Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 58-4 (2024) 676–681

Active Fault Detection Based on Tensor Train Decomposition ⋆

Ivo Punčochář * Ondřej Straka * Petr Tichavský **

* University of West Bohemia, 301 00 Pilsen, Czech Republic (e-mails: ivop@ntis.zcu.cz, straka30@ntis.zcu.cz).
** Institute of Information Theory and Automation, 182 00 Prague, Czech Republic (e-mail: tichavsk@utia.cas.cz)

Abstract: The paper deals with active fault detection of stochastic systems based on tensor train representation of the Bellman function. The faulty and faulty-free behavior of the system is represented using multiple models. The active fault detection problem is treated as an optimal design problem similar to optimal stochastic control. The original problem is reformulated as a perfect state information problem by introducing an information state that contains statistics computed by a state estimator. The Bellman function is computed using the value iteration algorithm over a rectilinear grid set up in the information state space. Within the value iteration algorithm, the Bellman function is represented using the tensor train decomposition, and considerable attention is devoted to designing a rectilinear grid that respects the constraints placed on the elements of the information state.

Keywords: active fault detection, tensor decomposition, value iteration.

1. INTRODUCTION

The main distinctive feature of active fault detection is an auxiliary input signal that excites a monitored system to improve the quality of decisions (Zhang, 1989). This approach can be formulated in stochastic (Kerestecioğlu and Zarrop, 1991; Blackmore and Williams, 2005; Punčochář and Šimandl, 2008), deterministic (Nikoukhah, 1998; Campbell and Nikoukhah, 2004; Raimondo et al., 2016) or combined (Streif et al., 2014) frameworks.

In the stochastic framework, the problem of active fault detection can be formulated as a functional optimization problem. The optimal solution can be found by solving the Bellman functional equation using the value iteration algorithm (Punčochář and Šimandl, 2014). The critical issues are finite-dimensional representations of the Bellman function (Punčochář et al., 2015) and the optimal input signal generator (Král and Punčochář, 2022). If the grid-based approach is used to represent the Bellman function, the memory and computational requirements increase exponentially with the dimension of the information state. In recent years, the tensor train decomposition was successfully used to represent a general function (Tichavský and Phan, 2023) and the Bellman function in optimal control (Gorodetsky et al., 2015).

The paper aims to employ the tensor train decomposition for the active fault detection problem. The main challenge is using a rectilinear grid over the information state that respects the constraints on some of its elements.

The paper is structured as follows: Section 2 provides problem formulation of active fault detection. Section 3 discusses the design of an active fault detector, including the reformulation to a perfect state information problem. Section 4 then describes the tensor train decomposition, grid design for the Bellman function, and the value iteration algorithm. A numerical illustration of the active fault detector using the proposed tensor train decomposition of the Bellman function is presented in Section 5, and concluding remarks are given in Section 6.

2. PROBLEM FORMULATION

The problem of active fault detection is considered for a stochastic system over an infinite time horizon $\mathcal{T} = \{0, 1, \dots\}$. The faults are modeled using a multiple model framework where the fault-free and each faulty behavior of the monitored system is described by its own model. Thus, it is assumed that the monitored system can be described at each time step $k \in \mathcal{T}$ be the following conditional probability density functions (PDFs)

$$p_{\mathbf{s}_{k+1}|\mathbf{s}_k,\mathbf{u}_k}(\mathbf{s}_{k+1}|\mathbf{s}_k, \mathbf{u}_k), \tag{1a}$$

$$p_{\mathbf{y}_k|\mathbf{s}_k}(\mathbf{y}_k|\mathbf{s}_k), \tag{1b}$$

where $\mathbf{s}_k^{\mathrm{T}} = [\mathbf{x}_k^{\mathrm{T}}, \mu_k] \in \mathcal{S} = \mathcal{X} \times \mathcal{M}$ is a state, $\mathbf{x}_k \in \mathcal{X} \subseteq \mathbb{R}^{n_x}$ is a continuous-valued part of the state, $\mu_k \in \mathcal{M} = \{1, 2 \dots, M\}$ is a discrete-valued part of the state that represents the index of the model ($\mu_k \in \{1\}$ for the faulty-free model and $\mu_k \in \{2, 3, \dots, M\}$ for faulty ones) being in effect at the time step k, $\mathbf{u}_k \in \mathcal{U} \subset \mathbb{R}^{n_u}$ is an auxiliary input that can be used for active fault detection, and $\mathbf{y}_k \in \mathcal{Y} \subseteq \mathbb{R}^{n_y}$ is an output. The conditional PDF (1a) can be factorized as

$$p_{\mathbf{s}_{k+1}|\mathbf{s}_k,\mathbf{u}_k}(\mathbf{s}_{k+1}|\mathbf{s}_k, \mathbf{u}_k) = p_{\mu_{k+1}|\mathbf{s}_k,\mathbf{u}_k}(\mu_{k+1}|\mathbf{s}_k, \mathbf{u}_k) \times p_{\mathbf{x}_{k+1}|\mathbf{s}_k,\mu_{k+1},\mathbf{u}_k}(\mathbf{x}_{k+1}|\mathbf{s}_k, \mu_{k+1}, \mathbf{u}_k). \tag{2}$$

We assume that the conditional PDFs $p_{\mathbf{x}_{k+1}|\mathbf{s}_k,\mu_{k+1},\mathbf{u}_k}$ and $p_{\mu_{k+1}|\mathbf{s}_k,\mathbf{u}_k}$ satisfy

⋆ The work was supported by the Czech Science Foundation under grant 22-11101S.

