Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

An Information Theoretic Approach to Prediction and Deliberative Decision Making of Embodied Systems

by

Nihat Ay, and Keyan Zahedi

Preprint no.: 22 2011

An Information Theoretic Approach to Prediction and Deliberative Decision Making of Embodied Systems

Nihat Ay[1,2] and Keyan Zahedi[1]

[1]Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig Germany
[2]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA
e-mail: {nay,zahedi}@mis.mpg.de

Abstract

This article deals with the causal structure of an agent's sensori-motor loop. Of particular interest are causal effects that can be identified from an agent-centric perspective based on in situ observations. Within this identification, the world model of the agent plays a central role. The prediction quality of an optimal world model is closely related to the notion of predictive information. Its maximization leads to interesting behavioral patterns such as coordination. This is studied in a virtual robotics scenario.

1 Introduction

Evaluating different possibilities and deliberatively choosing among them is an important ability, not only in humans and animals, but also in invertebrates [4]. A plausible model to describe such an ability is based on chaotic attractors as a mechanism to switch between different brain dynamics [5]. In this work, we propose another perspective, which combines embodied artificial intelligence and information theory. The former has demonstrated that cognitive processes are best understood if they are considered in the sensori-motor loop (SML) [8]. The latter allows to formulate first principle models and turned out to be beneficial in the context of self-organized learning [9], [1].

We use the formalism of Bayesian networks to combine both approaches and to study the causal relations in the SML (see previous work [6] and [2] in this context). A Bayesian network consists of two components, a directed acyclic graph G and a set of stochastic maps describing the individual mechanisms of the nodes in the graph. More precisely, G is assumed to have no directed cycles (see Fig. 1 as an example). Given a node Y with state set $\mathscr{Y}$, we write $pa(Y)$ for the set of nodes X that have an edge from X to Y and denote its state set by $\mathscr{X}$. The mechanism of Y is formalized in terms of a stochastic map $\kappa(x;y)$, $x \in \mathscr{X}$, $y \in \mathscr{Y}$. The stochasticity of κ refers to $\sum_y \kappa(x;y) = 1$ for all x.

The Fig. 1 shows the general causal diagram for the SML, where W_t, S_t, C_t, A_t denote the world, sensor, controller (memory), and action at some time t. We denote their state sets by $\mathscr{W}, \mathscr{S}, \mathscr{C}, \mathscr{A}$, respectively. The stochastic maps α, β, φ, and π describe the mechanisms that are involved in the sensori-motor dynamics. Here, φ and π are intrinsic to the agent. They are assumed to be modifiable in terms of a learning process. The mechanisms α and β are extrinsic and encode the agent's embodiment which sets constraints for the agent's learning (for details, see [9], [6]).

Pearl [7] proposes the concept of intervention to capture causal relationships between random variables in a given Bayesian network. We will show that the formalization of the SML allows to determine causal relations solely observational, although its derivation is based on intervention (see Sec. 2). In this identification of causal effects, the optimal world model plays a central role. It is given as the conditional probability $p(s|c,a)$ of observing the next sensor state s as a result of the current controller state c and the current action a of the agent.

In Sec. 3 we then ask the following question: Given an optimal world model (predictor), what is the maximal possible predictive information (PI) that the agent can extract from the SML? This is equivalent to the maximal information that the agent can utilize for prediction, and hence, for deliberative decision-making. Interestingly, only maximizing the PI in the SML already leads to interesting behavior and insights, which are discussed in Sec. 4.

We conclude by summarizing the findings of this paper and discussing their relations to deliberative decision making.

2 Causal effects in the sensori-motor loop

Fig. 1 illustrates the causal structure of the SML. This representation has been used in [6], [2].

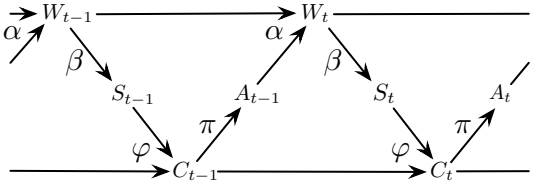

Figure 1: Causal diagram of the sensori-motor loop.

