*Int. J. Appl. Math. Comput. Sci., 2020, Vol. 30, No. 1, 23–34*
DOI: 10.34768/amcs-2020-0002

# AN INFORMATION BASED APPROACH TO STOCHASTIC CONTROL PROBLEMS

PIOTR BANIA [a]

[a]Faculty of Automatic Control and Robotics
AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Cracow, Poland
e-mail: pba@agh.edu.pl

An information based method for solving stochastic control problems with partial observation is proposed. First, information-theoretic lower bounds of the cost function are analysed. It is shown, under rather weak assumptions, that reduction in the expected cost with closed-loop control compared with the best open-loop strategy is upper bounded by a non-decreasing function of mutual information between control variables and the state trajectory. On the basis of this result, an *information based control* (IBC) method is developed. The main idea of IBC consists in replacing the original control task by a sequence of control problems that are relatively easy to solve and such that information about the system state is actively generated. Two examples of the IBC operation are given. It is shown that the method is able to find an optimal solution without using dynamic programming at least in these examples. Hence the computational complexity of IBC is substantially smaller than that of dynamic programming, which is the main advantage of the proposed method.

**Keywords:** stochastic control, feedback, information, entropy.

## 1. Introduction

Optimal controller synthesis in stochastic systems with partial observation can be performed by dynamic programming (DP). Unfortunately, although the theory of DP is well developed (see Zabczyk, 1996), its computational complexity grows exponentially with the number of variables and time steps. As a consequence, the problem is practically intractable.

To overcome the curse of dimensionality, a number of approximate methods have been developed. The *separation principle* and the *certainty equivalence* assumption have been used by Filatov and Unbehauen (2004), Åström and Wittenmark (1995), Tse (1974) or BarShalom and Tse (1976). As part of the theory of partially-observable Markov decision processes (POMDPs), various *policy-iteration* or *value-iteration* methods were developed by Thrun (2000), Porta *et al.* (2006), Brechtel *et al.* (2013), Dolgov (2017), Zhao *et al.* (2019), and many other researchers. These methods were initially developed for systems with a finite number of states and then adopted to more general problems with smooth dynamics. Therefore, as the numbers of variables and time steps increase, they suffer from the curse of dimensionality. Thus, there is still a need to develop methods of smaller computational complexity.

Analysis of the known optimal solutions (Zabczyk, 1996; Filatov and Unbehauen, 2004; Åström and Wittenmark, 1995; Tse, 1974; BarShalom and Tse, 1976; Bania, 2017) suggests that active exchange of information between the controller and the system is a distinctive feature of optimal controllers (Bania, 2018). Relationships between control of dynamical systems and available information are fundamental for understanding stochastic control theory. Since the pioneering work of Feldbaum (1965) the connections between control and information theory have been intensively studied. Hijab (1984) showed that the concept of entropy appears naturally in dual control. The entropic formulation of stochastic control was given by Saridis (1988) and Tsai *et al.* (1992). The works of Banek (2010) as well as Kozlowski and Banek (2011) suggest that information exchange, entropy reduction and stochastic optimality are related to one another.

An information and entropy flow in control systems was analyzed in the papers of Mitter and Newton (2005) as well as Sagawa and Ueda (2013). The

controllability, observability and stability of linear control systems with limitations of information contained in the measurements were investigated by Taticonda and Mitter (2004). Touchette and Lloyd (2004) showed that controllability and observability can be defined using the concepts of information theory. One of the most relevant results related to the subject of this article is the inequality of Touchette and Lloyd (2000). They proved that the one-step reduction in entropy of the final state is upper bounded by the mutual information between the control variables and the current state of the system. Delvenne and Sandberg (2013) suggested how to extend this result to more general cost functions.

The main contribution of this paper is as follows. First, the open and closed-loop strategies are defined in terms of mutual information between the system trajectory and control variables. Next, it is proved, under relatively weak assumptions, that

$$J_{\text{open}} - J_{\text{closed}}(\varphi) \leq \rho(I(X;U|\varphi)),$$

where $J_{\text{open}}$ is the expectation of the cost corresponding to the best open-loop control, $J_{\text{closed}}$ is the expectation of the cost corresponding to any closed-loop strategy $\varphi$, and $I(X;U|\varphi)$ is the mutual information between the system trajectory and control variables under the strategy $\varphi$. Function $\rho$ is non-decreasing and $\rho(0) = 0$.

Additionally, we prove that, under slightly stronger assumptions, $\rho$ is bounded by a linear function. Hence the condition $I(X;U) > 0$ is necessary for reduction in the cost below the best open-loop cost. On the basis of the above inequality, *information based control* (IBC) is proposed for finding an approximate solution of stochastic control problems. The phrase "approximate solution" means that the proposed method is able to find a strategy no worse than the *open-loop feedback optimal* (OLFO) algorithm given by Tse (1974).

The main idea consists in replacing the original control task by a sequence of control problems that are relatively easy to solve and such that the condition $I(X;U) > 0$ can be fulfilled. This can be done by introducing a penalty function for information deficiency. As a penalty function, the predicted mutual information between the system trajectory and the measurements is used. A similar idea was proposed by Alpcan *et al.* (2015), however, in this article, the process noise (input disturbances) is completely ignored, which is a very strong and often unrealistic assumption. Additional contributions include sufficient conditions for the above-mentioned bound, a one-step information-theoretic bound for the quadratic cost and two examples of the operation of IBC. In both examples, the optimal solution is found analytically by DP and then compared with the IBC solution. It is shown that IBC is able to find an optimal solution without using DP, which is the main advantage of the proposed method.

The rest of the paper is organized as follows. Section 2 formulates the stochastic control problem. Information-theoretic lower bounds of the cost are given in Section 3. Section 4 presents IBC and Section 5 contains examples of its application. A Monte Carlo approximation of the cost function and some computational issues are discussed in Section 6. The paper ends with conclusions and a list of references.

**Notation.** The abbreviation $\xi \sim p_\xi$ means that the variable $\xi$ has a density $p_\xi(\xi)$. The notation $\xi \sim N(m,S)$ means that $\xi$ has normal distribution with mean $m$ and covariance matrix $S$. If $S > 0$, then the density of normally distributed variable is

$$N(x,m,S) = (2\pi)^{-\frac{n}{2}}|S|^{-\frac{1}{2}}\exp(-0.5(x-m)^T S^{-1}(x-m)).$$

The symbol $\mathrm{col}(a_1,a_2,\ldots,a_n)$ denotes a column vector. The trace of matrix $A$ is denoted by $\mathrm{tr}(A)$. The inner product of matrices $A$ and $B$ is defined as $\langle A,B\rangle = \mathrm{tr}(A^TB)$. Let $\xi \in \mathbb{R}^n$ and let $Q$ be a square matrix of order $n$. The quadratic form $\xi^TQ\xi$ is denoted by $|\xi|_Q^2$. The entropy of variable $\xi$ is denoted by $H(\xi)$. The control strategy is denoted by $\varphi$. The symbol $H(\xi|\varphi)$ means that the entropy of variable $\xi$ is calculated with fixed strategy $\varphi$.