10.1016/j.ifacol.2024.07.297

$$p_{\mathbf{x}_{k+1}|\mathbf{s}_k,\mu_{k+1},\mathbf{u}_k}(\mathbf{x}_{k+1}|\mathbf{s}_k,\mu_{k+1},\mathbf{u}_k) = p_{\mathbf{x}_{k+1}|\mathbf{s}_k,\mathbf{u}_k}(\mathbf{x}_{k+1}|\mathbf{s}_k,\mathbf{u}_k), \quad (3)$$

$$p_{\mu_{k+1}|\mathbf{s}_k,\mathbf{u}_k}(\mu_{k+1}|\mathbf{s}_k,\mathbf{u}_k) = p_{\mu_{k+1}|\mu_k}(\mu_{k+1}|\mu_k), \quad (4)$$

where $p_{\mathbf{x}_{k+1}|\mathbf{s}_k,\mathbf{u}_k}$ and $P_{\mu_{k+1}|\mu_k}$ are known. The conditional PDF (1b) is also assumed to be known. The initial state $\mathbf{s}_0$ is described by the known PDF

$$p(\mathbf{s}_0) = p_{\mathbf{x}_0}(\mathbf{x}_0)P_{\mu_0}(\mu_0). \quad (5)$$

The aim is to design an active fault detector that is described at each time step as

$$\begin{bmatrix} d_k \\ \mathbf{u}_k \end{bmatrix} = \begin{bmatrix} \sigma_k(\mathbf{z}_k) \\ \boldsymbol{\gamma}_k(\mathbf{z}_k) \end{bmatrix}, \quad (6)$$

where $d_k \in \mathcal{M}$ is a decision and $\mathbf{z}_k = [\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}] \in \mathcal{Z}_k = \mathcal{Y}^{k+1} \times \mathcal{U}^k$ is an information vector [1] that contains all random variables whose realizations were observed up to the current time step k. The function $\sigma_k : \mathcal{Z}_k \mapsto \mathcal{M}$ describes a passive decision generator and the function $\boldsymbol{\gamma}_k : \mathcal{Z}_k \mapsto \mathcal{U}$ describes an input signal generator.

The sequence of functions $\sigma_{0:\infty}$ and $\boldsymbol{\gamma}_{0:\infty}$ should be designed such that the decision d_k is close to the model index μ_k at each time step of the infinite time horizon. Therefore, the functions σ_k and $\boldsymbol{\gamma}_k$ are designed to minimize the following discounted additive criterion

$$J = \lim_{F\to\infty} \mathrm{E}\left\{ \sum_{k=0}^{F} \eta^k L(\mu_k, d_k) \right\}, \quad (7)$$

where $\mathrm{E}\{\cdot\}$ is the expectation operator, $\eta \in (0,1)$ is a discount factor, and $L : \mathcal{M} \times \mathcal{M} \mapsto \mathbb{R}^+$ is a detection cost function selected by a user to assign a cost for selecting decision d_k while true model is μ_k. A reasonable detection cost function should satisfy

$$L(i,i) < L(i,j) \quad (8)$$

for each $i \in \mathcal{M}$ and $j \in \mathcal{M}$, $j \neq i$.

3. DESIGN OF ACTIVE FAULT DETECTOR

This section presents the design of an active fault detector. As the state $\mathbf{s}_k$ is observed only through the noisy measurement $\mathbf{y}_k$, the problem is first reformulated as a perfect state information problem that uses an information state consisting of statistics computed by a state estimator. The general solution to this reformulated problem is given in terms of the Bellman functional equation. Then, the tensor train decomposition is presented as an efficient approximation of a Bellman function over a rectilinear grid. As the information state includes covariance matrices and a probability vector, a rectilinear grid cannot be defined over the space of the information states. This issue is addressed by an auxiliary grid that serves as an intermediate between the information state space and the space of discrete indices of a tensor representing the Bellman function.

3.1 Perfect State Information Problem

Since the state $\mathbf{s}_k$ of the system (1) is not directly available, the problem of active fault detector design belongs to the class of imperfect state information problems. This paper addresses the issue by reformulating it as a perfect state information problem that uses the conditional PDF $p_{\mathbf{s}_k|\mathbf{z}_k}$ as an infinite-dimensional state (Striebel, 1965; Bertsekas, 2000). In principle, this PDF can be obtained recursively by the optimal state estimator that performs the measurement update using the Bayes functional relation

$$p_{\mathbf{s}_{k+1}|\mathbf{z}_{k+1}}(\mathbf{s}_{k+1}|\mathbf{z}_{k+1}) = \frac{p_{\mathbf{y}_{k+1}|\mathbf{s}_{k+1}}(\mathbf{y}_{k+1}|\mathbf{s}_{k+1})}{p_{\mathbf{y}_{k+1}|\mathbf{z}_k,\mathbf{u}_k}(\mathbf{y}_{k+1}|\mathbf{z}_k,\mathbf{u}_k)} \times p_{\mathbf{s}_{k+1}|\mathbf{z}_k,\mathbf{u}_k}(\mathbf{s}_{k+1}|\mathbf{z}_k,\mathbf{u}_k) \quad (9)$$

and the time update using the Chapman-Kolmogorov functional relation

$$p_{\mathbf{s}_{k+1}|\mathbf{z}_k,\mathbf{u}_k}(\mathbf{s}_{k+1}|\mathbf{z}_k,\mathbf{u}_k) = \int_{\mathcal{S}} p_{\mathbf{s}_{k+1}|\mathbf{s}_k,\mathbf{u}_k}(\mathbf{s}_{k+1}|\mathbf{s}_k,\mathbf{u}_k) p_{\mathbf{s}_k|\mathbf{z}_k}(\mathbf{s}_k|\mathbf{z}_k)\,\mathrm{d}\,\mathbf{s}_k. \quad (10)$$

