Preprints of the 18th IFAC World Congress
Milano (Italy) August 28 - September 2, 2011

Optimal experiment design for hypothesis testing applied to functional magnetic resonance imaging

X. Bombois * **A.J. den Dekker** * **C.R. Rojas** ** **H. Hjalmarsson** ** **P.M.J. Van den Hof** *

* *Delft Center for Systems and Control, Delft University of Technology, Delft (e-mail: x.j.a.bombois@tudelft.nl).*
** *Automatic Control, School of Electrical Engineering, KTH, Stockholm.*

Abstract: Hypothesis testing is a classical methodology of making decisions using experimental data. In hypothesis testing one seeks to discover evidence that either accepts or rejects a given null hypothesis $\mathcal{H}_0$. The alternative hypothesis $\mathcal{H}_1$ is the hypothesis that is accepted when $\mathcal{H}_0$ is rejected. In hypothesis testing, the probability of deciding $\mathcal{H}_1$ when in fact $\mathcal{H}_0$ is true is known as the false alarm rate, whereas the probability of deciding $\mathcal{H}_1$ when in fact $\mathcal{H}_1$ is true is known as the detection rate (or power) of the test. It is not possible to optimize both rates simultaneously. In this paper, we consider the problem of determining the data to be used for hypothesis testing that maximize the detection rate for a given false alarm rate. We consider in particular a hypothesis test which is relevant in functional magnetic resonance imaging (fMRI).

Keywords: optimal experiment design; detection; hypothesis testing

1. INTRODUCTION

A classical methodology to make a decision using experimental data is hypothesis testing (Kay, 1998). In hypothesis testing one seeks to discover evidence that either accepts or rejects a given null hypothesis $\mathcal{H}_0$. The alternative hypothesis $\mathcal{H}_1$ is the hypothesis that is accepted when $\mathcal{H}_0$ is rejected. Hypothesis testing is used in different fields of sciences. In the control literature, hypothesis testing is e.g. used for fault detection and isolation (see e.g. (Basseville, 1997, 1998; Huang and Tamayo, 2000)), model discrimination (Uosaki et al., 1984) or model validation (Ljung, 1999). In order to decide between the two (contradictory) hypotheses, a test (or decision rule) is designed based on a so-called test statistic. In designing such a test, it is important to find a good balance between its detection rate (i.e. the probability of deciding $\mathcal{H}_1$ when $\mathcal{H}_1$ is true indeed) and the false alarm rate (i.e. the probability of deciding $\mathcal{H}_1$ when in fact $\mathcal{H}_0$ is true). In other words, one has to find a good trade-off between the probability of making a so-called *Type I error* (deciding $\mathcal{H}_1$ when in fact $\mathcal{H}_0$ is true) and the probability of making a so-called *Type II error* (deciding $\mathcal{H}_0$ when in fact $\mathcal{H}_1$ is true). Note that reducing both error probabilities simultaneously is impossible. A typical approach in designing an optimal test is then to set the false alarm rate fixed while maximizing the detection rate (also known as the power of the test).

The choice of the decision rule and of the test statistic is of course very important when it comes to maximizing the detection rate (for a given false alarm rate). However, in this paper, we will suppose that the decision rule has been chosen a-priori and we will investigate another important factor affecting the power of the test: the data that have been collected to compute the test statistic. Indeed, if imposing an acceptable (i.e. low) false alarm rate is always possible, obtaining an acceptable (i.e. high) detection rate will highly depend on the richness of the data which are available to perform the hypothesis testing.

While hypothesis testing is now a state-of-the-art technique, until now, relatively few attention has been devoted to the optimal design of the experiment yielding the data which will be used for hypothesis testing. We will call this experiment the detection experiment in the sequel. In (Kerestecioglu and Zarrop, 1994), the optimal detection experiment design problem is formulated for hypothesis tests where the *true* hypothesis can frequently change over time and where the detection delay is therefore crucial. In this paper, we consider hypothesis tests where the *true* hypothesis does not change over time. In this case, we define the optimal detection experiment as that particular experiment (within an a-priori determined class of admissible detection experiments) which, for a given false-alarm rate, leads to the largest detection rate.