## 2. Stochastic control task

Consider the following stochastic system:

$$x_{k+1} = f(x_k,u_k,w_k), \quad k = 0,1,\ldots \tag{1}$$

$$y_k = h(x_k,v_k), \tag{2}$$

$$u_k \in U_{\text{ad}} = \{u \in \mathbb{R}^r : u_{\min} \leq u \leq u_{\max}\}, \tag{3}$$

where $x_k \in \mathbb{R}^n$, $y_k \in \mathbb{R}^m$, $w_k \in \mathbb{R}^{n_w}$, $v_k \in \mathbb{R}^{n_v}$, $w_k \sim p_w$, $v_k \sim p_v$. The inequalities in (3) are elementwise. It is also possible, in some justified cases, that $U_{\text{ad}} = \mathbb{R}^r$. Functions $f,h$ are $C^2$ with respect to all their arguments. The initial distribution of $x_0$ is denoted by $p_0^-(x_0)$. Variables $x_0,w_0,w_1,\ldots,w_k,v_0,v_1,\ldots,v_k$ are mutually independent for all $k$. Measurements until time $k$ are denoted by $Y_k = \mathrm{col}(y_0,y_1,\ldots,y_k) \in \mathbb{R}^{m(k+1)}$. Similarly, $X_k = \mathrm{col}(x_0,x_1,\ldots,x_k) \in \mathbb{R}^{n(k+1)}$, $U_k = \mathrm{col}(u_0,u_1,\ldots,u_k) \in \mathbb{R}^{r(k+1)}$. The control horizon is denoted by $N \geq 1$. We also introduce the following abbreviations: $Y = Y_{N-1}, U = U_{N-1}, X = X_{N-1}$.

Let $B(\mathbb{R}^{Nm},\mathbb{R}^{Nr})$ be the set of all bounded maps from $\mathbb{R}^{Nm}$ into $\mathbb{R}^{Nr}$. If $f_1,f_2 \in B$, $\alpha,\beta \in \mathbb{R}$ then $\alpha f_1 + \beta f_2 \in B$. Hence $B$ is a linear space. The set $B$ with the norm $\|f\|_{\mathbf{B}} = \sup_{Y\in\mathbb{R}^{Nm}}\|f(Y)\|_{\mathbb{R}^{Nr}}$, is a Banach space, which will be denoted by $\mathbf{B}$ and called the strategy space. The measurable map

$$\varphi_k : \mathbb{R}^{m(k+1)} \to U_{\text{ad}}, \quad u_k = \varphi_k(Y_k) \tag{4}$$

is the control strategy at time $k$. Let $U_{\text{ad}}^N = (U_{\text{ad}} \times U_{\text{ad}} \times, \ldots, \times U_{\text{ad}})_{N \text{ times}}$. The map

$$\varphi : \mathbb{R}^{mN} \to U_{\text{ad}}^N \subset \mathbb{R}^{Nr}, \quad U = \varphi(Y), \tag{5}$$

where

$$\varphi(Y) = \text{col}(\varphi_0(Y_0), \ldots, \varphi_{N-1}(Y_{N-1})), \tag{6}$$

is an admissible control strategy. The set of all admissible strategies is a denoted by $S_{\text{ad}}$. It follows from (4)–(6) that $S_{\text{ad}}$ is a bounded, closed and convex subset of $\mathbf{B}$.

Let $L : \mathbb{R}^n \to \mathbb{R}$ be a measurable $C^2$ function and let $J : S_{\text{ad}} \to R$ denote the cost functional. We are looking for a strategy $\varphi \in S_{\text{ad}}$ that minimizes the functional

$$J(\varphi) = E\{L(x_N)|\varphi\}, \tag{7}$$

where the expectation is calculated with respect to $x_0$, $w_0, \ldots, w_{N-1}$, $v_0, \ldots, v_{N-1}$. The optimal strategy will be denoted by $\varphi^*$ and the abbreviation $J(\varphi^*) = J^*$ will be used. We will assume that $\varphi^*$ exists. The optimal control corresponding to a realization of $Y_k$ will be denoted by $u_k^* = \varphi_k^*(Y_k)$.

## 3. Information-theoretic lower bounds of the cost function

If the strategy $\varphi \in S_{\text{ad}}$ is fixed, then relations between random variables $X, Y, U$ are described by their joint density $p(X, Y, U|\varphi)$. In particular, if $p(X, U|\varphi) = p(X|\varphi)p(U|\varphi)$, then $X$ and $U$ are independent and information contained in measurements $Y$ is not utilized. This is an open loop control strategy. A reduction in the cost (7), compared with the open-loop, is possible only if $X$ and $U$ are dependent. A natural measure of dependency is mutual information. We will show below that the cost (7) is lower-bounded by some non-increasing function of the mutual information between $X$ and $U$.

**3.1. General bounds.** The mutual information between $X$ and $U$ is given by

$$I(\varphi) = H(X|\varphi) - H(X|U, \varphi), \tag{8}$$

where the entropies $H(X|\varphi)$, $H(X|U, \varphi)$ are defined in a usual way, i.e.,

$$H(X|\varphi) = E(-\ln p(X|\varphi)), \tag{9}$$
$$H(X|U, \varphi) = E(-\ln p(X|U, \varphi)). \tag{10}$$

The expected value in (10) is calculated with respect to $X$ and $U$.

**Definition 1.** The strategy $\varphi$ is an open-loop strategy if, and only if, $I(\varphi) = 0$. Otherwise, $\varphi$ will be called a closed-loop or feedback strategy.

Let $s \in \mathbb{R}, s \geq 0$. The set

$$\Omega(s) = \{\varphi \in S_{\text{ad}} : I(\varphi) \leq s\} \tag{11}$$

contains all strategies for which the information $I(\varphi)$ is no greater than $s$. Let $\varphi \in S_{\text{ad}}$ be a constant map. Since $\varphi$ is constant, $U$ and $Y$ are independent and $I(\varphi) = 0$. Hence $\Omega(s)$ is non-empty for all $s \geq 0$. Consider now a family of optimization problems

$$\inf_{\varphi \in \Omega(s)} J(\varphi). \tag{12}$$

An optimal solution of (12) will be denoted by $\varphi_s^*$ and it is assumed that $\varphi_s^*$ exists for all $s$. The minimum open-loop cost is defined as

$$J_o = \inf_{\varphi \in \Omega(0)} J(\varphi). \tag{13}$$

**Lemma 1.** *If the solution to (12) exists for all $s \geq 0$, then there exists a non-decreasing, bounded function $\rho : [0, \infty) \to [0, J_o - J^*]$, $\rho(0) = 0$, such that*

$$J_o - J(\varphi) \leq \rho(I(\varphi)) \tag{14}$$

*for all $\varphi \in S_{ad}$.*

*Proof.* Define

$$\rho(s) = \sup_{\varphi \in \Omega(s)} (J_o - J(\varphi)). \tag{15}$$

For every $t, s \geq 0$, we have $\Omega(s) \subset \Omega(s + t)$. Hence $\rho$ is non-decreasing. If $s = 0$, then by (13) we have

$$\rho(0) = \sup_{\varphi \in \Omega(0)} (J_o - J(\varphi)) = J_o - J_o = 0.$$

Since $\varphi \in \Omega(I(\varphi))$,

$$J_o - J(\varphi) \leq \sup_{\psi \in \Omega(I(\varphi))} (J_o - J(\psi)) = \rho(I(\varphi)), \tag{16}$$

which proves (14). ∎

It follows from (14) that $J(\varphi) < J_o \Rightarrow I(\varphi) > 0$, but the function $\rho$ in (14) can be very irregular. To obtain a more accurate bound, additional conditions are needed. Let

$$d(\Omega(0), \varphi) = \inf_{\psi \in \Omega(0)} ||\psi - \varphi|| \tag{17}$$

denote the distance between $\Omega(0)$ and $\varphi$.