Combining these two updates, the perfect state information model can be written as

$$p_{\mathbf{s}_{k+1}|\mathbf{z}_{k+1}} = \varphi\left(p_{\mathbf{s}_k|\mathbf{z}_k}, \mathbf{u}_k, \mathbf{y}_{k+1}\right), \quad (11)$$

where $\varphi : \mathcal{P} \times \mathcal{U} \times \mathcal{Y} \mapsto \mathcal{P}$ is a functional mapping that transforms the conditional PDF $p_{\mathbf{s}_k|\mathbf{z}_k}$, input $\mathbf{u}_k$, and future output $\mathbf{y}_{k+1}$ to the conditional PDF $p_{\mathbf{s}_{k+1}|\mathbf{z}_{k+1}}$. The future output $\mathbf{y}_{k+1}$ is treated as a random disturbance described by the predictive conditional PDF

$$p_{\mathbf{y}_{k+1}|\mathbf{z}_k,\mathbf{u}_k} = \varphi_{\mathbf{y}}(p_{\mathbf{s}_k|\mathbf{z}_k}, \mathbf{u}_k), \quad (12)$$

where $\varphi_{\mathbf{y}} : \mathcal{P} \times \mathcal{U} \mapsto \mathcal{P}$ is a functional mapping that transforms the conditional PDF $p_{\mathbf{s}_k|\mathbf{z}_k}$ and input $\mathbf{u}_k$ into conditional PDF $p_{\mathbf{y}_{k+1}|\mathbf{z}_k,\mathbf{u}_k}$. Note that $\mathcal{P}$ denotes the set of all possible conditional PDFs. The functional mapping (11) includes the description of the monitored system through $p_{\mathbf{y}_{k+1}|\mathbf{s}_{k+1}}$ and $p_{\mathbf{s}_{k+1}|\mathbf{s}_k,\mathbf{u}_k}$ as its parameters.

The generalized pseudo Bayesian (GPB)2 state estimator together with the information state $\boldsymbol{\xi}_k$ can be used to build up an approximation to (11)

$$\boldsymbol{\xi}_{k+1} = \boldsymbol{\phi}\left(\boldsymbol{\xi}_k, \mathbf{u}_k, \mathbf{y}_{k+1}\right), \quad (13)$$

where $\boldsymbol{\phi} : \Xi \times \mathcal{U} \times \mathcal{Y} \mapsto \Xi$ is a mapping that represents the dynamics of an approximate perfect state information model and the future output $\mathbf{y}_{k+1}$ is treated as a random disturbance with the conditional PDF $p_{\mathbf{y}_{k+1}|\mathbf{z}_k,\mathbf{u}_k}$ that is approximated by a PDF with moments being computed using the state estimation algorithm from $\boldsymbol{\xi}_k$ and $\mathbf{u}_k$. The active fault detector for the perfect state information problem is time-invariant and is described as

$$\begin{bmatrix} d_k \\ \mathbf{u}_k \end{bmatrix} = \begin{bmatrix} \bar{\sigma}(\boldsymbol{\xi}_k) \\ \bar{\gamma}(\boldsymbol{\xi}_k) \end{bmatrix}. \quad (14)$$

Finally, the equivalent form of the additive discounted criterion (7) is

$$J = \lim_{F\to\infty} \mathrm{E}\left\{ \sum_{k=0}^{F} \eta^k \bar{L}(\boldsymbol{\xi}_k, d_k) \right\}, \quad (15)$$

where $\bar{L} : \Xi \times \mathcal{M} \mapsto \mathbb{R}^+$ is a detection cost function derived from the original detection cost function L. See e.g., (Punčochář et al., 2015) for details.

3.2 Bellman Equation for Active Fault Detector

The active fault detector optimal for the approximate perfect state information problem (13), (14), and (15) can

[1] The notation $\mathbf{y}_{i:j}$ represents the sequence of variables $\mathbf{y}_k$ from the time step i up to the time step j and $\mathcal{U}^n = \mathcal{U} \times \mathcal{U} \times \ldots \times \mathcal{U}$ denotes the n-ary Cartesian power of the set $\mathcal{U}$.

be determined by finding the value function $V : \Xi \mapsto \mathbb{R}^+$ that solves the following Bellman functional equation

$$V(\boldsymbol{\xi}) = \min_{\substack{d \in \mathcal{M} \\ \mathbf{u} \in \mathcal{U}}} \mathrm{E}\left\{\bar{L}(\boldsymbol{\xi}, d) + \eta V(\boldsymbol{\phi}(\boldsymbol{\xi}, \mathbf{u}, \mathbf{y}'))|\boldsymbol{\xi}, \mathbf{u}\right\}, \quad (16)$$

where $\mathrm{E}\{\cdot|\cdot\}$ is the conditional expectation operator. The conditional PDF $p_{\mathbf{y}'|\boldsymbol{\xi},\mathbf{u}}$ used for the expectation calculation is defined as

$$p_{\mathbf{y}'|\boldsymbol{\xi},\mathbf{u}}(\mathbf{y}'|\boldsymbol{\xi}, \mathbf{u}) = p_{\mathbf{y}_{k+1}|\mathbf{z}_k,\mathbf{u}_k}(\mathbf{y}'|\boldsymbol{\xi}, \mathbf{u}). \quad (17)$$

Once the Bellman function V is found, the decision and input signal generators are given as

$$d_k = \bar{\sigma}(\boldsymbol{\xi}_k) = \arg\min_{d \in \mathcal{M}} \mathrm{E}\left\{\bar{L}(\boldsymbol{\xi}_k, d)|\boldsymbol{\xi}_k\right\}, \quad (18)$$

$$\mathbf{u}_k = \bar{\gamma}(\boldsymbol{\xi}_k) = \arg\min_{\mathbf{u} \in \mathcal{U}} \mathrm{E}\left\{V(\boldsymbol{\phi}(\boldsymbol{\xi}_k, \mathbf{u}, \mathbf{y}'))|\boldsymbol{\xi}_k, \mathbf{u}\right\}. \quad (19)$$

The Bellman functional equation (16) can be solved approximately using iterative methods such as the value iteration, policy iteration, or policy search that successively compute an approximation to the Bellman function (Buşoniu et al., 2010). In recent years, the tensor train decomposition approach was successfully used to approximate the Bellman function in several optimal control problems, see e.g., Gorodetsky et al. (2015).