Pearl's formalism [7] allows to define and study causal effects in the SML, for instance the effect of actions on sensor inputs. Here, a fundamental understanding is that in order to reveal causal effects one has to test the system in experimental situations (ex situ). In this context, intervention is an operation that serves as an important building block in corresponding experiments. However, it is not always possible for an agent to perform an intervention. Therefore, it is important to know whether a particular causal effect can be identified purely based on in situ observations of the agent. In the proposition below, we list three causal effects in the SML that are identifiable by the agent without actual intervention. In order to be more precise, we have a closer look at the causal diagram of the transition from time $t-1$

to t.

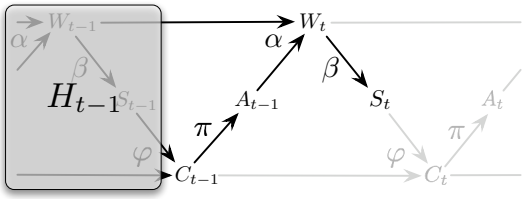

Figure 2: Reduction procedure of the causal diagram.

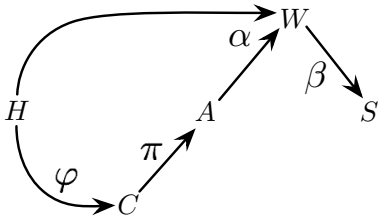

Figure 3: Reduced causal diagram for one time step.

Here, as shown in Fig. 2, we consider the future sensor value of only one time step and summarize the past process by a varible H_{t-1}. We focus on the resulting causal diagram of Fig. 3. The joint distribution in the reduced diagram is given as

$$p(h,c,a,w,s) \;=\; p(h)\,\varphi(h;c)\,\pi(c;a)\,\alpha(h,a;w)\,\beta(w;s)\;. \tag{1}$$

Given such a factorization of the joint distribution, one can define the intervention in a subset, which is referred to as *do*-operation. It is simply given by the corresponding truncation of the product, which formalizes the idea that the mechanisms of the intervened variables are changed from outside. As an illustration, we consider the product (1) and set the value A to a, that is we do a. The result of this conditioning is given as

$$p(h,c,w,s\,|\,do(a)) \;=\; p(h)\,\varphi(h;c)\,\alpha(h,a;w)\,\beta(w;s)\;.$$

Summing over the variables h, c, w, for example, gives us the probability of observing s after having set a. The corresponding stochastic map is referred to as the *causal effect* of A on S:

$$p(s\,|\,do(a)) \;=\; \sum_{h,c,w} p(h,c,w,s\,|\,do(a))\;.$$

Note that, in general, we do *not* have $p(s\,|\,do(a)) = p(s\,|\,a)$, which is an important property of causal effects. Applying the described procedure, one can compute various other causal effects. The following question plays a central role in Pearl's causality theory: Is it possible for an observer, such as an agent considered in this paper, to reveal a causal effect based on observations only? At first sight, this so-called identifiability problem appears meaningless, because causal effects are based on the concept of interventional. However, having some structural information sometimes allows to identify causal effects from observational data.

The following causal effects can be identified by the agent without any actual intervention.

Proposition 1. *Let the joint distribution (1) be strictly positive. Then the following equalities hold:*

$$\textbf{(a)}\;\; p(s\,|\,do(a),c) \;:=\; \frac{p(s,c\,|\,do(a))}{p(c\,|\,do(a))} \;=\; p(s\,|\,c,a)$$

$$\textbf{(b)}\qquad p(s\,|\,do(a)) \;=\; \sum_c p(s\,|\,c,a)\,p(c)$$

$$\textbf{(c)}\qquad p(s\,|\,do(c)) \;=\; \sum_a p(a\,|\,c)\sum_{c'} p(s\,|\,c',a)\,p(c')\;.$$

The proof of Proposition 1 is given in the appendix. In all three causal effects of this proposition, the conditional distribution $p(s\,|\,c,a)$ turns out to be essential as building block for the identification of the causal effects. Note that in the strictly positive case, according to Proposition 1 (a), it is not dependent on the agent's policy. In the next section, this distribution will be studied in more detail.

