**Preprints of the**
**15th IFAC Symposium on System Identification**
**Saint-Malo, France, July 6-8, 2009**

# On Implicit Approximation of the Bellman Equation ⋆

**Miroslav Pištěk** *

* *Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic (e-mail: mid@centrum.cz).*

**Abstract:** In this article, an efficient algorithm for an optimal decision strategy approximation is introduced. It approximates the Bellman equation without omitting the principal uncertainty stemming from incomplete knowledge. Thus, the approximated optimal strategy retains the ability to constantly verify the current knowledge. An integral part of the proposed solution is a reduction in memory demands using HDMR approximation. The result of this method is a linear algebraic system for an approximated upper bound on the Bellman function. The analysis of the approximation error has not been considered here. One illustrative example has been completely resolved.

Keywords: function approximation, Bellman equation

## 1. MOTIVATION

The main focus of this article is to develop an approximation tool suitable for enlarging the class of computationally feasible decision-making problems. This work copes with the principal problem within the stochastic dynamic programming, which is known as *curse of dimensionality*, see [3]. In the contemporary state of arts, there is a lack of approximation techniques capable of encompassing problems with a larger decision-making horizon. The aim of this work is to reduce memory demands necessary to store the optimal strategy. To this end, properties of an approximative tool called High Dimensional Model Representations (thereinafter "HDMR") are promising. It was stimulated by applications in chemistry, see [1], which focused on reducing enormous memory demands of the involved models. In its background, there stands a simple observation: only low-order correlations amongst the input variables have a significant impact upon the outputs of a typical model.

A general form of a HDMR expansion reads

$$g(x) \approx \tilde{g}(x) \equiv \tilde{g}(x_1, x_2, \dots, x_\mu) = \tilde{g}_0 + \sum_{m=1}^{\mu} \tilde{g}_m(x_m) + \sum_{m=1}^{\mu} \sum_{n=1}^{m-1} \tilde{g}_{mn}(x_m, x_n) + \dots \quad (1)$$

Here, a zero-order component $\tilde{g}_\emptyset$ denotes a constant scalar value over the domain of $g(x)$; the first-order components $\tilde{g}_m(x_m)$ describe an independent effect of each variable $x_m$; the second-order component $\tilde{g}_{mn}(x_m, x_n)$ represents the joint effect of the variables $x_m$ and $x_n$ and so on. The experience shows that even the low-order case often provides a sufficient description of $g(x)$.

Such a function approximation (representation) yields two main advantages. The first one is the data reduction. The memory space necessary to store all the values of the original function $g(x)$ grows exponentially with the dimension $\mu$, whereas the size growth of decomposition components is just polynomial in $\mu$. This property helps us to cope with high-dimensional problems of the real world. The second advantage is the reduction of computational complexity. In general, it allows us to split a high-dimensional linear problem into several low-dimensional subproblems.

The outline of this work is as follows. Section 2 deals with the current state of art in the decision making theory. A central point here is the presentation of the Bellman equation with its notorious difficulties, mainly the problem of a rapidly growing domain of the Bellman function. To cope with this inconvenience, an approximative technique of HDMR is introduced in detail within Section 3. Also, a system of linear equation determining an optimal function approximation is derived here. Its linearity does not match well with the non-linear Bellman equation. Thus, a linear equation for an upper bound on the Bellman function is derived, see Section 4. Connecting it with HDMR approximation, a viable technique for approximative decision making is obtained. In Section 5, there are concise instructions for the implementation of this approximation technique in real applications. As an illustration, one toy example is completely resolved. Section 6 is devoted to the conclusion.

Throughout this work, a few general conventions are followed. The domain of the variable $x$ is denoted $X$, $x \in X$. $|X|$ denotes the finite cardinality of the countable set $X$ or its Lebesque measure in case it is not countable. Next, $x_t$ is the quantity $x$ at the discrete time instant labeled by $t \in T$. The letter "$f$" is reserved for a probability density function (pdf). Its specific meaning is given through the names of its arguments. The same letter is used for conditioned pdfs, arguments in condition are separated by "—" in the argument list. Knowing $f(x|y)$, it is possible to introduce the expected value of the variable $x$ conditioned by $y$

⋆ This work was partially supported by the grant 102/08/0567 GA ČR, and by the grant 2C06001 MŠMT ČR

$$\mathcal{E}[\,x\,|\,y\,] \equiv \int_X x\, f(x|y)\, dx$$

For the vector $x \in X$, $X \subset \mathbb{R}^\mu$, and $m \in M \equiv \{1, \ldots, \mu\}$, $x^m$ denotes its $m$-th coordinate. Therefore, it reads $x = (x^1, \ldots, x^\mu)$. Taking some $N = \{n^1, \ldots, n^\nu\} \subset M$, a projection $x/_N \in \mathbb{R}^\nu$ is defined for all $x = (x^1, \ldots, x^\mu) \in X$ in this manner $x/_N \equiv (x^{n_1}, \ldots, x^{n_\nu}) \in \mathbb{R}^\nu$. A HDMR approximation of the function $h(x)$ is marked by $\tilde{h}(x)$. For the domain of $h(x)$, $\mathrm{dom}(h)$ is used.