4. TENSOR TRAIN DECOMPOSITION OF BELLMAN FUNCTION

4.1 Description of the tensor train decomposition

A survey of several tensor decompositions can be found, e.g., in (Kolda and Bader, 2009), and the tensor train decomposition used in this paper is elaborated in (Oseledets and Tyrtyshnikov, 2010). A real d-dimensional tensor $\mathbf{F}$ is a function defined over a discrete set

$$\mathbf{F} : \mathcal{I}_1 \times \mathcal{I}_2 \times \ldots \times \mathcal{I}_d \mapsto \mathbb{R}, \quad (20)$$

where $\mathcal{I}_j = \{1, 2, \ldots, n_j\}$ is a set of discrete indices and $n_j \in \mathbb{N}$ is the cardinality of the set $\mathcal{I}_j$. Since the domain of $\mathbf{F}$ is a finite set, we can represent the tensor $\mathbf{F}$ as a d-dimensional array (i.e., a function stored in its tabular form) [2] and write $\mathbf{F} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$. Since the total number of elements in the tensor $\mathbf{F}$ is

$$n = \prod_{j=1}^{d} n_j, \quad (21)$$

it is amenable for storing and computing only if the dimension d is low. This issue can be addressed using a tensor train decomposition (TTD) that exists for any tensor $\mathbf{F}$. This decomposition assumes that an element of the tensor $\mathbf{F}$ can be written as

$$\mathbf{F}(i_1, i_2, \ldots, i_d) = \sum_{\alpha_0=1}^{r_0} \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_d=1}^{r_d} \mathbf{F}_1(\alpha_0, i_1, \alpha_1)\mathbf{F}_2(\alpha_1, i_2, \alpha_2) \cdots \mathbf{F}_d(\alpha_{d-1}, i_d, \alpha_d), \quad (22)$$

where $\mathbf{F}_j : \mathcal{R}_{j-1} \times \mathcal{I}_j \times \mathcal{R}_j \mapsto \mathbb{R}$ is a real three-dimensional tensor, $\mathcal{R}_j = \{1, 2, \ldots, r_j\}$ is an auxiliary discrete set, and $r_j \in \mathbb{N}$ is a rank. The ranks r_0 and r_d can be selected as $r_0 = r_d = 1$ without loss of generality. The total number of elements in the TTD is

$$n_{\text{ttd}} = \sum_{j=1}^{d} r_{j-1} r_j n_j. \quad (23)$$

It suggests that a significant saving in memory requirements can be achieved if the ranks are not too high. Since the exact TTD of a given tensor might require high ranks, a low rank TTD $\hat{\mathbf{F}}$ that approximates the original tensor $\mathbf{F}$ is of interest. If the whole tensor $\mathbf{F}$ is available [3], its low-rank approximation can be found using numerically stable algorithms that build upon the singular value decompositions (SVDs) of unfolding matrices of the tensor $\mathbf{F}$ (Oseledets and Tyrtyshnikov, 2010). However, in most cases of interest, the tensor $\mathbf{F}$ is too large for these algorithms to be used. In such a case, the algorithms for finding a low rank TTD are based on the skeleton decomposition of unfolding matrices (Savostyanov and Oseledets, 2011). The algorithm presented in the paper (Savostyanov and Oseledets, 2011) can be used to find the approximation to the Bellman function at each iteration of the value iteration algorithm.

4.2 Grid design

Although the grid design is affected by several factors (e.g., grid extent and density), only one particular factor related to the constraint imposed by the covariance matrices and probability vector is addressed in this section. If the continuous-valued part of the state $\mathbf{x}_k$ has a dimension greater than one ($n_x > 1$) or the number of models is greater than two ($M > 1$), the set of admissible information states Ξ cannot be written as the Cartesian product of intervals due to restriction on the covariance matrix and the probability vector embedded in sets Ξ^{c} and Ξ^{p}, respectively. Therefore, it is impossible to define a rectilinear grid that would reasonably fill the whole set Ξ. This issue can be addressed by defining a rectilinear grid over a superset of Ξ and removing the inadmissible grid points from further computation. However, it is not clear if this approach is a viable option for tensor train algorithms that work with fibers of a tensor (i.e., values of tensor for all arguments fixed at all dimensions except for one). Therefore, the issue is addressed in this paper at the level of the grid design.

To facilitate further discussion, it is necessary to formalize the connection between the integer indices of the tensor and the grid points of a rectilinear grid. Let us assume a vector of indices $\mathbf{i}^{\mathrm{T}} = [i_1, i_2, \ldots, i_{n_\xi}] \in \mathcal{I} = \mathcal{I}_1 \times \mathcal{I}_2 \times \ldots \times \mathcal{I}_{n_\xi}$ and a rectilinear grid $\mathcal{G}$ over a superset of Ξ that is defined using the Cartesian product as

$$\mathcal{G} = \mathcal{G}_1 \times \mathcal{G}_2 \times \ldots \times \mathcal{G}_{n_\xi}, \quad (24)$$

where $\mathcal{G}_j = \left\{g_j^{(1)}, g_j^{(2)}, \ldots, g_j^{(n_j)}\right\}$ is a discrete ordered set for the j-th dimension and $g_j^{(m)} \in \mathbb{R}$. Then the relationship between a grid point $\mathbf{g} \in \mathcal{G}$ and vector of indices $\mathbf{i} \in \mathcal{I}$ can be written element-wise as

$$\mathbf{g}_j = \mathbf{h}_j(i_j), \quad (25)$$

where $\mathbf{h} : \mathcal{I} \mapsto \mathcal{G}$ is a bijective function that is naturally defined by the assignment $g_j^{(m)} = \mathbf{h}_j(m)$ for each $m \in \{1, \ldots, n_i\}$ and all $j \in \{1, \ldots, n_\xi\}$. Now, according

[2] These two interpretations will be used interchangeably without using different notation.