3 Predictive Information

The causal effects of Proposition 1 involve the conditional distribution $p(s\,|\,c,a)$. In this section we derive an interpretation of this conditional distribution as optimal world model that allows for the best possible prediction. In order to do so, we extend the causal diagram of Figure 3 by a world model γ which assigns a probability of observing s as a result of the action a in the context of the internal state c, formally $\gamma : (\mathscr{C}\times\mathscr{A})\times\mathscr{S} \to [0,1]$. The world model is a model of the agent's expectation, which can be used for a prediction $\widetilde{S}$ of the next sensor input S. We obtain the diagram of Figure 4.

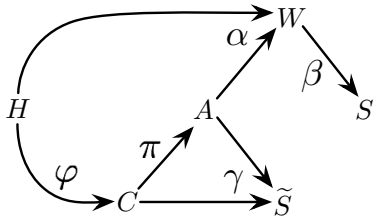

Figure 4: Causal diagram with world model γ.

The distribution of $\widetilde{S}$ given C is derived as

$$\begin{aligned}\widetilde{p}(s|c) \;&:=\; \mathsf{Prob}\{\widetilde{S}=s\,|\,C=c\}\\ &=\; \sum_{a'}\pi(c;a')\,\gamma(c,a';s)\;.\end{aligned}$$

(Here, Prob stands for probability.) In order to measure the quality of the world model γ, we use the entropic distance, also known as KL divergence, between $\widetilde{p}(s|c)$ and $\beta(w;s)$:

$$D(\beta\,\|\,\widetilde{p}) \;:=\; \sum_{c,w} p(c,w)\sum_s \beta(w;s)\ln\frac{\beta(w;s)}{\widetilde{p}(s|c)}\;.$$

The following proposition identifies the conditional probability $p(s|c,a)$ as best world model in terms of

this deviation measure.

Proposition 2. *If a world model $\widehat{\gamma}$ satisfies $\widehat{\gamma}(c,a;s) = p(s|c,a)$ whenever $p(c,a) > 0$ then it minimizes the distance $D(\beta \,\|\, \widetilde{p})$:*

$$\inf_{\gamma} \sum_{c,w} p(c,w) \sum_{s} \beta(w;s) \ln \frac{\beta(w;s)}{\sum_{a'} \pi(c;a')\,\gamma(c,a';s)}$$
$$= \sum_{c,w} p(c,w) \sum_{s} \beta(w;s) \ln \frac{\beta(w;s)}{\sum_{a'} \pi(c;a')\,p(s|c,a')} \; .$$

This implies that the minimal distance coincides with the conditional mutual information $I(W;S\,|\,C)$.

The proof of Proposition 2 is given in the appendix. The optimality of the world model can be achieved in a trivial manner. Assume, for example, that the map β is completely decoupled from the world, that is $\beta(w_1;s) = \beta(w_2;s)$ for all $w_1, w_2 \in \mathscr{W}$. In that case we have complete independence, and also $I(W;S\,|\,C) = 0$. In order to avoid such a decoupling from the world, we use the following estimate

$$I(W;S\,|\,C) \;\leq\; I(W;S).$$

Instead of minimizing the conditional mutual information on the left-hand side, we now consider the maximization of the difference between the upper bound on the right-hand side and the left-hand side:

$$I(W;S) - I(W;S\,|\,C) \; . \tag{2}$$

The first term quantifies the information flow from the world into the sensors. The second term quantifies the dependence of the sensors given the internal state C of the agent. If the second term is low then it is possible to actually predict the sensor state from the internal state. Note that although the two terms of expression (2) explicitly depend on aspects of the world that are not accessible to the agent, the difference is simply the mutual information $I(C;S)$, which only depends on local information that is intrinsic to the agent. This follows from the chain rule for mutual information and the conditional independence structure of the SML:

$$\begin{aligned} I(W;S) &= I(W,C;S) \\ &= I(C;S) + I(W;S|C) \; . \end{aligned}$$

In our previous work we studied the maximization of this mutual information which we refer to as *predictive information*, a notion that is related to the work [3]. Its maximization has some important implications in terms of behavioral patterns of embodied agents. We report on our results in the next section.