This definition is very general and can be applied to different hypothesis testing problems. However, in this paper, we restrict attention to a particular hypothesis testing problem where the objective is to determine whether there exists a causal relation between a signal $u(t)$ and another signal $y(t)$ or whether $y(t)$ is only made up of stochastic noise. This problem is e.g. relevant in functional magnetic resonance imaging (fMRI) as will be evidenced in the next section.

The optimal detection experiment design problem in that particular case is then very close to the problem of optimal input design for model discrimination (Uosaki et al., 1984).

In that paper, the hypothesis test consists of determining, among two model structures of different orders, the model structure which is better suited for a given plant. Besides the fact that the problems treated in these two papers are not identical, the main difference between (Uosaki et al., 1984) and the present paper is that the solution presented in (Uosaki et al., 1984) is based on asymptotic results (i.e. for detection experiments which are infinitely long) while, in the present paper, inspired by (Liu et al., 2007), we propose a finite-time solution for the optimal detection experiment design problem. Note that the problem of optimal input design for model discrimination was originally treated in the statistical literature (see e.g. (Atkinson and Fedorov, 1975; Fedorov and Khabarov, 1986)), but no explicit solution to the input design problem was given in this literature. Note finally that Nikoukhah (1998) proposes also a method for optimal input design for model discrimination, but in an non-stochastic framework.

2. DETECTION EXPERIMENT AND DECISION RULE

As mentioned in the introduction, we restrict attention to the following hypothesis testing problem where we would like to determine whether there exists a causal relation between a signal $u(t)$ and another signal $y(t)$. In particular, we would like to determine whether we are in hypothesis $\mathcal{H}_0$ or in hypothesis $\mathcal{H}_1$:

$$\mathcal{H}_0: \quad y(t) = H(z)e(t) \tag{1}$$

$$\mathcal{H}_1: \quad y(t) = G(z)u(t) + H(z)e(t) \tag{2}$$

where $G(z)$ and $H(z)$ are two stable and causal discrete-time transfer functions [1]. $H(z)$ is furthermore inversely stable and monic. The Gaussian noise $e(t)$ is white and independent of $u(t)$ and has a variance σ_e^2. For the case of simplicity, we will suppose that the only uncertainty in our problem is the question of knowing whether $\mathcal{H}_0$ or $\mathcal{H}_1$ is valid. This means that we suppose $G(z)$, $H(z)$ and σ_e^2 known.

We can interpret the considered problem as that of determining whether the transfer function between the signal $u(t)$ and $y(t)$ is equal to $G(z)$ ($\mathcal{H}_1$) or is equal to 0 ($\mathcal{H}_0$). This problem is relevant e.g. in functional magnetic resonance imaging (fMRI) in order to determine if a given zone (the so-called *voxel*) of the brain is active during the performance of a given action (stimulus). In this case, the signal $u(t)$ is the stimulus while the signal $y(t)$ is a physiological signal which is an image of the activity of the considered voxel. If the brain is active when performing the stimulus u, the signal $y(t)$ is as given in (2). If it is not active, the signal $y(t)$ is made up of only measurement noise such as in (1).

In order to discriminate between the two hypotheses, we perform a detection experiment. The detection experiment consists of generating a sequence $u(t)$ $(t = 0...N-1)$ (the stimulus in the fMRI application) and of measuring simultaneously the signal $y(t)$ $(t = 0...N-1)$. We will suppose that $u(t) = 0$ for $t < 0$. Based on the collected data $Z^N = \{\ u(t),\ y(t)\ |\ t = 0...N-1\}$, the most likely hypothesis can be determined using different approaches. In this paper, we consider an approach based on the identification with Z^N of a parameter θ distinguishing the two hypotheses. This technique that will be detailed in the sequel can be shown to be equivalent to the likelihood ratio test (the most powerful test in practice (Kay, 1998)) and is therefore the most commonly used approach in fMRI (see, e.g. Friston (1996)). The model structure that will be used for the identification is as follows:

$$\mathcal{M}: \quad y(t) = \theta\ \ (G(z)u(t))\ +\ H(z)e(t) \tag{3}$$

with $\theta \in \mathbf{R}$ the parameter to be determined. This parameter allows the distinction between $\mathcal{H}_0$ and $\mathcal{H}_1$. Indeed, the two hypotheses (1)-(2) are in this model structure equivalent to:

$$\mathcal{H}_0: \quad \theta = 0 \tag{4}$$

$$\mathcal{H}_1: \quad \theta = 1 \tag{5}$$

The prediction error estimate $\hat{\theta}_N$ of the parameter θ in $\mathcal{M}$ based on the data Z^N is given by (Ljung, 1999):

$$\hat{\theta}_N \triangleq arg \min_{\theta} \frac{1}{N} \sum_{t=0}^{N-1} H^{-1}(z)\,(y(t) - \theta\ \ (G(z)u(t)))$$

$$\Longrightarrow \hat{\theta}_N = \left(\frac{1}{N} \sum_{t=0}^{N-1} x^2(t) \right)^{-1} \frac{1}{N} \sum_{t=0}^{N-1} x(t)z(t)$$

with $x(t) = H^{-1}(z)G(z)u(t)$ and $z(t) = H^{-1}(z)y(t)$. It is important to note that $\hat{\theta}_N$ has the following properties under $\mathcal{H}_0$ and $\mathcal{H}_1$:

$$\hat{\theta}_N \sim \mathcal{N}(0, \sigma_\theta^2) \quad \text{under } \mathcal{H}_0 \tag{6}$$

$$\hat{\theta}_N \sim \mathcal{N}(1, \sigma_\theta^2) \quad \text{under } \mathcal{H}_1 \tag{7}$$

$$with\ \sigma_\theta^2 = \frac{\sigma_e^2}{N} \left(\frac{1}{N} \sum_{t=0}^{N-1} x^2(t) \right)^{-1} \tag{8}$$

The relations (6)-(8) follow from the fact that $z(t) = e(t)$ under $\mathcal{H}_0$ and $z(t) = x(t) + e(t)$ under $\mathcal{H}_1$.

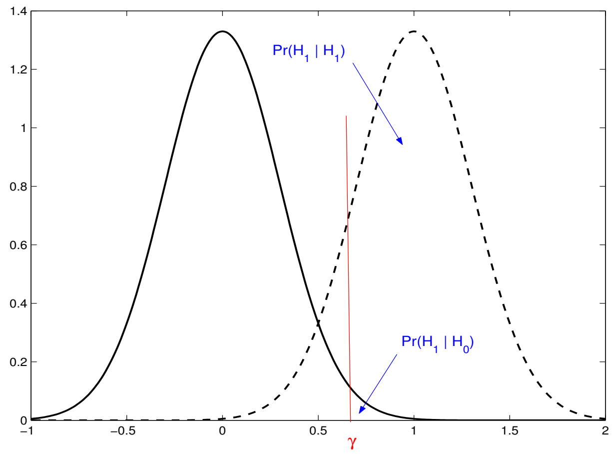

Fig. 1. Probability density function of $\hat{\theta}_N$ under $\mathcal{H}_0$ (solid) and under $\mathcal{H}_1$ (dashed) and indication of the false alarm and detection rates for an arbitrary value of γ.

[1] z being defined both as the Z-transform operator and the forward shift operator.

As mentioned above, we use the estimate $\hat{\theta}_N$ as a test statistic to formulate the decision rule to decide between $\mathcal{H}_0$ and $\mathcal{H}_1$:

$$\hat{\theta}_N \leq \gamma \implies \text{choose } \mathcal{H}_0$$
$$\hat{\theta}_N > \gamma \implies \text{choose } \mathcal{H}_1 \tag{9}$$

where $\gamma \in \mathbf{R}$ is a user-chosen threshold. Classically a decision rule is characterized by two probabilities. The first probability is the so-called false alarm rate $P_{FA} = Pr(\mathcal{H}_1|\mathcal{H}_0)$ i.e. the probability of choosing $\mathcal{H}_1$ when $\mathcal{H}_0$ is true. The second probability is the so-called detection rate $P_D = Pr(\mathcal{H}_1|\mathcal{H}_1)$ i.e. the probability of concluding $\mathcal{H}_1$ when $\mathcal{H}_1$ is true. For the decision rule (9), these probabilities are given by:

$$P_{FA} = Pr(\mathcal{H}_1|\mathcal{H}_0) = Pr(\hat{\theta}_N > \gamma \mid \hat{\theta}_N \sim \mathcal{N}(0, \sigma_\theta^2))$$
$$= \int_\gamma^\infty \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \; e^{\dfrac{-(\theta-0)^2}{2\sigma_\theta^2}} \; d\theta \tag{10}$$

$$P_D = Pr(\mathcal{H}_1|\mathcal{H}_1) = Pr(\hat{\theta}_N > \gamma \mid \hat{\theta}_N \sim \mathcal{N}(1, \sigma_\theta^2))$$
$$= \int_\gamma^\infty \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \; e^{\dfrac{-(\theta-1)^2}{2\sigma_\theta^2}} \; d\theta \tag{11}$$

with σ_θ^2 as in (8). The notions of false alarm and detection rates are illustrated in Figure 1.

It is obvious that we would like to determine the threshold γ in such a way that P_{FA} is as small as possible and P_D as large as possible. However, it is not possible to optimize both rates simultaneously. Indeed, there is only one degree of freedom in the decision rule i.e. γ. A typical approach then in designing a hypothesis test is to hold the false alarm rate fixed (by choosing γ in such a way that $P_{FA} = \alpha$) while maximizing the detection rate (also known as the power of the test).

3. OPTIMAL DESIGN OF THE DETECTION EXPERIMENT

3.1 General concepts

We have seen in the previous section that, for a given detection experiment, it is always possible to fix a small false alarm rate for the decision rule. However, the corresponding detection rate cannot be tuned a-priori. This detection rate will in fact depend on the richness of the data collected during the detection experiment. This observation leads to the following definition of the optimal detection experiment.

Definition 1. Suppose that we have a-priori chosen a decision rule based on a test statistic determined from a detection experiment. The optimal detection experiment is then that particular detection experiment (within an a-priori determined class of admissible detection experiments) which leads to the largest detection rate P_D for the chosen decision rule when this decision rule is formulated in such a way to guarantee a prespecified false alarm rate P_{FA}.

This definition is very general and applies to all detection problems. In the case considered in this paper, designing optimally the detection experiment consists of determining, within a predefined class $\mathcal{U}$ of admissible input sequences, that input sequence $u(t)$ ($t = 0...N-1$) leading to a decision rule (9) for which the detection rate $P_D = Pr(\hat{\theta}_N > \gamma \mid \hat{\theta}_N \sim \mathcal{N}(1, \sigma_\theta^2))$ is maximized when γ has been chosen in such a way to guarantee a false alarm rate $P_{FA} = Pr(\hat{\theta}_N > \gamma \mid \hat{\theta}_N \sim \mathcal{N}(0, \sigma_\theta^2))$ equal to an user-chosen α.

In the following proposition, we show that, for this particular situation, maximizing the detection rate is equivalent to minimizing the variance σ_θ^2 of the test statistic $\hat{\theta}_N$

Proposition 1. Consider the detection problem aiming at discriminating the two hypotheses (1) and (2) using a detection experiment with an input sequence $u(t)$ ($t = 0...N-1$) belonging to an user-chosen set of admissible sequences $\mathcal{U}$. Consider also the corresponding decision rule (9). Then the optimal input sequence for the detection experiment according to Definition 1 is the one solving the following optimization problem:

$$arg \min_{u(t) \in \mathcal{U}} \sigma_\theta^2 \tag{12}$$

with σ_θ^2 the variance (8) of the test statistic $\hat{\theta}_N$ used in the decision rule (9)

Proof. The proof is a direct consequence of Definition 1 and of relations (10)-(11). Indeed, the smaller σ_θ^2, the smaller the threshold γ to guarantee a pre-specified false alarm rate P_{FA} and therefore the larger the detection rate P_D. ■

Remark. From the proof of Proposition 1, it is obvious that the sequence solving (12) is also the optimal sequence for the following hypothesis test:

$$\mathcal{H}_0 : \; \theta = 0 \tag{13}$$
$$\mathcal{H}_1 : \; \theta \neq 0 \tag{14}$$

In the remainder of this section, we will determine the input sequence solving (12) for two different admissible sets $\mathcal{U}$.