## 2. DECISION MAKING THEORY

Within this section, the classical results are briefly summarized together with their classical troubles. A detailed discussion is to be found in [4], for example.

The decision-making task consists in selecting the decision-maker's strategy in order to reach decision-maker's aim with respect to the part of the world (so-called system). The decision maker observes or influences the system over the finite decision making horizon $\tau < \infty$. The data (system output) observed at the time instance $t \in T \equiv \{1, \ldots, \tau\}$ is denoted by $y_t \in Y_t$. It provides the decision maker with the information about the system behavior. Analogously, the decisions (actions) are denoted as $a_t \in A_t$. It is the value that can be directly chosen by the decision maker for reaching decision-maker's aims. The strategy $\{\mathcal{R}_t\}_{t\in T}$ is a collection of mappings transforming the current experience $d(t-1) \equiv (y_{t-1}, a_{t-1}, \ldots, y_1, a_1)$ into the choice of the next decision $a_t \in A_t$.

The next thing to do is to formalize a degree of achievements of the decision-maker's aims. The idea of loss function is promising. A loss value is assigned to each possible system trajectory $d(\tau)$ respecting just one rule: the more suitable the trajectory is, the lower loss value it possesses. This way, the loss function $Z(d(\tau))$ is obtained. Often, a less general concept of the additive loss function is introduced, i.e., the case when the losses accumulate with time

$$Z(d(\tau)) = \sum_{t=1}^{\tau} z_t\,(a_t, y_t) \quad \text{where} \quad z_t(a_t, y_t) \geq 0 \qquad (2)$$

Now, it is necessary to describe the involved system. In this work, a stochastic approach is held. Thus, the system is completely described in a probabilistic manner by the following collection of pdfs called the outer model of a system

$$\{f(y_t|a_t, d(t-1))\}_{t\in T} \qquad (3)$$

There are many ways how to find these formulae, see [5].

Knowing the loss function (2) altogether with the outer model (3), the optimal strategy is determined by the Bellman theorem. It claims: the strategy $\{\mathcal{R}_t\}_{t\in T}$ selecting such decisions $a_t^{opt}$ that $a_t^{opt}$ minimize

$$V_{t-1}(d(t-1)) = \min_{a_t \in A_t} \mathcal{E}[\,z_t(y_t, a_t) + V_t(d(t))\,|\,a_t, d(t-1)\,] \qquad (4)$$

at all times $t \in T$, minimizes also the expected value of the overall loss $Z(d(\tau))$ provided the boundary condition $V_\tau \equiv 0$ is satisfied.

The essential problem is to evaluate the Bellman function $V_t$ for all $t \in T$. Its exact recursive calculation is computationally infeasible in the majority of practical applications for the reason of geometrically growing size of its domain with the increasing decision making horizon $\tau$. This paper aims at reduction in the memory demands necessary to represent the approximated strategy.

### *2.1 Sufficient Statistic*

When operating with a large amount of data, it is meaningful to compress them into a set of a smaller dimension as follows

$$\sigma_t \equiv \sigma_t(d(t)) \qquad (5)$$

Such mapping is called statistic. For the random variable $x_t$, the statistic $\sigma_t$ is sufficient if there exists $f(x_t|\sigma_t(d(t)), t)$ satisfying the following condition for all the times $t \in T$ and all the possible trajectories $d(t)$, $t \in T$

$$f(x_t|d(t)) = f(x_t|\sigma_t(d(t)), t)$$

The explicit appearance of the time coordinate in the condition is for the sake of simplicity in sequel.