4 Prediction and Embodiment

The previous sections discussed the PI as the upper bound for the information available to the agent for prediction, and as a consequence, for task-orientated behavior. Given an optimal world model $\gamma(c,a;s)$, the maximal achievable PI is determined by the policy $\pi(c;s)$ and the kernels α and β. The latter two represent the embodiment of the agent, i.e. its behavior-relevant morphological properties. It is assumed here, that these are fixed, i.e. not open to variation by the agent itself. Hence, the PI can only be maximized with respect to the given embodiment by adaptation of the policy π. Consequently, an analysis and understanding of such information maximization principles in this context cannot be done solely based on the causal diagram as given in Fig. 3, but requires an realization in the SML. For this purpose we utilize virtual robots and determine the quality of the information maximization process also based on the observable behavior of the embodied robotic systems.

In previous experiments [9], a learning rule, which maximizes the PI was implemented in a chain of passively coupled, individual controlled mobile robots, placed in a bounded, but otherwise featureless environment (see Fig. 5). In the experiment chosen for presentation here, each wheel in a chain of five robots was controlled by a single controller. The only input and output to one controller is the current wheel velocity S_t, and the desired velocity A_t. No information about the actions and number of the other controller is available, other than through sensor stream S_t. Nevertheless, locally maximizing the PI leads to observable coordination among the robots (see Fig. 6). This not only demonstrates the potential of the PI, but also shows the necessity of applying concepts to embodied systems, as the observable behavior is not only a result of the PI maximization and but also the result of the exploitation of the embodiment by the policy, which is closely related to the concept of morphological computation [9].

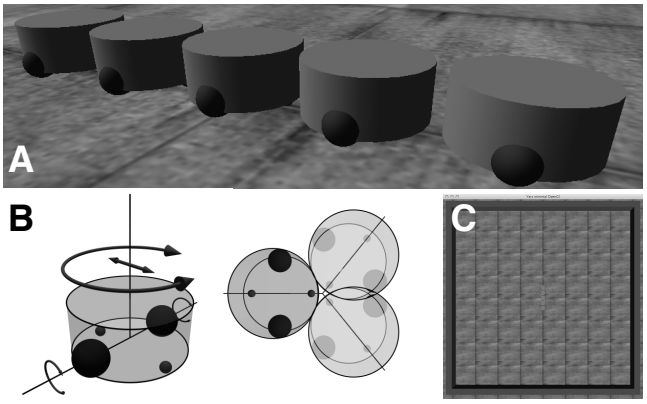

Figure 5: Experimental Set-up. A) Robot chain with 5 Robots. B) Schematics of the two-wheeled, differential drive robot, and the configuration of the passive connection (hinge joint with a deviation of $\pm 50°$). C) Bounded, but otherwise featureless environment, in which the robot was placed for learning.

5 Conclusions

Pearl writes in his book ([7], page 108): *Actions admit two interpretations: reactive and deliberative. The reactive interpretation sees action as a consequence of an agent's beliefs, disposition, and environmental inputs, as in "Adam ate the apple because Eve handed it to him." The deliberative interpretation sees action as an option of choice in contemplated decision making, usually involving comparison of consequences, as in "Adam was wandering what God would do if he ate the apple."*

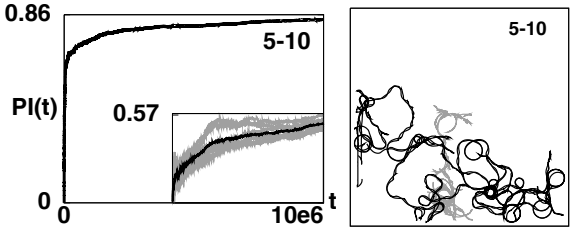

Figure 6: Results. Left-hand side: The maximization of the PI (avg. over all controllers) plotted over time for the robot chain with 5 robots and 10 controllers. The smaller box shows the initial learning phase, where the gray lines show the evolution of the individual controllers. Right-hand side: Trajectory plot for the same robot configuration. In gray: the initial learning phase, black: the converged exploration behavior as a result of the PI-maximization.

We have used a policy which refers to the Pearl's reactive interpretation of actions. It assigns an action a to the agent's state c, formalized in terms of a stochastic mapping $\pi : \mathscr{C} \times \mathscr{A} \to [0,1]$, $(c,a) \mapsto \pi(c;a)$. A deliberative policy would assign an action a to a given internal state c and an intended sensor state s, that is $\overline{\pi} : \big(\mathscr{C} \times \mathscr{S}\big) \times \mathscr{A}$, $(c,s,a) \mapsto \overline{\pi}(c,s;a)$. Here, the world model, which we considered in this paper, allows to choose actions that maximize the probability of a sensor state s. More precisely, if the system intends to generate a state s, given that the internal state is c, it would choose an action a^* with $p(s|c,a^*) = \max_a p(s|c,a)$. Such deliberative decision making is based on the optimal prediction of consequences of actions.