3.2 Optimal detection with an input sequence of fixed energy

When the set $\mathcal{U}$ of admissible input sequences is the set $\mathcal{U}_{N,\beta}$ of sequences $u(t)$ ($t = 0...N-1$) of given duration and bounded energy, the problem (12) becomes:

$$arg \min_{u(t)\ (t=0...N-1)} \sigma_\theta^2$$
$$\text{such that } \sum_{t=0}^{N-1} u^2(t) \leq \beta \tag{15}$$

with β the maximal allowed energy and σ_θ^2 as given in (8). The solution of this problem is given in the following proposition.

Proposition 2. Consider the detection problem aiming at discriminating the two hypotheses (1) and (2) using a detection experiment with an input sequence $u(t)$ ($t = 0...N-1$) of fixed duration N and maximal energy β.

Let us introduce the vector form of the input sequence:

$$U = \begin{pmatrix} u(0) \\ u(1) \\ \vdots \\ u(N-1) \end{pmatrix} \in \mathbf{R}^N$$

and let us define the matrix $P \in \mathbf{R}^{N\times N}$:

$$P = \begin{pmatrix} k(0) & 0 & 0 & 0 & ... & 0 \\ k(1) & k(0) & 0 & 0 & ... & 0 \\ k(2) & k(1) & k(0) & 0 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ k(N-1) & k(N-2) & k(N-3) & k(N-4) & ... & k(0) \end{pmatrix}$$

where $k(t)$ is the pulse response of the transfer function $H^{-1}(z)G(z)$.

Then, the vector form U_{opt} of the sequence $u_{opt}(t)$ solving (15), i.e. the optimal detection sequence according to Definition 1, is given by:

$$U_{opt} = \pm\sqrt{\beta}\; V_{max}(P^T P) \tag{16}$$

where $V_{max}(A)$ denotes the eigenvector corresponding to the largest eigenvalue of the matrix A. We suppose that the vector $V_{max}(A)$ is normed i.e. $V_{max}(A)^T V_{max}(A) = 1$

Proof. The signal $x(t) = H^{-1}(z)G(z)u(t)$ in the expression (8) of σ_θ^2 can be rewritten as follows in the vector form:

$$\begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{pmatrix} = PU$$

with the matrix P and the vector U as defined in Proposition 2. Consequently, we have that

$$\sigma_\theta^2 = \sigma_e^2 \left(U^T P^T P U\right)^{-1}$$

The problem (15) can thus be reformulated as:

$$\begin{aligned} &arg \max_{U\in\mathbf{R}^N} \; U^T P^T P U \\ &\text{such that } U^T U \leq \beta \end{aligned} \tag{17}$$

Since $U^T P^T P U$ is a quadratic cost function, it is obvious by definition of the eigenvectors that the optimal U is the scaled version (16) of the eigenvector corresponding to the largest eigenvector of $P^T P$. ■

Remark 1. In (Uosaki et al., 1984), for a model discrimination problem, it was also recognized that the power of the hypothesis test is maximized by minimizing the variance of an identified parameter. However, as opposed to the result in Proposition 2, the optimal input signal $u(t)$ in (Uosaki et al., 1984) is determined based on the asymptotic expression for the variance and not the exact expression (8). The duality between model discrimination and parameter estimation was also recognized in (Fedorov and Khabarov, 1986) and (Bohlin and Rewo, 1980), but without explicit solution for the optimal design problem such as in Proposition 2.

Remark 2. The last part of the proof of Proposition 2 is inspired by the philosophy of the proof of Theorem 3.1 in (Liu et al., 2007).