[3] It means that the tensor is small enough to be stored in memory and is amenable for computation.

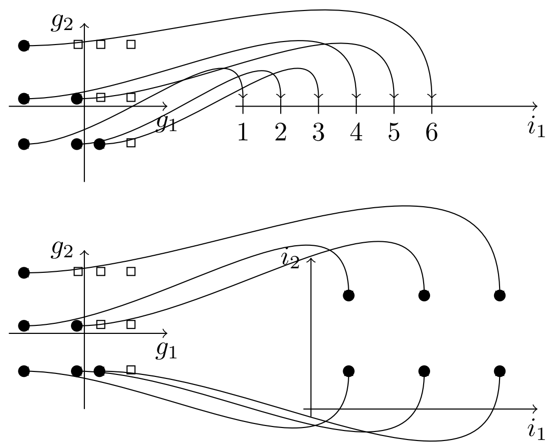

Fig. 1. Illustration of two possible nonlinear bijective functions that map grid points and indices of a tensor.

to the previous discussion, some of the grid points are not admissible as they lie outside the set Ξ. Thus, we must redefine the bijective function $\mathbf{h}$ to avoid such instances. One approach follows the idea of tensor reshaping used in multiresolution analysis (Kazeev and Oseledets, 2013). Instead of the original index set $\mathcal{I}$ an alternative index set $\mathcal{I}_\mathrm{m}$ with a possibly different dimension is selected and an alternative function $\mathbf{h}_\mathrm{m} : \mathcal{I}_\mathrm{m} \mapsto \mathcal{G} \cap \Xi$ is established. The underlying idea is illustrated in Fig. 1 for a two-dimensional rectilinear grid $\mathcal{G}$ by means of two different index sets $\mathcal{I}_\mathrm{m}$ and corresponding nonlinear functions $\mathbf{h}_\mathrm{m}$. The circles and squares represent the admissible and inadmissible grid points of $\mathcal{G}$, respectively. The upper part of the figure shows a mapping between a two-dimensional grid and a one-dimensional index. The lower part of the figure illustrates another mapping between a two-dimensional grid and two-dimensional indices. This approach gets more complex in a higher-dimensional case. Also, note that a possible structure existing in the original space can be lost. Therefore, this approach will not be pursued further in this paper.

Another approach is to introduce an intermediate rectilinear grid $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \ldots \times \mathcal{F}_{n_\xi}$, where $\mathcal{F}_j = \{f_j^{(1)}, f_j^{(2)}, \ldots, f_j^{(n_i)}\}$ and split the bijective function $\mathbf{h}$ into two compound bijective functions as

$$\mathbf{g} = \mathbf{h}_2\left(\mathbf{h}_1(\mathbf{i})\right), \tag{26}$$

where $\mathbf{h}_1 : \mathcal{I} \mapsto \mathcal{F}$ is a bijective function that maps the index set $\mathcal{I}$ to the intermediate rectilinear grid $\mathcal{F}$ and $\mathbf{h}_2 : \mathcal{F} \mapsto \mathcal{G} \subset \Xi$ is a bijective function that maps the intermediate rectilinear grid $\mathcal{F}$ to a grid $\mathcal{G}$ that might not be rectilinear. The function $\mathbf{h}_1$ is defined similarly to $\mathbf{h}$ by the assignment $f_j^{(m)} = \mathbf{h}_{1,j}(m)$. The definition of the bijective function $\mathbf{h}_2$ must consider the constraint on the admissible grid points in the set $\mathcal{G}$. Let us denote the elements of the function $\mathbf{h}_2$ that pertain to a single covariance matrix as $\mathbf{h}_2^\mathrm{c}$ and the elements representing the probability vector as $\mathbf{h}_2^\mathrm{p}$. The following two subsections discuss the selection of these two functions.

Covariance matrix The function $\mathbf{h}_2^\mathrm{c}$ is specified using the spherical parametrization of positive definite covariance matrices (Pinheiro and Bates, 1996). The Cholesky decomposition of a positive definite covariance matrix $\mathbf{P} \in \mathbb{R}^{n_n \times n_n}$ is

$$\mathbf{P} = \mathbf{L}^\mathrm{T}\mathbf{L}, \tag{27}$$

where $\mathbf{L} \in \mathbb{R}^{n_x \times n_x}$ is an upper triangular matrix. The matrix can be parameterized as

$$\mathbf{L} = \begin{bmatrix} \ell_1 & \ell_2\cos(\theta_{1,2}) & \ell_3\cos(\theta_{1,3}) & \ldots \\ 0 & \ell_2\sin(\theta_{1,2}) & \ell_3\sin(\theta_{1,3})\cos(\theta_{2,3}) & \ldots \\ 0 & 0 & \ell_3\sin(\theta_{1,3})\sin(\theta_{2,3}) & \ldots \\ \vdots & \vdots & 0 & \ldots \\ 0 & 0 & 0 & \ldots \end{bmatrix} \tag{28}$$

and the first i elements of the i-th column of the matrix for $i = 2, \ldots, n_x$ read