**Theorem 1.** *If there exist numbers $L_I, L_J > 0$, such that*

$$|J(\varphi) - J(\varphi_1)| \leq L_J ||\varphi_1 - \varphi||, \quad \varphi, \varphi_1 \in S_{ad}, \tag{18}$$

$$I(\varphi) \geq L_I d(\Omega(0), \varphi), \quad \varphi \in S_{ad}, \tag{19}$$

*then there exists a number $q > 0$ such that*

$$J_o - J(\varphi) \leq q I(\varphi), \quad \varphi \in S_{ad}. \tag{20}$$

*Proof.* Let $\varphi \notin \Omega(0)$ and let $\varphi_1 \in \Omega(0)$ be such that $||\varphi_1 - \varphi|| = d(\Omega(0), \varphi)$. If $qL_I - L_J \geq 0$, then on the basis of (18), (19) and (13) we get

$$\begin{aligned} &J(\varphi) + qI(\varphi) \\ &\quad \geq J(\varphi_1) - L_J||\varphi_1 - \varphi|| + qL_I d(\Omega(0), \varphi) \\ &\quad = J(\varphi_1) + (qL_I - L_J)||\varphi_1 - \varphi|| \\ &\quad \geq J(\varphi_1) \geq J_o, \quad \varphi \notin \Omega(0). \end{aligned} \tag{21}$$

If $\varphi \in \Omega(0)$, then $I(\varphi) = 0$ and it follows from (13) that $J(\varphi) \geq J_o$. Hence (20) holds for all $\varphi \in S_{\text{ad}}$. ■

**Remark 1.** The data processing inequality (see Cover and Thomas, 2006, p. 34) says that $I(X; F(Y)) \leq I(X; Y)$, for any function $F$. Since $U = \varphi(Y)$,

$$I(\varphi) = I(X; U|\varphi) \leq I(X; Y|\varphi). \tag{22}$$

As a consequence, Lemma 1 and Theorem 1 will still be true if we use $I(X; Y|\varphi)$ instead of $I(\varphi)$.

Since $S_{\text{ad}}$ is bounded and closed, then the Lipschitz continuity assumption (18) is not very restrictive. The assumption (19) says that information must grow linearly with the distance from the set $\Omega(0)$, which seems quite natural and not very restrictive. Let us also note that $I(\varphi)$ need not to be continuous.

**3.2. Entropy reduction of the final state.** Assume that the cost functional has the form

$$J(\varphi) = E\{-\ln p(x_N|\varphi)\}. \tag{23}$$

We shall call $J(\varphi)$ the closed-loop entropy and will write $H(\varphi) = J(\varphi)$. The minimum open-loop entropy of the final state is denoted by $H_o = J(\varphi_0^*)$. Touchette and Lloyd (2000; 2004) showed that one-step (i.e., $N = 1$) entropy reduction as compared to the best open-loop strategy is upper bounded by $I(x_0; u_0|\varphi)$. Their inequality (in our notation) has the form

$$H_o - H(\varphi) \leq I(\varphi), \quad \varphi \in S_{\text{ad}}. \tag{24}$$

It is a fundamental limitation in control systems but, unfortunately, the multi-step ($N > 1$) version of (24) is very weak (cf. Touchette, 2000, p. 47, Eqn. (3.74)). It only says that there *exists* a strategy $\varphi$ such that

$$H_o - H(\varphi) \leq \sum_{k=0}^{N-1} I(x_k; u_k|\varphi). \tag{25}$$

Since correlations between previous measurements and current control are omitted in (25), it may not be fulfilled for some $\varphi$. However, it is still possible on the basis of (24) to construct some one-step bound for (7).

**Theorem 2.** *Let $J(\varphi) = E\{L(x_1)|\varphi\}$, $x_1 \in \mathbb{R}^n$. If $L(x_1) \geq c|x_1|^2$, $c > 0$ then*

$$J(\varphi) \geq cn(2\pi e)^{-1} e^{2n^{-1}(H_o - I(\varphi))}. \tag{26}$$

*Proof.* Let us fix $S = \text{cov}(x_1, x_1|\varphi)$. Matrix $S$ fulfils the inequality $\text{tr}(S) \geq n|S|^{\frac{1}{n}}$ (cf. Cover and Thomas, 2006, Thm. 17.9.4, p. 680). Since the Gaussian distribution maximizes entropy over all distributions with the same covariance, it can be proved that $|S| \geq (2\pi e)^{-n} e^{2H(\varphi)}$ (Cover and Thomas, 2006, Thm. 8.6.5, p. 254). On the basis of these two inequalities and by using (24), we obtain

$$\begin{aligned} J(\varphi) &\geq cE|x_1|^2 \geq c\,\text{tr}(S) \geq cn|S|^{\frac{1}{n}} \\ &\geq cn(2\pi e)^{-1} e^{2n^{-1}H(\varphi)} \\ &\geq cn(2\pi e)^{-1} e^{2n^{-1}(H_o - I(\varphi))}. \end{aligned}$$

■

**3.3. Elementary example.** To illustrate the problem, consider a one-dimensional system

$$x_1 = x + u, \quad y = x + v. \tag{27}$$

Variables $x$ and $v$ are Gaussian, i.e., $x \sim N(0, s_x), s_x > 0$, $v \sim N(0, s_v), s_v > 0$. The cost functional has the form

$$J(\varphi) = E\{x_1^2|\varphi\}. \tag{28}$$

The best open-loop strategy is $\varphi_0^* = 0$ and $J_o = s_x$. The optimal strategy is given by a linear function of $y$,

$$\varphi^*(y) = -\frac{s_x}{s_x + s_v} y, \tag{29}$$

and the minimum cost is equal to

$$J(\varphi^*) = \frac{s_x s_v}{s_x + s_v} < s_x = J_o. \tag{30}$$

Since $x_1$ is Gaussian, its open-loop entropy is given by $H_o = \frac{1}{2}\ln(2\pi e J_o)$ and the inequality (26) yields

$$J(\varphi) \geq J_o e^{-2I(\varphi)} \tag{31}$$

for all $\varphi$. One can check by direct calculation that

$$I(\varphi^*) = \tfrac{1}{2}\ln\left(1 + \frac{s_x}{s_v}\right), \tag{32}$$

and then $J(\varphi^*) = J_o e^{-2I(\varphi^*)}$. Hence the bound (31) is tight. The entropy of $x_1$, under the optimal strategy, is given by $H(\varphi^*) = \frac{1}{2}\ln(2\pi e J(\varphi^*))$, and one can check that $H_o - H(\varphi^*) = I(\varphi^*)$. Hence, the strategy (29) is also optimal for entropy reduction.

## 4. Information based control

Minimum of $J(\varphi)$ can be found by dynamic programming (DP), but the computational complexity of DP grows exponentially with the number of time steps and control

variables. As a consequence, DP is often impractical and there is a need to construct approximate methods with a lower computational complexity (Filatov and Unbehauen, 2004, pp. 14–32; Åström and Wittenmark, 1995, pp. 354–370). It is possible, on the basis of the previous section, to construct such an approximate method. The easiest way to simplify the problem is to replace the original control task with a sequence of open-loop control problems. These control problems consist in minimization of

$$J_k(u^{(k)}) = E\{L(x_N)|Y_k, u^{(k)}\}, \tag{33}$$

where $u^{(k)} = \mathrm{col}(u_k, \ldots, u_{N-1})$ denote the future control sequence.

The minimizer of (33) will be denoted by $\bar{u}^{(k)}(Y_k)$. To control the system, only the first element of $\bar{u}^{(k)}$ is used and the procedure is repeated in subsequent steps. Hence, the control strategy generated by sequential minimization of (33) has the form

$$\varphi_k(Y_k) = \bar{u}_1^{(k)}(Y_k), \tag{34}$$

and this may or may not be feedback in the sense Definition 1.

The above simplification is known as *open loop feedback optimal* (OLFO), and it is well known that does not generate information and cannot be optimal, except linear Gaussian systems (cf. Section 3.3; Example 2; Tse, 1974; Filatov and Unbehauen, 2004). On the other hand, it follows from Section 3, and particularly from (20) and (22), that

$$J(\varphi) \geq J_o - qI(X;Y|\varphi), \tag{35}$$

which implies that every controller better than the open-loop one must actively generate information. This can be enforced by adding to (33) a penalty function for information deficiency. Such a penalty function can be constructed by using the mutual information between future states and measurements. It is also possible to use $I(X;U)$ as a penalty; however, calculation of $I(X;U)$ is much more difficult than that of $I(X;Y)$. Therefore it is computationally more convenient to use $I(X;Y)$. This is the basic idea of *information based control* (IBC).