3.3 Optimal detection with an input sequence relevant for the fMRI application

When considering the fMRI application, the class $\mathcal{U}_{N,fMRI}$ of admissible input sequences of duration N is more restricted than the one considered in the previous subsection. Indeed, the values of the input sequence are restricted to be equal either to zero or one and, in particular, to sequences $u(t)$ ($t = 0...N-1$) of the following form:

$$\begin{cases} u(t) = 1 \;\; for\; t = 0,\; 1, ...,\; n-1 \\ u(t) = 0 \;\; for\; t = n,\; n+1, ...,\; 2n-1 \end{cases}$$

and such that $u(t+2n) = u(t)$ for $t = 0...N-2n-1$

The input sequence is thus a truncated periodic signal of period $2n$. The half period n is generally chosen in such a way that the sequence $u(t)$ ($t = 0...N-1$) contains an integer number of periods. This discrete set of values for n will be denoted $\mathcal{N}_N$. (e.g., for $N = 60$, $\mathcal{N}_{N=60}$ is the set $\{1,\ 2,\ 3,\ 5,\ 6,\ 10,\ 15,\ 30\}$).

The optimal detection problem (12) when $\mathcal{U} = \mathcal{U}_{N,fMRI}$ is thus equivalent to finding the value $n \in \mathcal{N}_N$ which leads to the sequence $u(t) \in \mathcal{U}_{N,fMRI}$ corresponding to the smallest σ_θ^2. A way to solve this optimization problem is to compute the corresponding σ_θ^2 for each value $n \in \mathcal{N}_N$ and determine in this way the value of n corresponding to the smallest σ_θ^2.

Remark. Note that all sequences $u(t)$ in $\mathcal{U}_{N,fMRI}$ have an energy $\sum_{t=0}^{N-1} u^2(t)$ equal to $\frac{N}{2}$. The optimal sequence in $\mathcal{U}_{N,\beta=\frac{N}{2}}$ (see Section 3.2) can thus be used to see how conservative the restriction to $\mathcal{U}_{N,fMRI}$ is.

4. NUMERICAL ILLUSTRATIONS

In fMRI applications, it is generally supposed that the transfer function $G(z)$ in (2) has a finite pulse response $g(t)$ (the so-called hemodynamic response function). In this example, we consider the pulse response $g(t)$ represented in Figure 2. We observe that $g(t)$ is non-zero in the time interval [1, 9]. For the noise description $H(z)$ in (1)-(2), we will consider two possibilities in this example:

$$H_1(z) = 1$$
$$H_2(z) = 1 + c_1\; z^{-1} + c_2\; z^{-2} + c_3\; z^{-3} + c_4\; z^{-4}$$

with $c_1 = -1.99185$, $c_2 = 2.20265$, $c_3 = -1.84083$ and $c_4 = 0.89413$.

Let us first consider the case $H(z) = H_1(z)$ and let us choose $\sigma_e^2 = 20$.

We wish to design the detection experiment which will maximize the detection rate P_D when determining the

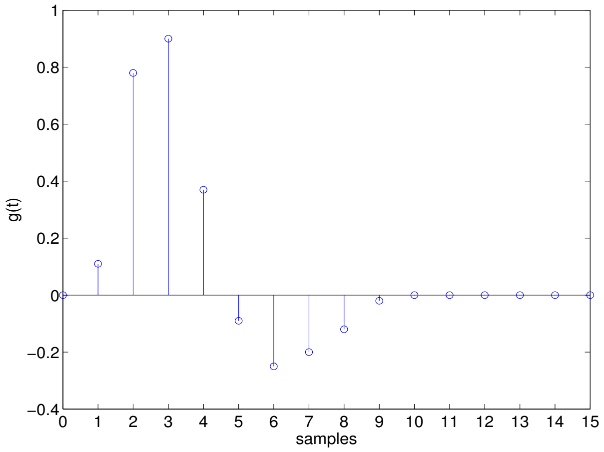

Fig. 2. Pulse response $g(t)$ of $G(z)$.

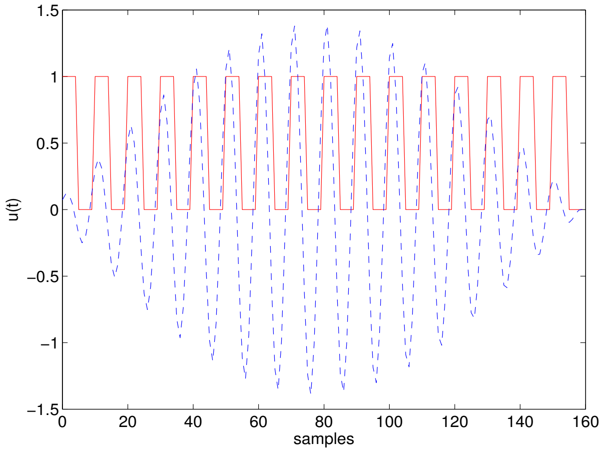

Fig. 3. $H(z) = H_1(z)$: optimal detection sequence $u(t)$ in $\mathcal{U}_{N=160,\beta=80}$ (blue dashed) and optimal detection sequence in $\mathcal{U}_{N=160,fMRI}$ (red solid).