$$\begin{bmatrix} \mathbf{L}_{1,i} \\ \mathbf{L}_{2,i} \\ \mathbf{L}_{3,i} \\ \vdots \\ \mathbf{L}_{i,i} \end{bmatrix} = \begin{bmatrix} \ell_i\cos(\theta_{1,i}) \\ \ell_i\sin(\theta_{1,i})\cos(\theta_{2,i}) \\ \ell_i\sin(\theta_{1,i})\sin(\theta_{2,i})\cos(\theta_{3,i}) \\ \vdots \\ \ell_i\sin(\theta_{1,i})\sin(\theta_{2,i})\cdots\sin(\theta_{i-1,i}) \end{bmatrix}, \tag{29}$$

where $\ell_i \in \mathbb{R}^{++}$ is a non-zero norm of the i-th column of $\mathbf{L}$ and $\theta_{j,i} \in [0, \pi]$ for $j = 1, \ldots, i-1$ are angles of the spherical parametrization. From the parametrization (29), the following convenient relationship between the diagonal elements of the covariance matrix $\mathbf{P}$ and parameters ℓ_i follows

$$\mathbf{P}_{ii} = \ell_i^2. \tag{30}$$

Let us denote the part of the intermediate rectilinear grid $\mathcal{F}$ that pertains to the covariance matrix $\mathbf{P}_{k|k}(i)$ as $\mathcal{F}^{\mathrm{c},i} \subset \mathbb{R}^{n_x(n_x+1)/2}$. It is specified as a Cartesian product of discrete sets $\mathcal{L}_i \subset \mathbb{R}^{++}$ and $\Theta_{j,i} \subset [0, \pi]$ for the parameters ℓ_i, and $\theta_{1,i}$ to $\theta_{i-1,i}$ for all $i = 1, \ldots, n$, respectively.

The inverse function $\mathbf{h}_2^{-1}$ is given as follows. For a given positive definite covariance matrix $\mathbf{P}$, the Cholesky decomposition is found. For the i-th column of the Cholesky factor $\mathbf{L}$, the parameters ℓ_i, and $\theta_{1,i}$ to $\theta_{i-1,i}$ are computed as follows. The parameter ℓ_i is given as

$$\ell_i = \|\mathbf{L}_{1:i,i}\|_2. \tag{31}$$

Then, the column is normalized

$$\mathbf{L}_{1:i,i} := \frac{\mathbf{L}_{1:i,i}}{\ell_i} \tag{32}$$

and the angles can be computed successively as

$$\theta_{j,i} = \arccos\left(\mathbf{L}_{j,i}\right), \qquad \mathbf{L}_{j+1:i,i} := \frac{\mathbf{L}_{j+1:i,i}}{\sin(\theta_{j,i})}, \tag{33}$$

where $j = 1, 2, \ldots, i-1$.

Probability vector The elements of the function $\mathbf{h}_2$ that correspond to the probability vector $\boldsymbol{\pi} \in \Pi$ are specified in the following way. A commonly used bijective function that maps the hypercube $(0,1)^{M-1}$ to the interior of the probability simplex (Leger, 2023) is

$$\boldsymbol{\pi}_i = \mathbf{t}_i \prod_{j=i+1}^{M-1} (1 - \mathbf{t}_j), \tag{34}$$

where $\mathbf{t} \in (0,1)^{M-1}$. If this function is extended to the closure of the hypercube $(0,1)^{M-1}$, it also maps to the boundary of the probability simplex, but it fails to

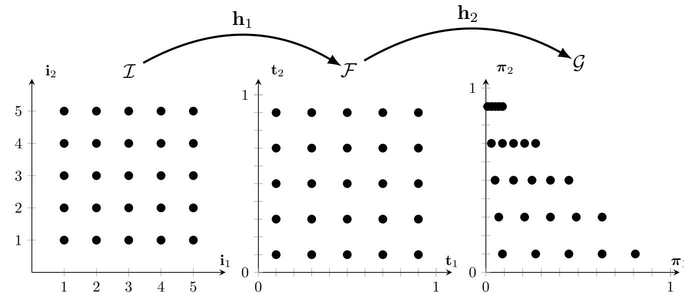

Fig. 2. Illustration of two compound bijective functions that map indices of a tensor and grid points.

be bijective. If the conditional probabilities $P(\mu_k|\mathbf{z}_k)$ do not converge to the boundary values of the probability simplex, it seems reasonable to limit the grid points only to the interior of the probability simplex. Then the inverse function $\mathbf{h}_2^{-1}$ is given as

$$\mathbf{t}_i = \frac{\boldsymbol{\pi}_i}{1 - \sum_{j=i+1}^{M-1} \boldsymbol{\pi}_j}, \tag{35}$$

where $\boldsymbol{\pi}$ is in the interior of Ξ^{P}.

An illustration of the grid for the index set $\mathcal{I}$, intermediate rectilinear grid $\mathcal{F}$, and the grid with admissible points $\mathcal{G}$ for a probability vector in a two-dimensional simplex is shown in Fig. 2.

4.3 Tensor Train Value Iteration

When the TTD is used to represent the Bellman function $V^{(i)}$ in the value iteration algorithm, the single iterate is described as

$$V^{(i+1)}(\mathbf{i}) = \min_{\mathbf{u}\in\mathcal{U}} \mathrm{E}\left\{\bar{L}\left(\mathbf{h}_2(\mathbf{h}_1(\mathbf{i})), \mathbf{u}\right) + \eta V^{(i)}\left(\mathbf{h}_1^{-1}(\mathbf{h}_2^{-1}(\boldsymbol{\phi}(\mathbf{h}_2(\mathbf{h}_1(\mathbf{i})), \mathbf{u}, \mathbf{y})|\mathbf{i}, \mathbf{u}))\right)\right\}. \tag{36}$$

The iteration starts with a zero Bellman function $V^{(0)}$ in the tensor train format. At each iteration step, the right-hand side of (36) can be evaluated for any selected index $\mathbf{i}$. The TT-RC method (Savostyanov and Oseledets, 2011) that is implemented in the TT-Toolbox is used to construct the TTD of the right-hand side and is stored in $V^{(i+1)}$. The iterations are performed until convergence criteria are satisfied.