A practically realizable implementation of IBC is as follows. Let $X_k^+ = \mathrm{col}(x_{k+1}, \ldots, x_{N-1})$, $Y_k^+ = \mathrm{col}(y_{k+1}, \ldots, y_{N-1})$ denote the future states and observations. Define, for $k = 0, 1, \ldots, N-2$,

$$I_k(u^{(k)}|Y_k) = \int p(X_k^+, Y_k^+|Y_k) \ln \frac{p(X_k^+, Y_k^+|Y_k)}{p(X_k^+|Y_k)p(Y_k^+|Y_k)} \, \mathrm{d}X_k^+ \mathrm{d}Y_k^+. \tag{36}$$

This is the mutual information between $X_k^+$ and $Y_k^+$, predicted at time $k$ and conditioned on $Y_k$. Since $y_N$ is irrelevant from the control point of view, one can assume that $I_{N-1} = 0$. Now, at every time instant, we are looking for the minimum of the functional

$$J_k(u^{(k)}) = E\{L(x_N)|Y_k\} - \nu_k I_k(u^{(k)}|Y_k), \tag{37}$$

where

$$u_i^{(k)} \in U_{\mathrm{ad}}, \quad \nu_k \geq 0, \quad k = 0, \ldots, N-1, \quad N \geq 2. \tag{38}$$

The expectation in (37) is calculated with respect to $x_0$ and $w_k, \ldots, w_{N-1}$, but not with reference to $v_k, \ldots, v_{N-1}$, which substantially simplifies the problem. The minimizer of (37) will be denoted by $\bar{u}^{(k)}$. To control the system, only the first element of $\bar{u}^{(k)}$ is used and the whole procedure is repeated in subsequent steps. Note that $\bar{u}^{(k)}$ depends on $Y_k$ as required in (4). As a consequence, $X$ depends on $U$ and it is possible that IBC generates a feedback strategy in the sense of Definition 1. The minimizer of (37) can be considered a compromise between open-loop control (first term) and learning (second term). The intensity of learning is given by $\nu_k$. If $\nu_k = 0$, then IBC becomes an *open-loop feedback* strategy, which is generally not optimal.

**Remark 2.** If the system (1), (2) is linear and the disturbances are additive Gaussian white noise signals, then the mutual information in (37) does not depend on control (Bania, 2018, Thm. 3.1). As a consequence, application of IBC to linear Gaussian systems with quadratic cost gives a well-known result, i.e., the Kalman filter and the LQ controller.

## 5. Examples

**Example 1.** To illustrate the main idea of IBC, let us start from the very simple example of the integrator with unknown gain. Let

$$x_{k+1} = x_k + \theta u_k, \qquad y_k = x_k, \qquad \theta \in \{-1, 1\}, \qquad x_0 = 1. \tag{39}$$

The cost function is given by

$$J(\varphi_0, \varphi_1) = E(x_2^2).$$

The initial distribution of $\theta$ has the form $P(\theta = -1) = p$, $P(\theta = 1) = 1 - p$, $p \in [0, 1]$. Since $\theta$ can be treated as a second component of the state vector, (39) can be viewed as a special case of (1) and (2).

The optimal solution, obtained by dynamic programming, has the form

$$\varphi_0^* \neq 0, \quad \varphi_1^*(y_1) = \frac{\varphi_0^* y_1}{1 - y_1}. \tag{40}$$

It follows from (39) and (40) that $x_2 = 0$. Hence the minimal value of the cost is $J^* = 0$. The observation $y_1$ contains information about $\theta$ if, and only if, $u_0 \neq 0$. Hence $I(y_1;\theta) > 0$ if, and only if, $u_0 \neq 0$.

Let $\nu_0 = 1$. According to (37), in the first step the cost

$$J(u_0, u_1) = E(x_2^2|y_0) - I(y_1;\theta)$$

should be minimized. Calculation of the expectation gives

$$J(u_0, u_1) = (u_0 + u_1)^2 + 2(1-2p)(u_0 + u_1) + 1 - I(y_1;\theta).$$

We know that $I(y_1, \theta) > 0$ if, and only if, $u_0 \neq 0$. Hence the optimal solution in the first step is

$$u_0 \neq 0, \quad u_1 = 2p - 1 - u_0.$$

In the second step we minimize

$$J(u_1) = E\{x_2^2|(y_0, y_1)\} = (y_1 + \hat{\theta}u_1)^2,$$

where

$$\hat{\theta} = \frac{y_1 - 1}{u_0}$$

denotes the estimate of $\theta$ obtained on the basis of $u_0$ and $y_0$. Minimization gives

$$u_1 = \frac{u_0 y_1}{1 - y_1},$$

which is exactly the optimal solution given by (40). Thus, the IBC method allowed us to find an optimal solution, without using dynamic programming. ♦

**Example 2.** Due to various modelling inaccuracies, in real-life applications the parameters are not constant, but they are rather stochastic processes. As an example of a system with parametric noise we will first consider the one-dimensional deterministic system

$$\dot{\eta}(t) = -a_c\eta(t) + (b_c + \epsilon(t))u(t) + g_{2c}\zeta(t), \qquad (41)$$

where $\epsilon(t)$ and $\zeta(t)$ represent changes in the gain and the input disturbances, respectively. The control input is denoted by $u(t) \in \mathbb{R}$. If we assume that $\epsilon$ is a Wiener process and $\zeta$ is white noise, then (41) can be written as a system o two Ito equations,

$$\mathrm{d}x = (A_c(u)x + B_c u)\mathrm{d}t + G_c\,\mathrm{d}w, \qquad (42)$$

$$A_c(u) = \begin{bmatrix} 0 & 0 \\ u & -a_c \end{bmatrix}, \quad B_c = \begin{bmatrix} 0 \\ b_c \end{bmatrix}, \quad G_c = \begin{bmatrix} g_{1c} & 0 \\ 0 & g_{2c} \end{bmatrix}. \qquad (43)$$

Processes $w_1(t)$ and $w_2(t)$ are mutually independent standard Wiener processes. Parameters $a_c, b_c, g_{1c}, g_{2c}$ are positive numbers. The observation equation has the form

$$y_k = x_2(t_k) + v_k, \quad k = 0, 1, 2, \ldots, \qquad (44)$$

where $v_k = N(0, s_v), s_v > 0$, $t_k = kT_0$, $T_0 > 0$. If control is piecewise constant, i.e., $u(t) = u_k, t \in [t_k, t_{k+1})$, then the discrete-time version of (42) and (44) is given by

$$x_{k+1} = A(u_k)x_k + Bu_k + \sqrt{D(u_k)}w_k, \qquad (45)$$

$$y_k = Cx_k + v_k, \qquad (46)$$

where

$$A(u_k) = A_0 + A_1 u_k, \qquad (47)$$

$$D(u_k) = D_0 + D_1 u_k + D_2 u_k^2, \qquad (48)$$

$$A_0 = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & 0 \\ a_3 & 0 \end{bmatrix}, \qquad (49)$$

$$D_0 = \begin{bmatrix} d_1 & 0 \\ 0 & d_3 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & d_2 \\ d_2 & 0 \end{bmatrix}, \qquad (50)$$

$$D_2 = \begin{bmatrix} 0 & 0 \\ 0 & d_4 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ b \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 \end{bmatrix}. \qquad (51)$$

The matrices $A, B, D$ can be calculated by using the well-known discretization rules:

$$A = e^{A_c T_0},$$

$$B = \int_0^{T_0} e^{A_c\tau} B_c\,\mathrm{d}\tau,$$

$$D = \int_0^{T_0} e^{A_c\tau} G_c^2 e^{A_c^T\tau}\,\mathrm{d}\tau.$$

The input noise is a sequence of mutually independent Gaussian random variables, i.e., $w_k \sim N(0, I_{2\times 2})$, where $I_{2\times 2}$ denotes the identity matrix of order 2. The initial condition is given by $x_0 \sim N(m_0^-, S_0^-)$.