The collection of the following mappings $\{S_t\}_{t\in T}$ is necessary to effectively update the statistic. $S_1(y_1, a_1) \equiv \sigma_1(y_1, a_1)$, and for all $t \in \{2, \ldots, \tau\}$, the new data $y_t \in Y_t$ observed after the decision $a_t \in A_t$ is carried out and the old statistic value $\sigma_{t-1} \equiv \sigma_{t-1}(d(t-1))$, it reads

$$S_t(y_t, a_t, \sigma_{t-1}) \equiv \sigma_t(d(t)) \qquad (6)$$

In this article, the function approximation will be searched over a statistic domain. Therefore, the knowledge of the exact statistic domain $\Sigma_t$ for all the times $t \in T$ is necessary. For $\Sigma_1$, it obviously holds $\Sigma_1 \equiv \Sigma_1(Y_1, A_1)$, and for all $t \in \{2, \ldots, \tau\}$, the domains are introduced in a recursive manner

$$\Sigma_t \equiv S_t(Y_t, A_t, \Sigma_{t-1})$$

In the context of the Bellman equation (4), the existence of the statistic $\{\sigma_t \in \Sigma_t\}_{t\in T}$ sufficient for the system model (3) is assumed. It suggests to rewrite the Bellman equation valid over all $\Sigma_t$, $t \in T$, using a shortcut

$$\sigma_{t-1} \equiv \sigma_{t-1}(d(t-1))$$

with the condition $V_\tau \equiv 0$. For all $t \in \{2, \ldots, \tau\}$ it holds

$$V_{t-1}(\sigma_{t-1}) = \min_{a_t \in A_t} \mathcal{E}[\,z_t(y_t, a_t) + V_t(S_t(y_t, a_t, \sigma_{t-1}))\,|\,a_t, \sigma_{t-1}, t-1\,] \qquad (7)$$

The previous compression of the domain of the Bellman function is a crucial step towards the solution of the problem.

## 3. HIGH DIMENSIONAL MODEL REPRESENTATION

This section is to prepare a HDMR approximation technique to reduce memory demands to represent the Bellman function defined by (7). There are many ways how to construct the decomposition like (1), see [1]. To reduce this ambiguity, it is necessary to formalize the desired properties of the decomposition.

The function Hilbert space $L^2(X)$ is an useful concept for the function approximation. Generally, it is a space

of real functions defined over $X$ with the finite norm $\|g\| \equiv \sqrt{\langle g\,,\,g\rangle}$ inducted by the following scalar product

$$\langle g\,,\,h\rangle_X \equiv \int_X g(x)\,h(x)\,dx \tag{8}$$

The optimal HDMR decomposition $\tilde{g}$ of the function $g \in L^2(X)$ is a minimizer of an approximation error evaluated in this norm, i.e., it is a function minimizing $\|g - \tilde{g}\|$.

*Partial Constancy of Decomposition Components*

To get rid of the ellipsis "..." in (1), it is suitable to index decomposition components by elements of a general index set. Consider $\mu \in \mathbb{N}$ equal to the dimension of $X$, $X \subset \mathbb{R}^\mu$. Introducing $M \equiv \{1, \ldots, \mu\}$, the decomposition component can be addressed by an element of the following index set

$$D \subsetneq \{N | N \subset M\}$$

The set's elements are the indices determining which variables the decomposition component depends on. This way, it is possible to prescribe a different component order for different variables (or groups of variables). It could be useful if there is some a priori information on their influence. The resulting HDMR decomposition of $g(x)$ has the following general form

$$\tilde{g}(x) \equiv \sum_{K \in D} \tilde{g}_K(x) \tag{9}$$

Obviously, considering decomposition components within the space $L^2(X)$ is not strict enough. For any $K \in D$, the HDMR decomposition component $\tilde{g}_K(x)$ must not depend on $x^m$ for $m \in M \setminus K$. A space of constant functions will be useful. For all $K \subset M$, they can be introduced as

$$C_K(X) \equiv \{h \,|\, \mathrm{dom}(h) = X, \underset{x,y \in X}{\forall} (x/_K = y/_K \rightarrow h(x) = h(y))\} \tag{10}$$

These functions are constant in all the variables but $x^k$, $k \in K$. Such a restriction is non-optional when talking about the HDMR approximation.

*Support Restriction of Decomposition Components*

Another restriction is necessary to guarantee the uniqueness of each separate decomposition component. The problem stems from the fact that only the overall sum of the decomposition components enters the minimization task. For instance, the constant value $\tilde{g}_\emptyset$ can be nullified and added to any higher-order decomposition component. There are many ways how to manage this ambiguity. The one proposed here aims to decrease the resulting memory. The key idea is to nullify the decomposition components on the specific border parts of their domains. Thus, for $K \subset M$ a $X_K \subset X$ is defined

$$X_K \equiv \bigcap_{m \in M \setminus K} \left\{ x \in X \;\middle|\; x^m > \min_{y \in X} y^m \right\} \tag{11}$$

Reminding the concept of the function support

$$\mathrm{supp}(h) \equiv \{x \in \mathrm{dom}(h), h(x) \neq 0\}$$

the supports of decomposition components are reduced in the way that for all $K \in D$ it reads $\mathrm{supp}(\tilde{g}_K) \subset X_K$. With the following condition put on $D$

$$\underset{K \in D}{\forall} \underset{L \subset K}{\forall} L \in D \tag{12}$$

the uniqueness of each separate decomposition component $\tilde{g}_K$, $K \in D$, is guaranteed. It is an easy exercise to verify this fact. The resulting decomposition would give the same error of approximation with or without these conditions. Next, it is necessary to take the general optimal decomposition $\{\tilde{g}_K\}_{K \in D}$, to complete the index set $D$ in the sense of (12), and, by induction from the largest to the lowest component orders to restrict their support appropriately. The only thing to take deal with is the overall sum of the components, which have to be fixed during these operations. Within this process, the exact value of each restricted component is directly calculated, i.e., the collection of the restricted components is determined uniquely.