threshold γ in (9) in order to guarantee a false alarm rate $P_{FA} = 2.5\%$. For this purpose, we have to determine the detection sequence $u(t)$ $(t = 0...N-1)$ in the set $\mathcal{U}_{N,fMRI}$ (see Section 3.3) which minimizes the variance σ_θ^2 of the detection parameter. Here, we choose $N = 160$.

We know that the sequences in $\mathcal{U}_{N=160,fMRI}$ all have an energy equal to $\frac{N}{2} = 80$. In order to have a reference point, we first determine, using Proposition 2, the optimal sequence in the set $\mathcal{U}_{N=160,\beta=80}$ of sequences of duration $N = 160$ and of maximal energy $\beta = 80$. The optimal sequence is represented in blue dashed in Figure 3. The variance σ_θ^2 of $\hat{\theta}_N$ is, for this optimal sequence, equal to 0.0456 and the obtained detection rate for this σ_θ^2 is equal to $P_D = 99.7\%$.

Let us now consider the sequences in $\mathcal{U}_{N=160,fMRI}$. As mentioned in Section 3.3, in order to find the optimal sequence, we need to compute the variance σ_θ^2 for each possible value of the half period $n \in \mathcal{N}_{N=160}$. This leads to Table 1 where we also give for each $n \in \mathcal{N}_{N=160}$, the corresponding detection rate.

We observe that the minimal variance σ_θ^2 and thus the maximal detection rate is obtained for $n = 5$. This optimal sequence is represented in red solid in Figure 3. We observe

n	σ_θ^2	P_D
1	0.2257	55.8 %
2	0.1785	65.8%
4	0.0768	95%
5	0.0733	95.8%
8	0.0819	93.7%
10	0.0868	92.4%
16	0.0954	89.9%
20	0.0986	89%
40	0.1058	86.7%
80	0.1098	85.5%

Table 1. $H(z) = H_1(z)$: σ_θ^2 and P_D for different values of the half period n of the sequences in $\mathcal{U}_{N=160,fMRI}$.

also that, for sequences within $U_{N=160,fMRI}$, the resulting detection rates vary from 55.8% (when $n = 1$) to 95.8% (when $n = 5$). We see thus that designing intelligently the detection experiment is important.

Because $\mathcal{U}_{N=160,fMRI} \subset \mathcal{U}_{N=160,\beta=80}$, the optimal sequence in $\mathcal{U}_{N=160,fMRI}$ cannot attain the detection rate $P_D = 99.7\%$ obtained when considering the class $\mathcal{U}_{N=160,\beta=80}$. However, the difference between the two detection rates is not very important. When comparing the two optimal sequences in Figure 3, we observe that the one has an oscillatory behaviour with a period of ten samples and the other one is a periodic signal with the same period.

Let us now consider the other noise model $H(z) = H_2(z)$ and let us choose $\sigma_e^2 = 10000$. The other parameters of the problem remain unchanged. For this new case, the optimal sequence in $\mathcal{U}_{N=160,fMRI}$ is the one corresponding to $n = 8$. This optimal sequence is represented in red solid in Figure 4 and corresponds to a detection rate $P_D = 99.2\%$. As shown in Table 2, the detection rate corresponding to the sequence with all other values of n are here much smaller which shows even more the importance of an optimal design of the detection experiment.