The cost functional is given by

$$J(\varphi) = \tfrac{1}{2}E\{q_1 x_{1,2}^2 + r_0\varphi_0^2 + q_2 x_{2,2}^2 + r_1\varphi_1^2\}, \qquad (52)$$

where $x_{k,2}$ denotes the second component of $x_k$ and $q_k \geq 0$, $r_k > 0$. Since this problem is solved by Bania (2017), only the main results will be presented and some laborious transformations will be omitted. To simplify the notation, we will skip some of the function's arguments; in particular, instead of $m_k(Y_k, U_k), S_k(U_k), \varphi_k(Y_k)$, we will write briefly $m_k, S_k, \varphi_k$ etc. It is shown by Bania

(2018) that the joint density of $x_k, Y_k$ and the conditional density of $x_{k+1}$ are given by

$$p(x_k, Y_k) = N(x_k, m_k, S_k) \times \prod_{i=0}^{k} N(y_i, Cm_i^-, W_i), \quad (53)$$

$$p(x_{k+1}|Y_k) = N(x_{k+1}, m_{k+1}^-, S_{k+1}^-), \quad (54)$$

where

$$W_i = (s_v + CS_i^- C^T), \quad (55)$$

$$S_i = S_i^- - S_i^- C^T W_i^{-1} C S_i^-, \quad (56)$$

$$m_i = m_i^- + S_i C^T s_v^{-1}(y_i - Cm_i^-), \quad (57)$$

$$m_{i+1}^- = A(u_i)m_i + Bu_i, \quad (58)$$

$$S_{i+1}^- = A(u_i)S_i A(u_i)^T + D(u_i), \quad (59)$$

$$i = 0, 1, \ldots, k.$$

Note that Eqns. (55)–(59) describe the Kalman filter for (45) and (46).

**Optimal solution.** According to (4)–(6), the strategy $\varphi$ consists of two mappings, $u_0 = \varphi_0(y_0)$ and $u_1 = \varphi_1(y_0, y_1)$. The optimal solution can be found by dynamic programming. It is shown by Bania (2017) that the optimal strategy is given by

$$\varphi_0^*(y_0) = \arg\min_{u_0 \in \mathbb{R}} R_0(u_0, y_0), \quad (60)$$

$$\varphi_1^*(u_0, y_0, y_1) = -\frac{\beta_1(u_0, y_0, y_1)}{\alpha_1(u_0, y_0, y_1)}, \quad (61)$$

where

$$R_0(u_0, y_0) = \tfrac{1}{2}\alpha_0 u_0^2 + \beta_0 u_0 + \gamma_0 + V_1(u_0, y_0), \quad (62)$$

$$V_1(u_0, y_0) = \int N(y_1, Cm_1^-, W_1) \times R_1(u_0, y_0, y_1, \varphi_1^*(u_0, y_0, y_1))\, \mathrm{d}y_1, \quad (63)$$

$$R_1(u_0, y_0, y_1, \varphi_1) = \tfrac{1}{2}\alpha_1 \varphi_1^2 + \beta_1 \varphi_1 + \gamma_1, \quad (64)$$

$$\alpha_0(y_0) = (A_1 m_0 + B)^T Q_1 (A_1 m_0 + B) + \langle A_1 S_0 A_1^T + D_2, Q_1\rangle + r_0,$$

$$\beta_0(y_0) = (A_1 m_0 + B)^T Q_1 A_0 m_0 + \tfrac{1}{2}\langle A_0 S_0 A_1^T + A_1 S_0 A_0^T + D_1, Q_1\rangle,$$

$$\gamma_0(y_0) = \tfrac{1}{2} m_0^T A_0^T Q_1 A_0 m_0 + \tfrac{1}{2}\langle A_0^T S_0 A_0 + D_0, Q_1\rangle,$$

$$\alpha_1(u_0, y_0, y_1) = (A_1 m_1 + B)^T Q_2 (A_1 m_1 + B) + \langle A_1 S_1 A_1^T + D_2, Q_2\rangle + r_1,$$

$$\beta_1(u_0, y_0, y_1) = (A_1 m_1 + B)^T Q_2 A_0 m_1 + \tfrac{1}{2}\langle A_0 S_1 A_1^T + A_1 S_1 A_0^T + D_1, Q_2\rangle,$$

$$\gamma_1(u_0, y_0, y_1) = \tfrac{1}{2} m_1^T A_0^T Q_2 A_0 m_1 + \tfrac{1}{2}\langle A_0^T S_1 A_0 + D_0, Q_2\rangle,$$

$$Q_k = \mathrm{diag}(0, q_k), \quad k = 1, 2.$$

Matrices $S_k$ and vectors $m_k$ are given by (56) and (57), respectively. The inner product of matrices $A$ and $B$ is denoted by $\langle A, B\rangle = \mathrm{tr}(A^T B)$.

**Information based solution.** We will first calculate the conditional expectation. Write $\xi = q_1 x_{1,2}^2 + r_0 u_0^2 + q_2 x_{2,2}^2 + r_1 u_1^2$. After calculation of the integrals we get

$$E(\xi|Y_0) = \sum_{i=1}^{2} \left(\mu_i^T Q_i \mu_i + r_{i-1} u_{i-1}^2 + \mathrm{tr}(Q_i \Sigma_i)\right), \quad (65)$$

where

$$\mu_{i+1} = A(u_i)\mu_i + Bu_i, \quad \mu_0 = m_0, \quad (66)$$

$$\Sigma_{i+1} = A(u_i)\Sigma_i A(u_i)^T + D(u_i), \quad \Sigma_0 = S_0, \quad (67)$$

$$Q_i = \mathrm{diag}(0, q_i), \quad i = 0, 1. \quad (68)$$

The conditional mean $m_0$ and covariance $S_0$ are given by (56) and (57), where $S_0^-$, $m_0^-$ are known *a priori*.

Now the mutual information will be calculated. It follows from (53) that

$$p(x_1, y_1|y_0) = N(x_1, m_1, S_1) N(y_1, Cm_1^-, W_1). \quad (69)$$

According to Section 4, we have $X_0^+ = x_1$, $Y_0^+ = y_1$, $Y_0 = y_0$, $u^{(0)} = (u_0, u_1)^T$. Hence $p(X_0^+, Y_0^+|Y_0) = p(x_1, y_1|y_0)$ and calculation of the integral (36) yields

$$I_0(u_0|y_0) = \tfrac{1}{2}\ln\left(1 + \frac{C\Sigma_1(u_0)C^T}{s_v}\right). \quad (70)$$

By assumption we have $I_1(u^{(1)}|Y_1) = 0$. According to (37), in the first step, we minimize the cost

$$J_0(u_0, u_1, y_0) = \tfrac{1}{2}\sum_{i=1}^{2}\left(|\mu_i|_{Q_i}^2 + r_{i-1}u_{i-1}^2 + \mathrm{tr}(Q_i\Sigma_i)\right) - \nu_0 I_0(u_0|y_0). \quad (71)$$

After performing calculations we get

$$J_0(u_0, u_1, y_0) = \tfrac{1}{2}\left(|\mu_1|_{Q_1}^2 + r_0 u_0^2 + \mathrm{tr}(Q_1\Sigma_1(u_0))\right) - \nu_0 I_0(u_0|y_0) + \tfrac{1}{2}\bar{\alpha}_0 u_1^2 + \bar{\beta}_0 u_1 + \bar{\gamma}_0, \quad (72)$$

where

$$\begin{aligned}\bar{\alpha}_0(u_0,y_0) &= (A_1\mu_1+B)^TQ_2(A_1\mu_1+B)\\ &\quad+\langle A_1S_1A_1^T+D_2,Q_2\rangle+r_1,\\ \bar{\beta}_0(u_0,y_0) &= (A_1\mu_1+B)^TQ_2A_0\mu_1\\ &\quad+\tfrac{1}{2}\langle A_0S_1A_1^T+A_1S_1A_0^T+D_1,Q_2\rangle,\\ \bar{\gamma}_0(u_0,y_0) &= \tfrac{1}{2}\mu_1^TA_0^TQ_2A_0\mu_1\\ &\quad+\tfrac{1}{2}\langle A_0^TS_1A_0+D_0,Q_2\rangle.\end{aligned}$$