A small example will help to clarify the used notation. If the aim is to obtain just the first order decomposition of the function $g(x_1, x_2, x_3)$, $\mathrm{dom}(g) = X_1 \times X_2 \times X_3 \subset \mathbb{R}^3$, the following choice of an index set is the right one

$$D = \{\emptyset, \{1\}, \{2\}, \{3\}\}$$

Then, $g$ is going to be approximated in this way

$$\tilde{g}(x_1, x_2, x_3) \equiv \tilde{g}_\emptyset + \tilde{g}_1(x_1) + \tilde{g}_2(x_2) + \tilde{g}_3(x_3)$$

Compare with the general form (9). If a hypothesis exists that the biggest influence originates from the cooperation of $x_2$ with $x_3$, an addition of the set $\{2, 3\}$ into the index set $D$ is a good idea. It would change the searched HDMR decomposition into this form

$$\tilde{g}_\emptyset + \tilde{g}_1(x_1) + \tilde{g}_2(x_2) + \tilde{g}_3(x_3) + \tilde{g}_{23}(x_2, x_3)$$

Afterwards, the presence of the second-order decomposition component $\tilde{g}_{23}(x_2, x_3)$ should result in a noticeable decrease in the approximation error. For the purpose of readability, $\tilde{g}_{\{2,3\}}(x_2, x_3)$ is shorten into $\tilde{g}_{23}(x_2, x_3)$, etc.

At the moment, the main function spaces are defined for any $K \subset M$ in this way

$$H_K(X) \equiv \{h \in L^2(X) \cap C_K(X) \,|\, \mathrm{supp}(h) \subset X_K\} \tag{13}$$

where $C_K(X)$ is defined by (10). The functions within this space depend only on $x^k$ for $k \in K$ and, moreover, they are nullified on the part of the border of their domains, see (11). It leads to the observation that all the (possibly) non-zero values of $h \in H_K(X)$ are fully determined by its values on the following set $X_K^\perp \subset \mathbb{R}^{|K|}$

$$X_K^\perp \equiv \{x/_K \,|\, x \in X_K \subset \mathbb{R}^\mu\} \tag{14}$$

This definition is important in the next section.

### *3.1 Optimality Conditions*

Taking $\tilde{g}_K \in H_K(X)$, $K \in D$, for the optimal decomposition $\tilde{g}$ defined in (9) it holds

$$\tilde{g} \in \bigcup_{K \in D} H_K(X)$$

On the right hand side, there is the closed subspace of $L^2(X)$ and therefore a classical result for projection on the closed subspace of the Hilbert space can be applied. It guarantees the existence and uniqueness of the function $\tilde{g}$ minimizing the approximation error $\|g - \tilde{g}\|$. And more, it prescribes conditions for the optimal decomposition $\tilde{g}$ defined in (9). For all $K \in D$ and all $h \in H_K(X)$, it holds

$$\langle\, \tilde{g} - g\,,\, h\rangle = 0$$

According to the definition of the scalar product, see (8), this equation reads

$$\int_X (\tilde{g}(x) - g(x))\, h(x)\, dx = 0 \tag{15}$$

The Dirac distribution $\delta_y$, $y \in \mathbb{R}$, denotes a linear functional defined over a space of all real functions. Consider a real function $p(x)$, it operates in this way

$$\delta_y[p] \equiv \int_{\mathbb{R}} \delta_y(x)\, p(x)\, dx \equiv p(y)$$

Its extension to a higher dimension is straightforward. Its integral, formally uncorrect representation however provides better readability for the following text. Note that, if necessary, it could be formalized directly by means of the theory of distribution.