Acknowledgements

Both authors thank Ralf Der, Daniel Polani, and Bastian Steudel for valuable discussions on causal effects in the sensori-motor loop. This work has been supported by the Santa Fe Institute.

References

[1] Nihat Ay, Nils Bertschinger, Ralf Der, Frank Güttler, and Eckehard Olbrich. Predictive information and explorative behavior of autonomous robots. *The European Physical Journal B - Condensed Matter and Complex Systems*, 63(3):329–339, 2008.

[2] Nihat Ay and Daniel Polani. Information flows in causal networks. *Advances in Complex Systems*, 1(11):17–41, 2008.

[3] William Bialek, Ilya Nemenman, and Naftali Tishby. Predictability, complexity, and learning. *Neural Computation*, 13(11):2409–2463, 2001.

[4] Björn Brembs. Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates. *Proceedings of the Royal Society B: Biological Sciences*, 2011.

[5] Kunihiko Kaneko and Ichiro Tsuda. *Complex Systems: Chaos and Beyond*. Springer Berlin / Heidelberg, 2001.

[6] Alexander S. Klyubin, Daniel Polani, and Chrystopher L. Nehaniv. Tracking information flow through the environment: Simple cases of stigmergy. *Proceedings of Artificial Life IX*, 2004.

[7] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[8] Rolf Pfeifer and Josh C. Bongard. *How the Body Shapes the Way We Think: A New View of Intelligence*. The MIT Press (Bradford Books), 2006.

[9] Keyan Zahedi, Nihat Ay, and Ralf Der. Higher coordination with less control – a result of information maximization in the sensori-motor loop. *Adaptive Behavior*, 18(3–4):338–355, 2010.

Appendix

Proof of Proposition 1:

(a) $$\begin{aligned} & p(h,c,w,s \,|\, do(a)) \\ &= p(h)\,\varphi(h;c)\,\alpha(h,a;w)\,\beta(w;s) \end{aligned}$$

This implies

$$\begin{aligned} & p(s,c \,|\, do(a)) \\ &= \sum_{h,w} p(h)\,\varphi(h;c)\,\alpha(h,a;w)\,\beta(w;s) \end{aligned}$$

$$\begin{aligned} & p(c \,|\, do(a)) \\ &= \sum_{s}\sum_{h,w} p(h)\,\varphi(h;c)\,\alpha(h,a;w)\,\beta(w;s) \\ &= p(c) \end{aligned}$$

$$\begin{aligned} & p(s \,|\, do(a),c) \\ &= \frac{p(s,c \,|\, do(a))}{p(c \,|\, do(a))} \\ &= \sum_{h,w} \tfrac{p(h)}{p(c)}\,\varphi(h;c)\,\alpha(h,a;w)\,\beta(w;s) \\ &= \sum_{h,w} p(h \,|\, c)\,p(w \,|\, h,a)\,p(s \,|\, w) \\ &= \sum_{h,w} p(h \,|\, c,a)\,p(w \,|\, h,a,c)\,p(s \,|\, w,h,a,c) \\ &\quad \text{(conditional independence,} \\ &\quad \text{see diagram in Figure 3)} \\ &= p(s \,|\, a,c)\ . \end{aligned}$$

The second and third equations of the proposition follow from the general theory (see [7], Theorem 3.2.2 (Adjustment for Direct Causes, and Theorem 3.3.4 (Front-Door Adjustment)). For completeness, we prove them directly.