5. NUMERICAL EXAMPLE

A preliminary evaluation of the proposed approach is performed through a simplified numerical example of a linear Gaussian second-order system with a measurable state $\mathbf{x}_k$ and two models. The first model represents fault-free behavior, and the second represents faulty behavior. Although this simplified model does not require the proposed grid design for the covariance matrices, it is suitable for an initial evaluation. The PDFs describing the system are given as

$$p_{\mathbf{x}_{k+1}|\mathbf{s}_k,\mathbf{u}_k}(\mathbf{x}_{k+1}|\mathbf{s}_k, \mathbf{u}_k) = \mathcal{N}\{\mathbf{x}_{k+1} : \mathbf{A}_{\mu_k}\mathbf{x}_k + \mathbf{B}_{\mu_k}u_k, \mathbf{Q}\},$$
$$p_{\mathbf{y}_k|\mathbf{s}_k}(\mathbf{y}_k|\mathbf{s}_k) = \delta(\mathbf{y}_k - \mathbf{x}_k),$$

where $\mathcal{N}\{\mathbf{x} : \mathbf{m}, \mathbf{P}\}$ denotes the Gaussian distribution of the random variable $\mathbf{x}$ with the mean value $\mathbf{m}$ and the covariance matrix $\mathbf{P}$, and δ denotes the Dirac delta function. The matrices of the system are

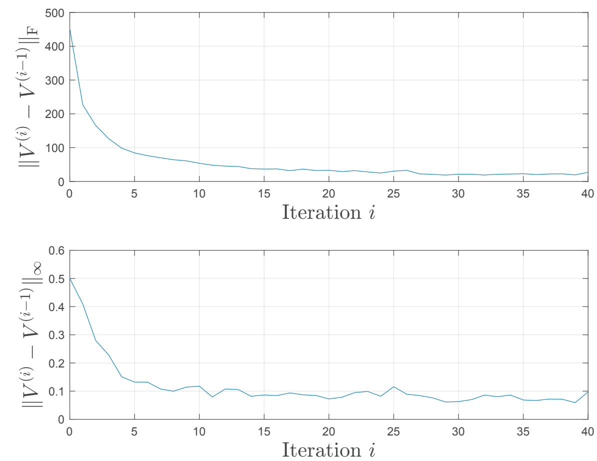

Fig. 3. The convergence of the value iteration algorithm.

$$\mathbf{A}_1 = \begin{bmatrix} 0.956 & 0.085 \\ -0.834 & 0.7 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} 0.957 & 0.083 \\ -0.814 & 0.67 \end{bmatrix}, \tag{37}$$

$$\mathbf{B}_1 = \begin{bmatrix} 0.002 \\ 0.042 \end{bmatrix}, \mathbf{B}_2 = \begin{bmatrix} 0.002 \\ 0.041 \end{bmatrix}, \mathbf{Q} = 9 \times 10^{-4}\mathbf{I}_2. \tag{38}$$

The transition probability is

$$P(\mu_{k+1}|\mu_k) = \begin{bmatrix} 0.95 & 0.02 \\ 0.05 & 0.98 \end{bmatrix}. \tag{39}$$

The initial state is described as $p_{\mathbf{x}_0}(\mathbf{x}_0) = \mathcal{N}\{\mathbf{x}_0 : \mathbf{0}_{2\times1}, 0.002\,\mathbf{I}_2\}$ and $P(\mu_0 = 1) = 1$. The set of admissible inputs is selected as $\mathcal{U} = \{-10, 0, 10\}$, the discount factor is $\eta = 0.9$, and the detection cost function is

$$L(\mu_k, d_k) = \begin{cases} 1 & \text{if } \mu_k \neq d_k, \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

Since the continuous-valued state is directly available, the information state $\boldsymbol{\xi}_k \in \mathbb{R}^3$ is selected as

$$\boldsymbol{\xi}_k^{\mathrm{T}} = \left[\mathbf{x}_k^{\mathrm{T}}\ P(\mu_k = 1|\mathbf{z}_k)\right] \tag{41}$$

and a simpler state estimator is used to compute $P(\mu_k = 1|\mathbf{z}_k)$. The grid is selected as $\mathcal{G} = \{-1 : 0.005 : 1\} \times \{-3 : 0.1 : 3\} \times \{0 : 0.01 : 1\}$ and it contains 2 470 561 points. The mean value in the Bellman equation (36) is computed approximately using the unscented transform (Julier et al., 2000). The TTD representation of the Bellman function is found in the fixed number of 40 iterations of the value iteration algorithm. During the value iteration, the ranks were around $r_1 = 6$ and $r_2 = 4$, which translates approximately to 4274 elements of the TTD. It is considerable less then the number of grid points. The convergence of the algorithm expressed as $\|V^{(i)} - V^{(i-1)}\|_{\mathrm{F}}$ is depicted in Fig. 3. Since the dimension of the information state space is small and full tensor can be constructed, it is possible to compute the max norm $\|V^{(i)} - V^{(i-1)}\|_\infty$ that is also given in Fig. 3. These norms are not monotonic, probably due to the random selection of fibers performed within the TT-RC algorithm. One particular realization of the input, state, and decision trajectories is given in Fig. 4.