The previously written optimality conditions (15) are valid for all $K \in D$ and for all the test functions $h \in H_K(X)$. To rewrite them in the $\delta$-distribution formalism, it is necessary to think over an effective domain of $h$ carefully. Formerly, it was deduced that such a function is fully determined by its values on $X_K^{\perp}$, see (14). On that account, a more complex distribution $\delta_{K,y}$ is defined for all $K \in D$ and all $y \in X_K^{\perp}$ in this way

$$\delta_{K,y}(x) \equiv \delta_y(x/_K)$$

Next, considering the $\delta$-distribution index as an element of

$$\mathcal{I} \equiv \left\{ (K, y) \,|\, K \in D, y \in X_K^{\perp} \right\} \tag{16}$$

it is possible to rewrite the previous optimality conditions (15) in the equivalent form valid for all $\kappa \in \mathcal{I}$

$$\int_X (\tilde{g}(x) - g(x))\, \delta_\kappa(x)\, dx = 0$$

Expanding the HDMR approximation $\tilde{g}$ in accordance with its definition, see (9), the last equations turn into

$$\sum_{L \in D} \int_X \tilde{g}_L(x)\, \delta_\kappa(x)\, dx = \int_X g(x)\, \delta_\kappa(x)\, dx \tag{17}$$

Again, considering the support of the decomposition components altogether with its constancy in some variables, see (13), this system could be represented by the linear operators $P$, $R$ and the system of equations valid for all $\kappa \in \mathcal{I}$

$$\sum_{L \in D} \int_{X_L^{\perp}} P_{\kappa,(L,x)}\, \tilde{g}_L(x)\, dx = R_\kappa[g] \tag{18}$$

where for operator elements $P_{\kappa,\lambda}$, resp. $R_\kappa[g]$, and all $\kappa, \lambda \in \mathcal{I}$ it holds

$$P_{\kappa,\lambda} \equiv \int_X \delta_\lambda(x)\, \delta_\kappa(x)\, dx \tag{19}$$

$$R_\kappa[g] \equiv \int_X g(x)\, \delta_\kappa(x)\, dx \tag{20}$$

In sequel, the linear system (18) can be written

$$P \star \tilde{g} = R[g] \tag{21}$$

This is a linear system determining precisely one optimal HDMR decomposition of $g$ minimizing its approximation error in the norm of $L^2(X)$. From the numerical point of view, an important feature of this system is the symmetry of the operator $P$.

## 4. STOCHASTIC DYNAMIC PROGRAMMING APPROXIMATION

In the previous section, the HDMR tool was introduced. It is based on linear equations determining the optimal decomposition (21), whereas the Bellman equation (7) is highly nonlinear due to the operator of minimization. This fact obstructs the direct use of the HDMR. For that reason, it is necessary to find some linear approximation of the Bellman equation first.

As a mean value of some function has to be higher or equal to its minimum, the following upper estimate holds for all $t \in \{2, \ldots, \tau\}$, all $\sigma_{t-1} \in \Sigma_{t-1}$ and the Bellman function defined in (7)

$$V_{t-1}(\sigma_{t-1}) \leq \frac{1}{|A_t|} \times$$

$$\int_{A_t} \mathcal{E}\left[\, z_t(y_t, a_t) + V_t(S_t(y_t, a_t, \sigma_{t-1})) \,|\, a_t, \sigma_{t-1}, t-1 \,\right] da_t$$

This inequality can be rewritten in a more compact way by introducing a few shortcuts. At first, the following uniform pdf will be useful $f(a_t|\sigma_{t-1}, t-1) \equiv \frac{1}{|A_t|}$. It is a mere shortcut, but it could be also interpreted as the simplest possible optimal strategy predictor. It permits to introduce $Z_t(\sigma)$, the function evaluating expected one-step-ahead loss

$$Z_t(\sigma) \equiv \mathcal{E}\left[\, z_{t+1}(y_{t+1}, a_{t+1}) \,|\, \sigma, t \,\right]$$

Then, introducing the following conditioned pdf

$$f(\sigma_{t+1}|y_{t+1}, a_{t+1}, \sigma_t) \equiv \delta_{\sigma_{t+1}}(S_{t+1}(y_{t+1}, a_{t+1}, \sigma_t))$$

representing a model of statistic dynamic, and using the chain rule, see for instance [5], gives the pdf

$$f(\sigma_{t+1}|\sigma_t, t) \equiv \int_{Y_{t+1}} \int_{A_{t+1}} f(\sigma_{t+1}|y_{t+1}, a_{t+1}, \sigma_t, t) \times$$
$$f(y_{t+1}|a_{t+1}, \sigma_t, t) \times f(a_{t+1}|\sigma_t, t)\, da_{t+1}\, dy_{t+1}$$

Now, the previous inequality can be rewritten as follows

$$V_{t-1}(\sigma_{t-1}) \leq Z_{t-1}(\sigma_{t-1}) + \mathcal{E}\left[\, V_t(\sigma_t) \,|\, \sigma_{t-1}, t-1 \,\right]$$

Thanks to the recursive nature of the Bellman equation, see (7), this inequality spreads over the whole domain of $V$. Considering just the equality part, it turns into a recursive equation for a function $U$, which is an upper bound on the Bellman function

$$U_{t-1}(\sigma_{t-1}) = Z_{t-1}(\sigma_{t-1}) + \mathcal{E}\left[\, U_t(\sigma_t) \,|\, \sigma_{t-1}, t-1 \,\right] \tag{22}$$

It is a linear equation, and therefore it can be solved easier than the exact Bellman equation.