The optimal value of $u_1$ as a function of $u_0$ and $y_0$ is found by minimization of (72) with respect to $u_1$,

$$u_1(u_0,y_0)=-\frac{\bar{\beta}_0(u_0,y_0)}{\bar{\alpha}_0(u_0,y_0)}. \tag{73}$$

Substitution of (73) into (72) gives an analogue of Eqn. (62),

$$\begin{aligned}\Psi(u_0,y_0) &= J_0(u_0,u_1(u_0,y_0),y_0)\\ &= \tfrac{1}{2}\big(\mu_1^TQ_1\mu_1+r_0u_0^2\\ &\quad+\operatorname{tr}(Q_1\Sigma_1(u_0))\big)\\ &\quad-\nu_0I_0(u_0|y_0)+\bar{\gamma}_0-\frac{\bar{\beta}_0^2}{2\bar{\alpha}_0},\end{aligned} \tag{74}$$

where, for simplicity, the function $J_0(u_0,u_1(u_0,y_0),y_0)$ is denoted by $\Psi(u_0,y_0)$. Minimization of (74) with respect to $u_0$ gives $\bar{u}_0(y_0)$, which is the information-based strategy in the first step. After this, new information contained in $y_1$ is used by the filter (53)–(59) and the new state and covariance estimates ($m_1$ and $S_1$) are available. Thus, according to Section 4, in the second step we minimize

$$J_1(u_1,Y_1)=\tfrac{1}{2}\left(\mu_2^TQ_2\mu_2+r_1u_1^2+\operatorname{tr}(Q_2\Sigma_2)\right), \tag{75}$$

where

$$\mu_2=A(u_1)m_1+Bu_1, \tag{76}$$

$$\Sigma_2=A(u_1)S_1A(u_1)^T+D(u_1), \tag{77}$$

and the control value $u_0$ (optimal or not) is treated as a fixed parameter. Completing the calculations in much the same way as above, we get

$$J_1(u_1)=\tfrac{1}{2}\bar{\alpha}_1u_1^2+\bar{\beta}_1u_1+\bar{\gamma}_1, \tag{78}$$

where

$$\begin{aligned}\bar{\alpha}_1(u_0,y_0,y_1) &= (A_1m_1+B)^TQ_2(A_1m_1+B)\\ &\quad+\langle A_1S_1A_1^T+D_2,Q_2\rangle+r_1,\\ \bar{\beta}_1(u_0,y_0,y_1) &= (A_1m_1+B)^TQ_2A_0m_1\\ &\quad+\tfrac{1}{2}\langle A_0S_1A_1^T+A_1S_1A_0^T+D_1,Q_2\rangle,\\ \bar{\gamma}_1(u_0,y_0,y_1) &= \tfrac{1}{2}m_1^TA_0^TQ_2A_0m_1\\ &\quad+\tfrac{1}{2}\langle A_0^TS_1A_0+D_0,Q_2\rangle.\end{aligned}$$

The optimal information-based solution in the second step is given by

$$\bar{u}_1=-\frac{\bar{\beta}_1(u_0,y_0,y_1)}{\bar{\alpha}_1(u_0,y_0,y_1)}. \tag{79}$$

Comparing (61) and (79), we conclude that $\bar{u}_1$ will equal the optimal control $\varphi_1^*(y_0,y_1)$, provided that $\bar{u}_0$ is equal to the optimal control $\varphi_0^*(y_0)$. If this last condition is fulfilled, then the optimal strategy can be recovered by IBC. We will show below that this is possible provided that parameter $\nu_0$ in (71) is appropriately chosen.

**Numerical example.** The parameters of the continuous-time system (41)–(43) were $a_c=1$, $b_c=1$, $g_{1c}=g_{2c}=\sqrt{2}$, $s_v=0.01$, $T_0=0.1$. The parameters of the corresponding discrete-time system (45)–(51) were equal to $a_1=1.0$, $a_2=0.90483$, $a_3=b=0.09516$, $d_1=0.2$, $d_2=9.674\ 10^{-3}$, $d_3=0.18126$, $d_4=6.189\ 10^{-4}$. The weights were $r_0=r_1=10^{-3}$, $q_0=0$, $q_1=1$. The initial conditions were equal to $m_0=(0,0)^T$, $S_0=\operatorname{diag}(s_{0,1},s_{0,2})$, $s_{0,1}=5$, $s_{0,2}=0.1$. For simplicity, an assumption was made that $y_0=0$. The results of numerical calculations of functions $R_0$, (62) and $\Psi$, (74), are shown in Fig. 1.

The optimal control $\bar{u}_0$ is ambiguous and equal to $\pm2.0352$. Although the initial condition is concentrated around zero, the optimal control is non-zero. This is a dual effect, described first by Feldbaum (1965). Let us observe that parameter $\nu_0$ can be chosen such that function $\Psi$; cf. (74), has minima at the same points as function $R_0$; cf. (62). This implies the main conclusion that optimal feedback can be realized by information based

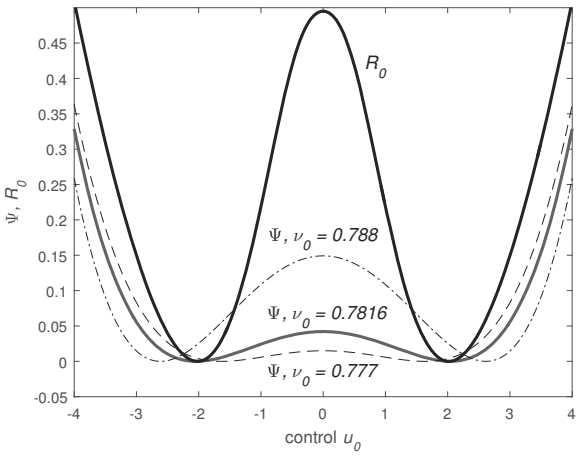


Fig. 1. Graph of functions $R_0$, (62), and $\Psi$, (74), for three values of $\nu_0$. If $\nu_0\approx0.7816$, then function $\Psi$ has minima at the same points as $R_0$ and the optimal strategy (60), (61) can be recovered by IBC. For clarity, the graphs of both the functions are scaled and shifted vertically.

control, at least in this example. It is important to that the information based solution has been found without using dynamic programming, which substantially reduces computational complexity. ♦

## 6. Computational issues and practical implementation of IBC

Minimization of the cost (37) requires in advance the solution of the following problems:

1. calculation of the filtering distribution $p(x_k|Y_k)$,
2. calculation of the expectation in (37),
3. calculation of the mutual information (36).

