}$$

The main result of Theorem 1 can be applied on both the sequential and the consecutive MI values by assigning the notations in a slightly different manner, such that the result of Lemma 2 is transformed into

$${I_0^t}^{in} = \sum_{i=1}^{t}\left[\mathop{\mathbb{E}}_{\mathcal{Z}_{1:i-1}}\left[{I_{i-1}^i}^{in}\right]\right], \tag{53}$$

where ${I_{i-1}^i}^{in} = I_{aug}\left[X_{i-1}^{in} \boxplus x_i; Z_i \mid h_i^-\right]$ is the consecutive MI over the involved subset of the state X_{i-1}.

Theorem 3 *Let us define a new reward, $\rho_t' = \sum_{i=1}^{t+1} {I_{i-1}^i}^{in}$. Solving the ρ-POMDP optimization problem with this reward is equivalent to solving it with the original reward, $\rho_t = IG_0^t$, such that*

$$J_t^\star = \max_{a_t}\left\{\rho_t' + \mathop{\mathbb{E}}_{\mathcal{Z}_{t+1}}\left[J_{t+1}^\star\right]\right\}. \tag{54}$$

Proof

We remind eq. (29) is

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \sum_{t=0}^{T} {I_0^t}^{in} \right\}. \tag{55}$$

Since by definition ${I_0^0}^{in} = 0$, we can start the summation from $t = 1$

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \sum_{t=1}^{T} {I_0^t}^{in} \right\}. \tag{56}$$

Plugging the result from eq. (53) into the above yields

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \sum_{t=1}^{T} \left[\sum_{i=1}^{t} \left[\mathop{\mathbb{E}}_{\mathcal{Z}_{1:i-1}} \left[{I_{i-1}^i}^{in} \right] \right] \right] \right\}. \tag{57}$$

Due to commutativity, we can switch between the order of expectation and summation, which yields

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \mathop{\mathbb{E}}_{\mathcal{Z}_{1:T-1}} \left[\sum_{t=1}^{T} \left[\sum_{i=1}^{t} \left[{I_{i-1}^i}^{in} \right] \right] \right] \right\}. \tag{58}$$

We then denote $\rho'_{t-1} \triangleq \sum_{i=1}^{t} \left[{I_{i-1}^i}^{in} \right]$, and get

$$J_0^\star = \max_{a_{0:T-1}} \left\{ \mathop{\mathbb{E}}_{\mathcal{Z}_{1:T-1}} \left[\sum_{t=1}^{T} \rho'_{t-1} \right] \right\} = \max_{a_{0:T-1}} \left\{ \mathop{\mathbb{E}}_{\mathcal{Z}_{1:T-1}} \left[\sum_{t=0}^{T-1} \rho'_t \right] \right\}. \tag{59}$$

We then separate the first action a_0 from the rest of the actions $a_{1:T-1}$. We also observe that Z_1 is not a function of $a_{1:T-1}$, and that $\rho'_0 = {I_0^1}^{in}$ is not a function of both $a_{1:T-1}$ and Z_1 (since I_0^1 is already an expectation over Z_1). This yields

$$J_0^\star = \max_{a_0} \left\{ \rho'_0 + \mathop{\mathbb{E}}_{\mathcal{Z}_1} \left[\max_{a_{1:T-1}} \left\{ \mathop{\mathbb{E}}_{\mathcal{Z}_{2:T-1}} \left[\sum_{t=1}^{T-1} \rho'_t \right] \right\} \right] \right\}. \tag{60}$$

We then observe that the term inside the expectation over Z_1 is equal to $J_1^\star$, which yields the following recursive form

$$J_0^\star = \max_{a_0} \left\{ \rho'_0 + \mathop{\mathbb{E}}_{\mathcal{Z}_1} \left[J_1^\star \right] \right\}, \tag{61}$$

and, in general, $\forall t \in [1, T-1]$

$$J_t^\star = \max_{a_t} \left\{ \rho'_t + \mathop{\mathbb{E}}_{\mathcal{Z}_{t+1}} \left[J_{t+1}^\star \right] \right\}. \tag{62}$$

We observe that this is the Bellman optimality equation with the new reward, ρ'_t. This eventually means that Solving the ρ-POMDP optimization problem with this reward is equivalent to solving it with the original reward we have started with, $\rho_t = IG_0^t$. We note another slight difference between the formulations, for which the latter formulation does not include a terminal reward.

This allows the usage of estimators which directly estimate MI, as our suggested estimator `MI-SMC` does, together with the usage of tree-based solvers of ρ-POMDP. However, we emphasize that instead of sequential MI values, we will calculate consecutive MI values.

We note that ${I^i_{i-1}}^{in} = \mathop{\mathbb{E}}_{\mathcal{Z}_i}\left[IG^i_{i-1}\right]$. This means that the calculation of the MI values is not limited only to the observations that are used for constructing the tree, thus the calculation can be more accurate, which is another added value of this formulation.

And, lastly, we note that $\rho'_t = \sum_{i=1}^{t+1}\left[{I^i_{i-1}}^{in}\right] = \rho'_{t-1} + {I_t^{t+1}}^{in}$. Meaning that for each node, we can calculate the reward based on the previous reward and just update the new information incrementally, without having to calculate the entire reward from scratch. The belief tree which resembles this optimization problem is shown in Figure 1b.

Using one-time marginalization, i.e. determining ahead all the involved variables (together with variables which are required for other reward functions), and marginalizing out the rest of the variables, the above analysis suggests that the entire tree can be constructed considering only the marginalized beliefs rather than the entire-state beliefs. This, in turn, reduces also the complexity of constructing this tree, since we avoid maintaining and propagating the beliefs over unnecessary states. Care should be taken, however, when using this approach, since marginalizing out a variable which would in retrospect be found to be involved would mean that the tree should be updated from the root. Also note that this approach might prevent the usage of calculation re-use approaches (e.g. [2], [3]) since we only consider a subset of the state for the whole planning process.

References

1. Jan Beirlant, Edward J Dudewicz, László Györfi, Edward C Van der Meulen, et al. Nonparametric entropy estimation: An overview. *International Journal of Mathematical and Statistical Sciences*, 6(1):17–39, 1997.
2. E. I. Farhi and V. Indelman. Towards efficient inference update through planning via jip - joint inference and belief space planning. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2017.
3. E. I. Farhi and V. Indelman. ix-bsp: Belief space planning through incremental expectation. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2019.
4. G. Rotman and V. Indelman. involve-mi: Informative planning with high-dimensional non-parametric beliefs. In *Intl. Workshop on the Algorithmic Foundations of Robotics (WAFR)*, 2022. Submitted.
5. Dan Stowell and Mark D Plumbley. Fast multidimensional entropy estimation by k-d partitioning. *IEEE Signal Processing Letters*, 16(6):537–540, 2009.