With the knowledge of $U$, the approximated optimal decision at the time step $t \in T$ is $a_t^{opt} \in A_t$ satisfying

$$a_t^{opt} = \operatorname*{argmin}_{a_t \in A_t} \mathcal{E}\left[\, z_t(y_t, a_t) + U_t(S_t(y_t, a_t, \sigma_{t-1})) \,|\, a_t, \sigma_{t-1}, t \,\right]$$

Here, again the shortcut $\sigma_{t-1} \equiv \sigma_{t-1}(d(t-1))$ was used.

### *4.1 HDMR-based Approximation*

The linearity of equation (22), which describes the upper bound on the Bellman function $U$, allows applying the HDMR approximation directly. For all the times $t \in T$, the optimal HDMR decomposition component $\tilde{U}_{t,K}$, $K \in D$,

has to be searched within $H_K(\Sigma_t)$. Firstly, the common index set $M \equiv \{1, \ldots, \mu\}$ is selected obeying the condition

$$\bigcup_{t \in T} \Sigma_t \subset \mathbb{R}^\mu$$

Next, an appropriate set of the decomposition components $D$ is chosen satisfying (12). Its choice fully determines a structure of the following approximation. Some a priori knowledge can be applied here, or all the components can be selected up to the same order. The typical choice is the second order decomposition, i.e., the case when $D$ is chosen as follows

$$D = \{\emptyset\} \cup \{\{m\} | m \in M\} \cup \{\{m, n\} | m, n \in M, m < n\}$$

Finally, it is necessary to prepare the index sets $\mathcal{I}_t$, $t \in T$, see (16). Not only the index set $D$, bud also a geometry of each approximation domain $\Sigma_t$ plays a role here.

Respecting the recursive nature of equation (22), and also the border condition $U_\tau \equiv 0$, it is necessary to start from $t = \tau - 1$, find the collection $\{\tilde{U}_{\tau-1,K}\}_{K \in D}$ determining approximated values of $U$ at the time $t = \tau - 1$, decrease $t$ by one and repeat this procedure until $t = 1$. Inserting (22) into (21) and respecting condition $U_\tau \equiv 0$ the following equation is obtained

$$P_{\tau-1} \star \tilde{U}_{\tau-1} = R_{\tau-1}[Z_{\tau-1}]$$

with $P_{\tau-1}$, resp. $R_{\tau-1}$, defined analogously to (19), resp. (20). Its solution is the collection $\{\tilde{U}_{\tau-1,K}\}_{K \in D}$ fully determining the HDMR approximation of the upper bound on the Bellman function for time $t = \tau - 1$.

Now, knowing $\{\tilde{U}_{t+1,K}\}_{K \in D}$ for some $t + 1 \in T$, the analogous procedure is performed to find $\{\tilde{U}_{t,K}\}_{K \in D}$. It leads to the equation

$$P_t \star \tilde{U}_t = R_t[Z_t + \mathcal{E}[U_{t+1}(\sigma_{t+1}) | \sigma, t]]$$

This equation is the exact equation for the optimal HDMR decomposition components of $\tilde{U}_t$ having only one, but crucial problem. On the right hand side, there occurs the exact value of $U_{t+1}(\sigma_{t+1})$, which is unknown at the moment. To avoid this, again, its HDMR decomposition

$$\tilde{U}_{t+1}(\sigma) \equiv \sum_{K \in D} \tilde{U}_{t+1,K}(\sigma)$$

is substituted instead. This way, the previous equation turns into

$$P_t \star \tilde{U}_t = R_t[Z_t] + Q_{t+1} \star \tilde{U}_{t+1} \tag{23}$$

where

$$Q_{t+1} \star \tilde{U}_{t+1} \equiv R_t \left[ \sum_{K \in D} \mathcal{E}\left[ \tilde{U}_{t+1,K}(\sigma_{t+1}) \,\middle|\, \sigma, t \right] \right]$$

Reminding the definition of $R[g]$, see (20), altogether with the detailed meaning of the "starred" product, see (21), for all $\kappa \in \mathcal{I}_t$, $\lambda \in \mathcal{I}_{t+1}$ and the operator element $Q_{t+1,\kappa,\lambda}$ it holds

$$Q_{t+1,\kappa,\lambda} = \int_{\Sigma_{t+1}} \int_{\Sigma_t} f(\sigma_{t+1}|\sigma, t)\, \delta_\kappa(\sigma)\, d\sigma\, \delta_\lambda(\sigma_{t+1})\, d\sigma_{t+1}$$

The solution of the series of linear systems (23) is equivalent to finding approximative solution $\tilde{U}$ of the upper bound of the Bellman equation (22) using the HDMR technique.