The filtering distribution can be calculated by an unscented Kalman filter (UKF), a particle filter (PF) or a Gaussian sum filter (GSF) (see the works of Särkä (2013), Alspach and Sorenson (1972) for details). Since both the theory and practical implementations of these filters are well developed, we will assume below that $p(x_k|Y_k)$ or its approximation is known. Let $x_{k,i}, i = 1, \ldots, n_s$, denote samples from $p(x_k|Y_k)$, and let $x_{N,i}$, $Y^+_{k,i}$ denote the final state and observations generated by (1), (2) with the initial condition $x_{k,i}$. Then it is easy to observe that samples $x_{N,i}$, $Y^+_{k,i}$ are drawn from $p(x_N|Y_k, u^{(k)})$ and $p(Y^+_k|Y_k, u^{(k)})$, respectively. Hence, the Monte Carlo approximation of the expectation in (37) is given by

$$E\{L(x_N)|Y_k\} \approx \frac{1}{n_s}\sum_{i=1}^{n_s} L(x_{N,i}). \tag{80}$$

Calculation of the mutual information (36) cannot be easily done without additional simplifications. Therefore, below we will briefly discuss special cases that are relatively easy to solve. Let us assume that Eqn. (2) has the form

$$y_k = h(x_k) + v_k, \tag{81}$$

where $v_k \sim N(0, S_v)$. By direct calculation we get

$$I_k(u^{(k)}|Y_k) = H_k(u^{(k)}|Y_k) - \frac{n_k}{2}\ln 2\pi e|S_V|, \tag{82}$$

where $n_k$ denotes the size of $Y^+_k$ and

$$\begin{aligned} &H_k(u^{(k)}|Y_k) \\ &= -\int p(Y^+_k|Y_k, u^{(k)}) \ln p(Y^+_k|Y_k, u^{(k)}) dY^+_k \end{aligned} \tag{83}$$

is an entropy of $Y^+_k$, predicted at time $k$. The kernel density estimator (KDE) of $p(Y^+_k|Y_k, u^{(k)})$ has the form

$$\hat{p}_{n_s}(Y^+_k|Y_k, u^{(k)}) = \frac{1}{n_s}\sum_{i=1}^{n_s} N(Y^+_k, Y^+_{k,i}, \sigma^2 I_{n_k}), \tag{84}$$

where $I_{n_k}$ is the identity matrix of order $n_k$ and the bandwidth parameter is given by

$$\sigma = \left(\frac{4}{n_s(n_k+2)n_k^2}\right)^{\frac{1}{n_k+4}}. \tag{85}$$

Now, the entropy estimator can be constructed as follows:

$$\begin{aligned} H_k(u^{(k)}|Y_k) &= E(-\ln p(Y^+_k|Y_k, u^{(k)})) \\ &\approx -\frac{1}{n_s}\sum_{i=1}^{n_s} \ln \hat{p}_{n_s}(Y^+_{k,i}|Y_k, u^{(k)}) \\ &= \frac{n_k}{2}\ln(2\pi\sigma^2) \\ &\quad - \frac{1}{n_s}\sum_{i=1}^{n_s} \ln\left(\frac{1}{n_s}\sum_{j=1}^{r} e^{-D_{i,j}}\right), \end{aligned} \tag{86}$$

where

$$D_{i,j} = \frac{1}{2\sigma^2}||Y^+_{k,i} - Y^+_{k,j}||^2. \tag{87}$$

Combining (86) and (82), we get

$$\begin{aligned} &I_k(u^{(k)}|Y_k) \\ &\quad \approx \frac{n_k}{2}\ln\frac{\sigma^2}{e|S_V|} - \frac{1}{n_s}\sum_{i=1}^{n_s} \ln\left(\frac{1}{n_s}\sum_{j=1}^{r} e^{-D_{i,j}}\right). \end{aligned} \tag{88}$$

On the basis of (37), (80) and (88), we have

$$\begin{aligned} &J_k(u^{(k)}) \\ &= E\{L(x_N)|Y_k\} - \nu_k I_k(u^{(k)}|Y_k) \\ &\approx \frac{1}{n_s}\sum_{i=1}^{n_s}\left(L(x_{N,i}) + \nu_k \ln\left(\frac{1}{n_s}\sum_{j=1}^{r} e^{-D_{i,j}}\right)\right) \\ &\quad - \frac{\nu_k n_k}{2}\ln\frac{\sigma^2}{e|S_V|}. \end{aligned} \tag{89}$$

Since the last term in (89) does not depend on $u^{(k)}$, finally, the cost function to be minimized is given by

$$\begin{aligned} &\bar{J}_k(u^{(k)}) \\ &= \frac{1}{n_s}\sum_{i=1}^{n_s}\left(L(x_{N,i}) + \nu_k \ln\left(\frac{1}{n_s}\sum_{j=1}^{r} e^{-D_{i,j}}\right)\right). \end{aligned} \tag{90}$$

Convergence conditions for (84) and (86) are given by Jiang (2017) and Joe (1989). These conditions can be fulfilled assuming that $p_w, p_v, f, h$ are sufficiently regular. In particular, if $p(Y^+_k|Y_k, u^{(k)})$ is bounded, globally Lipschitz, $C^4$ and its second order partial derivatives are all upper bounded by an integrable function, then (84) converges uniformly and the variance of (86) tends to zero as $n_s \to \infty$. The convergence rate is $O(n^{-\alpha})$, $\alpha \in (0, 1/2]$.

Since $f, h, L$ are $C^2$, then cost (90) is also $C^2$ with respect to $u^{(k)}$ and its gradient can be effectively calculated by solving the associated adjoint equation. Then minimization of (90) can be performed by combining global search algorithms (e.g., differential evolution, simulated annealing, genetic algorithms) with stochastic quasi-Newton methods as local solvers Byrd *et al.* (2016).

Control of a linear system with a finite number of unknown parameters and with a quadratic cost function is another special case that is tractable by IBC. Analytical formulas describing the cost function and the filtering distribution were given by Bania (2018). Various types of recursive filters are also analyzed by Bania and Baranowski (2016; 2017) or Baranowski *et al.* (2017). A computationally effective lower bound to the mutual information (36) that can be utilized to construct an upper bound to the cost is given by Bania (2019). Thus, in this particular case, the cost (37) and its gradient can be calculated without using Monte Carlo sampling, and the control problem is relatively easy to solve.

## 7. Conclusions

Lower bounds of the cost function in stochastic optimal control problems were analysed in terms of information exchange between the system and the controller. It was proved, under weak assumptions, that the cost function is lower bounded by some decreasing function of mutual information between the system trajectory and control variables. Under some additional regularity conditions, the lower bound obtained above is a linear function of information, but the constant $q$ appearing in (20) depends on system dynamics. It also follows from Theorem 1 and (22) that the minimum value of the cost is determined by the capacity of the measurement channel (i.e., the maximal value of $I(X;Y)$). Next, on the basis of the Touchette–Lloyd inequality, a new one-step lower bound (26) was established, provided that the cost function is quadratic. This bound is independent of system dynamics and in that sense universal.

The inequalities (20) and (22) indicate that restrictions in communication between parts of the system prevent certain states from being reached. One of the examples of such a phenomenon is synchronization in dynamical networks. Since the synchronization problem can be interpreted as a stochastic control task, communication constraints of the form $I(X;Y) < C$ imply that $J_o - J(\varphi) \leq C$. In consequence, synchronization may be lost if $C$ is too small. This was confirmed by Huang *et al.* (2012).

The conclusion resulting from the analysis of information-theoretic bounds is that the feedback controller must actively (if possible) generate information about the state of the system. On the basis of these results, the *information based control* approach to stochastic control was proposed. The main idea of IBC consists in replacing the original control problem with a sequence of simpler, auxiliary control problems. The cost function to be minimized in these auxiliary problems consists of two parts: the predicted expectation of the cost conditioned on available measurements and the penalty function for information deficiency. As the penalty function, the predicted mutual information between the trajectory and measurements was used. Hence the method enforces active generation of information about the system state and is able to generate a feedback strategy. The IBC method can be also viewed as a modification of the OLFO (Tse, 1974) algorithm or as a compromise between control and state estimation.

It follows from Section 6 that minimization of the cost (37) can be performed by standard optimization algorithms, without using dynamic programming. Hence the computational complexity of the IBC is substantially smaller than that of DP. This feature of IBC introduces the possibility of solving large-scale tasks, which is impossible with DP. It was shown that IBC is able to find optimal solutions, provided that learning intensity (parameter $\nu_k$) is appropriately selected. The optimal value of $\nu_k$ can be tuned experimentally but, at the current stage of research, this problem is not resolved. The ability of IBC to find an optimal solution is surprising but, due to the complexity of the problem, convergence to an optimal solution is difficult to investigate and is not proven.

Effective calculation of the mutual information or the development of its approximation is a crucial issue and some methods from optimal experimental design and fault detection theory can be adopted here (see Bania, 2019; Uciński, 2004; Korbicz *et al.*, 2004). It is also possible to use the information lower bound proposed by Kolchinsky and Tracey (2017).