The figures show that the TTD approximate representation of the Bellman function can provide quality decisions

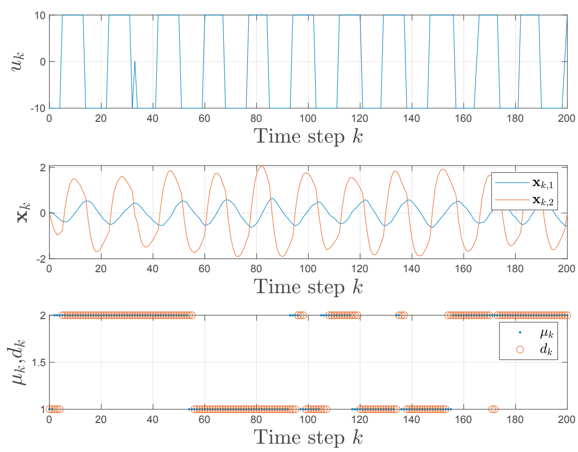

Fig. 4. A sample trajectory of input u_k, state $\mathbf{x}_k$, μ_k, and decision d_k.

even though its number of elements is small compared to the number of elements if no approximation is involved for the Bellman function representation. Such savings can be crucial when dealing with higher dimensional problems.

6. CONCLUSION

The paper dealt with the TTD representation of the Bellman function for the active fault detection problem. The main challenge addressed was the design of a grid that can be used in connection with the tensor train representation. The use of TTD for the Bellman function representation is crucial for high-dimensional problems for which the full tensor representation of the Bellman function cannot be constructed due to enormous memory requirements.

REFERENCES

Bertsekas, D.P. (2000). *Dynamic Programming and Optimal Control: Volume I*, volume I. Athena Scientific, Belmont, MA, USA, 2 edition.

Blackmore, L. and Williams, B.C. (2005). Finite horizon control design for optimal model discrimination. In *Proceedings of the 44th IEEE Conference on Decision and Control, and European Control Conference*, 3795–3802. Seville, Spain.

Buşoniu, L., Babuška, R., Schutter, B.D., and Ernst, D. (2010). *Reinforcement Learning and Dynamic Programming Using Function Approximators.* CRC Press, Boca Raton, FL, USA.

Campbell, S.L. and Nikoukhah, R. (2004). *Auxiliary Signal Design for Failure Detection.* Princeton University Press, Princeton, NJ, USA.

Gorodetsky, A., Karaman, S., and Marzouk, Y. (2015). Efficient High-Dimensional Stochastic Optimal Motion Control using Tensor-Train Decomposition. In *Proceedings of the 11^{th} Robotics Science and Systems*, 1–9. Rome, Italy.

Julier, S., Uhlmann, J., and Durrant-Whute, H.F. (2000). A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators. *IEEE Transactions on Automatic Control*, 45(3), 477–482.

Kazeev, V. and Oseledets, I. (2013). The tensor structure of a class of adaptive algebraic wavelet transforms. Technical report, ETH Zürich.

Kerestecioğlu, F. and Zarrop, M.B. (1991). Optimal Input Design for Change Detection in Dynamical Systems. In *Proceedings of the 1991 European Control Conference*, 321–326. Grenoble, France.

Kolda, T.G. and Bader, B.W. (2009). Tensor Decompositions and Applications. *SIAM Review*, 51(3), 455–500.

Král, L. and Punčochář, I. (2022). Policy search for active fault diagnosis with partially observable state. *International Journal of Adaptive Control and Signal Processing*, 36(9), 2190–2216.

Leger, J.B. (2023). Parametrization cookbook: A set of bijective parametrizations for using machine learning methods in statistical inference. arXiv:2301.08297.

Nikoukhah, R. (1998). Guaranteed Active Failure Detection and Isolation for Linear Dynamical Systems. *Automatica*, 34(11), 1345–1358.

Oseledets, I. and Tyrtyshnikov, E. (2010). TT-Cross approximation for multidimensional arrays. *Linear Algebra and its Applications*, 432, 70–88.

Pinheiro, J.C. and Bates, D.M. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6, 289–296.

Punčochář, I., Škach, J., and Šimandl, M. (2015). Infinite Time Horizon Active Fault Diagnosis based on Approximate Dynamic Programming. In *Proceedings of the 54th IEEE Conference on Decision and Control*, 4456–4461. Osaka, Japan.

Punčochář, I. and Šimandl, M. (2008). Active fault detection and dual control in multiple model framework. In *Proceedings of the 17^{th} IFAC World Congress*, 7227–7232. Seoul, Korea.

Punčochář, I. and Šimandl, M. (2014). On infinite horizon active fault diagnosis for a class of non-linear non-Gaussian systems. *International Journal of Applied Mathematics and Computer Science*, 24(4), 795–807.

Raimondo, D.M., Marseglia, G.R., Braatz, R.D., and Scott, J.K. (2016). Closed-loop input design for guaranteed fault diagnosis using set-valued observers. *Automatica*, 74, 107–117.

Savostyanov, D. and Oseledets, I. (2011). Fast adaptive interpolation of multi-dimensional arrays in tensor train format. In *Proceedings of the 2011 International Workshop on Multidimensional Systems*, 1–8. Poitiers, France.

Streif, S., Petzke, F., Mesbah, A., Findeisen, R., and Braatz, R.D. (2014). Optimal Experimental Design for Probabilistic Model Discrimination Using Polynomial Chaos. In *Proceedings of the 19th IFAC World Congress*, 4103–4109. Cape Town, South Africa.

Striebel, C. (1965). Sufficient Statistics in the Optimum Control of Stochastic Systems. *Journal of Mathematical Analysis and Applications*, 12, 576–592.

Tichavský, P. and Phan, A.H. (2023). Tensor Chain Decomposition and Function Interpolation. In *IEEE Statistical Signal Processing Workshop*, 557—-561. Hanoi, Vietnam.

Zhang, X.J. (1989). *Auxiliary Signal Design in Fault Detection and Diagnosis.* Springer-Verlag, Berlin, Germany.