## 5. TOY PROBLEM EXAMPLE

Tossing of an unknown coin is an appropriate example to depict the core of this work. A decision maker plays a hazard game with a (two-sided) coin. Only one side is the winning one. The coin is unfair and pay-off probabilities of its sides are unknown. Also, it is not clear whether the result of tossing depends on the starting orientation of the coin. The only, but crucial knowledge is that the pay-off probabilities are fixed, i.e., the coin is rigid.

The decision-maker's problem is: how to find the best strategy to pick the winning side of the coin? Even though this problem could be formulated so easily, it is a real issue for a longer game horizon as it is hard to balance exploration and exploitation. Winning in the first round does not imply the decision maker should choose the same coin side as it prevents the pay-off probability of the opposite coin side.

Consider the finite decision making horizon of $\tau$ steps. Using the previous notation, $y_t$ represents the observed value (upper side of the coin when it has landed) for each time step and $a_t$ decision (selected coin side before tossing) of a player (decision maker). As the game rules are fixed and even the coin itself is rigid, the range of system input and/or output is still the same. For all $t \in T$, it holds $a_t \in A_t \equiv A = \{0, 1\}$ and similarly $y_t \in Y_t \equiv Y = \{0, 1\}$, where "0" stands for the "Tails" side of the coin and "1" for the "Heads" side.

Note, for computation of the expected value in (4), the knowledge of the outer system model (3) is crucial. It can be composed of two separate probability densities, a parametric system model pdf and a describing internal unknown parameter (pay-off probability of coin sides). For more detailed information, see [2].

Before writing the resulting formulae, it is necessary to introduce a sufficient statistic. Introducing the Kronecker's symbol for $j, k \in \mathbb{N}$, $\delta_{j,k} \equiv 1$ if $j = k$ and $\delta_{j,k} \equiv 0$ otherwise, a sufficient statistic can be identified with a three-dimensional vector $\sigma_t(d(t)) = (\sigma_t^1, \sigma_t^2, \sigma_t^3)$ as follows

$$\sigma_t \equiv \left( \sum_{i=1}^{t} \delta_{y_i,0}\, \delta_{a_i,0}, \sum_{i=1}^{t} \delta_{y_i,0}\, \delta_{a_i,1}, \sum_{i=1}^{t} \delta_{y_i,1}\, \delta_{a_i,0} \right)$$

These values are simply the sums of different game results. For instance, $\sigma_t^1$ equals to the count of the previous game rounds starting with the coin on the "Tails" side (denoted by 0) and landing on the same side. Then, the outer system model is

$$f(0|0, \sigma, t) = \frac{\sigma_t^1 + 1}{\sigma_t^1 + \sigma_t^3 + 2}$$

$$f(1|0, \sigma, t) = \frac{\sigma_t^3 + 1}{\sigma_t^1 + \sigma_t^3 + 2}$$

$$f(0|1, \sigma, t) = \frac{\sigma_t^2 + 1}{t + 2 - \sigma_t^1 - \sigma_t^3}$$

$$f(1|1, \sigma, t) = \frac{t + 1 - \sigma_t^1 - \sigma_t^2 - \sigma_t^3}{t + 2 - \sigma_t^1 - \sigma_t^3}$$

Naturally, for the statistic values it holds $\sigma_t^1 + \sigma_t^2 + \sigma_t^3 \leq t$. This constraint implies these statistic domains $\Sigma_t$, $t \in T$

$$\Sigma_t \equiv \left\{(\sigma_t^1, \sigma_t^2, \sigma_t^3) \in \{0, \ldots, t\}^3 | \, \sigma_t^1 + \sigma_t^2 + \sigma_t^1 \leq t\right\}$$

To completely formalize the problem, let prescribe a form of the statistic updating mapping postulated in (6). In the context of the toy problem, it is time independent

$$S(y, a, \sigma^1, \sigma^2, \sigma^3) \equiv$$
$$\left(\sigma^1 + \delta_{0,y}\, \delta_{0,a}, \sigma^2 + \delta_{0,y}\, \delta_{1,a}, \sigma^3 + \delta_{1,y}\, \delta_{0,a}\right)$$

In the experiments, the coin tossing was simulated with an use of pseudo-random generator simulating a coin with the pay-off probability of the "Heads" side fixed at 60% and the pay-off probability of the "Tails" side sampled from 0% to 100% by a 1% step. At first, the short-horizon experiments were made to compare the results of the different orders of the used HDMR approximation. The index sets are

$$D_1 \equiv \{\emptyset, \{1\}, \{2\}, \{3\}\}$$
$$D_2 \equiv \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}\}$$

On the short game horizon, i.e., $\tau = 10$, each experiment was repeated 5000 times for the various strategies: the exact optimal strategy prepared according to (7) and for both approximated optimal strategies derived according equations (23) using index set $D_1$, resp. $D_2$. The results of these experiments are depicted in Figure 1.