Application of the IBC method to solve more realistic control problems and development of information-based model predictive control algorithms is planned as a part of future works.

## References

Alpcan, T., Shames, I., Cantoni, M. and Nair, G. (2015). An information-based learning approach to dual control, *IEEE Transactions on Neural Networks and Learning Systems* **26**(11): 2736–2748.

Alspach, D. and Sorenson, H. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations, *IEEE Transactions on Automatic Control* **17**(4): 439–448.

Åström, K. and Wittenmark, B. (1995). *Adaptive Control, Second Edition*, Dover Publications, New York, NY.

Banek, T. (2010). Incremental value of information for discrete-time partially observed stochastic systems, *Control and Cybernetics* **39**(3): 769–781.

Bania, P. (2017). Simple example of dual control problem with almost analytical solution, *Proceedings of the 19th Polish Control Conference, Krakow, Poland*, pp. 55–64, DOI: 10.1007/978-3-319-60699-6-7.

Bania, P. (2018). Example for equivalence of dual and information based optimal control, *International Journal of Control* **38**(5): 787–803, DOI: 10.1080/00207179.2018.1436775.

Bania, P. (2019). Bayesian input design for linear dynamical model discrimination, *Entropy* **21**(4): 1–13, DOI: 10.3390/e21040351.

Bania, P. and Baranowski, J. (2016). Field Kalman filter and its approximation, *55th IEEE Conference on Decision and Control, Las Vegas, NV, USA*, pp. 2875–2880, DOI: 10.1109/CDC.2016.7798697.

Bania, P. and Baranowski, J. (2017). Bayesian estimator of a faulty state: Logarithmic odds approach, *22nd International Conference on Methods and Models in Automation and Robotics (MMAR), Miedzyzdroje, Poland*, pp. 253–257, DOI: 10.1109/MMAR.2017.8046834.

Baranowski, J., Bania, P., Prasad, I. and T., C. (2017). Bayesian fault detection and isolation using field Kalman filter, *EURASIP Journal on Advances in Signal Processing* **79**(1), DOI: 10.1186/s13634-017-0514-8.

BarShalom, Y. and Tse, E. (1976). Caution, probing, and the value of information in the control of uncertain systems, *Annals of Economic and Social Measurement* **5**(3): 323–337.

Brechtel, S., Gindele, T. and Dillmann, R. (2013). Solving continuous POMDPs: Value iteration with incremental learning of an efficient space representation, *Proceedings of the 30th International Conference on International Conference on Machine Learning, ICML'13, Atlanta, GA, USA,* Vol. 28, pp. III–370–III–378.

Byrd, R., Hansen, S., Nocedal, J. and Singer, Y. (2016). A stochastic quasi-Newton method for large-scale optimization, *SIAM Journal on Optimization* **26**(2): 1008–1031.

Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory, Second Edition*, John Wiley & Sons, Inc., Hoboken, NJ.

Delvenne, J.C. and Sandberg, H. (2013). Towards a thermodynamics of control: Entropy, energy and Kalman filtering, *52nd IEEE Conference on Decision and Control, Florence, Italy*, pp. 3109–3114.

Dolgov, M. (2017). *Approximate Stochastic Optimal Control of Smooth Nonlinear Systems and Piecewise Linear Systems*, PhD thesis, Karlsruhe Institute of Technology, Karlsruhe.

Feldbaum, A.A. (1965). *Optimal Control Systems*, Academic Press, New York, NY.

Filatov, N.M. and Unbehauen, H. (2004). *Adaptive Dual Control: Theory and Applications*, Springer-Verlag, Berlin/Heidelberg.

Hijab, O. (1984). Entropy and dual control, *23rd Conference on Decision and Control, Las Vegas, NV, USA*, pp. 45–50.

Huang, C., Ho, D.W.C., Lu, J. and Kurths, J. (2012). Partial synchronization in stochastic dynamical networks with switching communication channels, *Chaos: An Interdisciplinary Journal of Nonlinear Science* **22**(2): 023108, DOI: 10.1063/1.3702576.

Jiang, H. (2017). Uniform convergence rates for kernel density estimation, *Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia*, pp. 1694–1703.

Joe, H. (1989). Estimation of entropy and other functionals of a multivariate density, *Annals of the Institute of Statistical Mathematics* **41**(4): 683–697.

Kolchinsky, A. and Tracey, B.D. (2017). Estimating mixture entropy with pairwise distances, *Entropy* **19**(361): 1–17.

Korbicz, J., Koscielny, J.M., Kowalczuk, Z. and Cholewa, W. (2004). *Fault Diagnosis: Models, Artificial Intelligence, Applications*, Springer-Verlag, Berlin/Heidelberg.

Kozlowski, E. and Banek, T. (2011). Active learning in discrete time stochastic systems, *in* J. Jozefczyk and D. Orski (Eds), *Knowledge-Based Intelligent System Advancements: Systemic and Cybernetic Approaches*, Information Science References, New York, NY, pp. 350–371.

Mitter, S.K. and Newton, N.J. (2005). Information and entropy flow in the Kalman–Bucy filter, *Journal of Statistical Physics* **118**(1): 145–176.

Porta, J.M., Vlassis, N., Spaan, M.T. and Poupart, P. (2006). Point-based value iteration for continuous POMDPs, *Journal of Machine Learning Research* **7**(1): 2329–2367.

Sagawa, T. and Ueda, M. (2013). Role of mutual information in entropy production under information exchanges, *New Journal of Physics* **15**(125012): 2–23.

Saridis, G.N. (1988). Entropy formulation of optimal and adaptive control, *IEEE Transactions on Automatic Control* **33**(8): 713–721.

Särkä, S. (2013). *Bayesian Filtering and Smoothing*, Cambridge University Press, New York, NY.

Taticonda, S. and Mitter, S.K. (2004). Control under communication constraints, *IEEE Transactions on Automatic Control* **49**(7): 1056–1068.

Thrun, S. (2000). Monte Carlo POMDPs, *in* S. Solla *et al.* (Eds), *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, pp. 1064–1070.

Touchette, H. (2000). *Information-theoretic Aspects in the Control of Dynamical Systems* Master's thesis, MIT, Cambridge, MA, `https://pdfs.semanticscholar.org/c915/088f514d937f5d1c666221c95d731532101e.pdf`.

Touchette, H. and Lloyd, S. (2000). Information-theoretic limits of control, *Physical Review Letters* **84**(6): 1156–1159.

Touchette, H. and Lloyd, S. (2004). Information-theoretic approach to the study of control systems, *Physica A* **331**(1): 140–172.

Tsai, Y.A., Casiello, F.A. and Loparo, K.A. (1992). Discrete-time entropy formulation of optimal and adaptive control problems, *IEEE Transactions on Automatic Control* **37**(7): 1083–1088.

Tse, E. (1974). Adaptive dual control methods, *Annals of Economic and Social Measurement* **3**(1): 65–82.

Uciński, D. (2004). *Optimal Measurement Methods for Distributed Parameter System Identification*, CRC Press, Boca Raton, FL.

Zabczyk, J. (1996). *Chance and Decision. Stochastic Control in Discrete Time*, Quaderni Scuola Normale di Pisa, Pisa.

Zhao, D., Liu, J., Wu, R., Cheng, D. and Tang, X. (2019). An active exploration method for data efficient reinforcement learning, *International Journal of Applied Mathematics and Computer Science* **29**(2): 351–362, DOI: 10.2478/amcs-2019-0026.

**Piotr Bania** received his PhD degree from the AGH University of Science and Technology, Cracow, Poland, in 2008. He is currently an associate professor there. His research interests cover predictive control, stochastic control and filtering theory. He has published over 30 papers in refereed journals and conferences.

Received: 9 March 2019
Revised: 17 October 2019
Accepted: 31 October 2019