(b) $$\begin{aligned} & p(s \,|\, do(a)) \\ &= \sum_{h,c,w} p(h,c,w,s \,|\, do(a)) \\ &= \sum_{h,c,w} p(h)\,\varphi(h;c)\,\pi(c;a)\,\alpha(h,a;w)\,\beta(w;s)\frac{1}{p(a|c)} \\ &= \sum_{h,c,w} \frac{p(h,c,a,w,s)}{p(c,a)}\,p(c) \\ &= \sum_{c} p(s|c,a)\,p(c)\ . \end{aligned}$$

(c) $p(s \,|\, do(c))$

$$
\begin{aligned}
&= \sum_{h,a,w} p(h,a,w,s \,|\, do(c)) \\
&= \sum_{a} \pi(c;a) \sum_{h,w} p(h)\, \alpha(h,a;w)\, \beta(w;s) \\
&= \sum_{a} p(a|c) \sum_{h,w} \left(\sum_{c'} p(c')\, p(h|c') \right) p(w|h,a)\, p(s|w) \\
&= \sum_{a} p(a|c) \sum_{c'} p(c') \sum_{h,w} p(h|c')\, p(w|h,a)\, p(s|w) \\
&= \sum_{a} p(a|c) \sum_{c'} p(c') \sum_{h,w} p(h|c',a)\, p(w|h,a,c')\, p(s|w) \\
&= \sum_{a} p(a|c) \sum_{c'} p(c')\, p(s|c',a)\ . \qquad \square
\end{aligned}
$$

Proof of Proposition 2: We first varify the relation to conditional mutual information in the case where $\widehat{\gamma}(c,a;s) := p(s \,|\, c,a)$ whenever $p(c,a) > 0$:

$$
\begin{aligned}
&\sum_{c,w} p(c,w) \sum_{s} \beta(w;s) \ln \frac{\beta(w;s)}{\sum_{a'} \pi(c;a')\, p(s|c,a')} \\
&= \sum_{c,w} p(c,w) \sum_{s} p(s|w) \ln \frac{p(s|w)}{p(s|c)} \\
&= \sum_{c,w} p(c,w) \sum_{s} p(s|c,w) \ln \frac{p(s|c,w)}{p(s|c)} \\
&\quad \text{(conditional independence,} \\
&\quad \text{see diagram in Figure 3)} \\
&= I(W;S \,|\, C)
\end{aligned}
$$

It remains to prove that the above choice $\widehat{\gamma}$ is indeed a minimizer. To this end, we introduce Lagrange multiplier $\lambda_{c,a}$ and consider the following partial derivative (here we use $\gamma(c,a;s)$, $c \in \mathscr{C}$, $a \in \mathscr{A}$, $s \in \mathscr{S}$, as independent variables):

$$
\begin{aligned}
&\frac{\partial}{\partial\, \gamma(\bar{c},\bar{a};\bar{s})} \Bigg\{ \sum_{c,w} p(c,w) \sum_{s} \beta(w;s) \\
&\qquad\qquad \times \ln \frac{\beta(w;s)}{\sum_{a'} \pi(c;a')\, \gamma(c,a';s)} \\
&\qquad\qquad + \sum_{c,a} \lambda_{c,a} \left(\sum_{s} \gamma(c,a;s) - 1 \right) \Bigg\} \\
&= -p(\bar{c},\bar{s}) \, \frac{\pi(\bar{c};\bar{a})}{\sum_{a'} \pi(\bar{c};a')\, \gamma(\bar{c},a';\bar{s})} \ + \ \lambda_{\bar{c},\bar{a}}.
\end{aligned}
$$

The partial derivatives vanish if

$$
\lambda_{\bar{c},\bar{a}} \ = \ p(\bar{c},\bar{s}) \, \frac{\pi(\bar{c};\bar{a})}{\sum_{a'} \pi(\bar{c};a')\, \gamma(\bar{c},a';\bar{s})} \ .
$$

The constraints imply $\lambda_{\bar{c},\bar{a}} \ = \ p(\bar{c})\, \pi(\bar{c};\bar{a})$, and therefore

$$
\sum_{a'} \pi(\bar{c};a')\, x_{\bar{c},a';\bar{s}} \ = \ p(\bar{s} \,|\, \bar{c}) \ . \tag{3}
$$

If we choose

$$
\gamma(c,a;s) \ := \ p(s \,|\, c,a)
$$

whenever $p(c,a) > 0$ then the equation (3) is satisfied. The convexity of the function $D(\beta \,\|\, \widetilde{p})$ in γ implies that $\widehat{\gamma}$ is a minimizer. $\square$