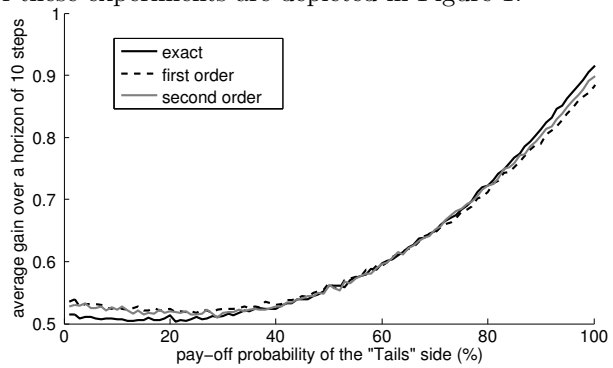


Fig. 1. The average gain of the optimal strategy compared with the gains obtained from approximated strategies based on index sets $D_1$, resp. $D_2$.

Comparing the results of the approximated strategies based on the index set $D_1$ and $D_2$, the first-order approximation driven by $D_1$ seems to be good enough in the context of the toy problem. Therefore, it is used also in the long horizon experiments. To illustrate the power of the newly introduced technique, the game horizon of 200 steps is to be solved. It is no more possible to compare these results of the approximated suboptimal solution with the optimal one. As a basic illustration, results obtained by the "receding horizon" technique are attached. It ran with the receding horizon of 5 steps. Both strategies ran in 100 repetitions, for the results see Figure 2.

## 6. CONCLUSION

The aim of this work was to cope with infeasible memory demands necessary to represent the optimal decision making strategy. The upper bound of the Bellman function was found capable of applying the HDMR approximation easily. To obtain the best possible approximation, the HDMR technique was tuned to work with a general shape of approximation domain. Combining both these approaches, a series of linear systems appears, that implicitly determines the approximated quantity.

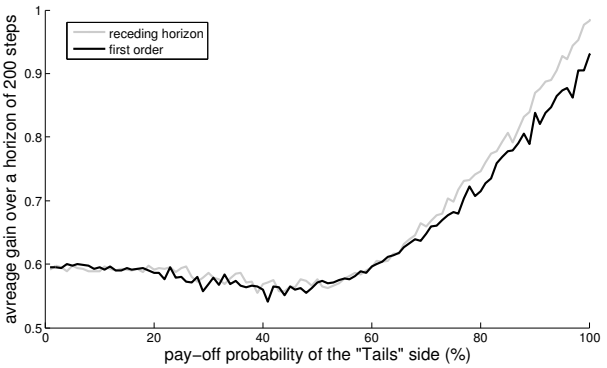


Fig. 2. The average gain of the first order approximation based on index set $D_1$ compared with the average gain obtained from the "receding horizon" approximation.

As illustrated in the example of the tossing of an unknown coin, the correspondence of the results produced by the approximated strategy on the one hand with the optimal results on the other hand was very good. The extended experiments on more complex systems are needed to confirm this observation.

The bottle-neck of this approximation technique is the complicated construction of the central matrices (23). It still needs to pass through the whole solution domain, never-the-less it can be parallelized easily. Also, a promising variant seems to be the recycling of these matrices into a new step of decision making, i.e., introducing a receding-horizon-like concept with a much longer horizon enabled by the use of the HDMR approximation. It is a topic of the future.

## REFERENCES


[1] H.J. Rabitz and O.F. Alis. General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry*, 25:197–233, 1999.

[2] V. Peterka. Bayesian system identification. In P. Eykhoff, editor, *Trends and Progress in System Identification*, pages 239–304. Pergamon Press, Oxford, 1981.

[3] Warren B. Powell. *Approximate Dynamic Programming: solving the curses of dimensionality.* John Wiley & Sons, Inc., Hoboken, New Jersey, 2007.

[4] H. Kushner. *Introduction to Stochastic Control.* Holt, Rinehart and Winston, New York, 1971.

[5] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař. *Optimized Bayesian Dynamic Advising: Theory and Algorithms.* Springer, London, 2005.