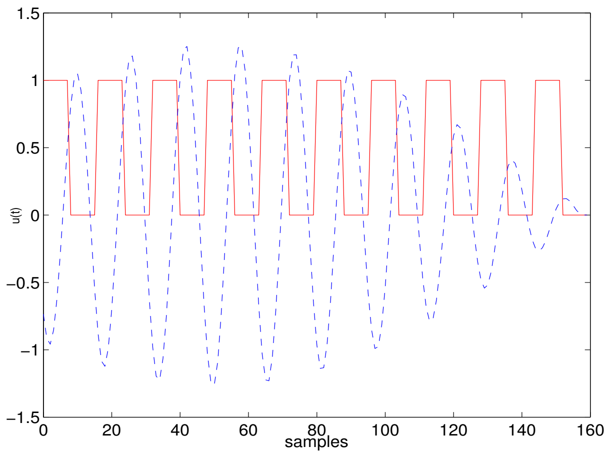

Fig. 4. $H(z) = H_2(z)$: optimal detection sequence $u(t)$ in $\mathcal{U}_{N=160,\beta=80}$ (blue dashed) and optimal detection sequence in $\mathcal{U}_{N=160,fMRI}$ (red solid).

Using Proposition 2, we can also determine the optimal sequence in the set $\mathcal{U}_{N=160,\beta=80}$. When comparing, in Figure 4, the optimal sequences in $\mathcal{U}_{N=160,\beta=80}$ and $\mathcal{U}_{N=160,fMRI}$, we observe, like in the case $H(z) = H_1(z)$,

n	σ_θ^2	P_D
1	6.1351	6 %
2	5.3053	6.3 %
4	3.7527	7.4%
5	2.3022	9.7%
8	0.0445	99.7%
10	0.5386	27.5%
16	2.2137	9.9%
20	1.7389	11.5%
40	0.7481	21%
80	2.2926	9.7%

Table 2. $H(z) = H_2(z)$: σ_θ^2 and P_D for different values of the half period n of the sequences in $\mathcal{U}_{N=160,fMRI}$.

that these two signals are characterized by the same period of 16 samples.

We have observed above that the (frequency content of the) optimal sequence is different for the two choices of $H(z)$. It is interesting to understand the reason of this difference. For this purpose, let us observe that σ_θ^2 is inversely proportional to the power of the signal $x(t) = H^{-1}(z)G(z)u(t)$. The transfer function $H^{-1}(z)G(z)$ seems thus instrumental in determining the optimal $u(t)$. When comparing the modulus of $H^{-1}(e^{j\omega})G(e^{j\omega})$ for the two choices for $H(z)$, we observe that these two modulus attain their maximum at the frequency corresponding to the period of the optimal sequence. Indeed, we have $\omega_1 = arg \max_\omega |H_1^{-1}(e^{j\omega})G(e^{j\omega})| = 0.63$ and $\omega_2 = arg \max_\omega |H_2^{-1}(e^{j\omega})G(e^{j\omega})| = 0.39$. The corresponding fundamental period is $T_1 = \frac{2\pi}{\omega_1} \approx 10$ and $T_2 = \frac{2\pi}{\omega_2} \approx 16$. The optimal sequence excites the transfer function $H^{-1}G$ at the frequency of maximal gain.

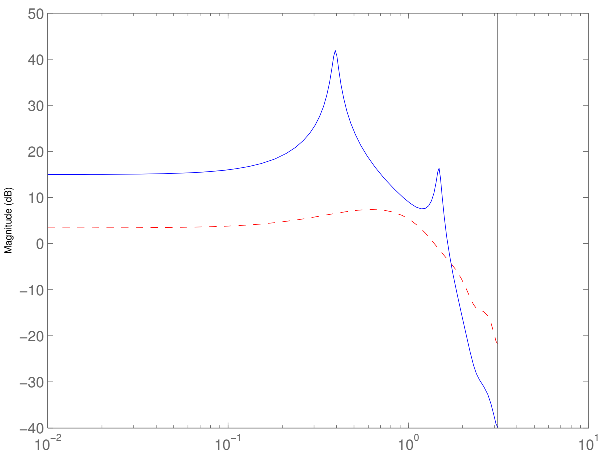

Fig. 5. Modulus of $H^{-1}(e^{j\omega})G(e^{j\omega})$ for $H(z) = H_1(z)$ (red dashed) and for $H(z) = H_2(z)$ (blue solid).

5. CONCLUSIONS

In this paper, we have considered an hypothesis testing problem which is frequently used in functional magnetic resonance imaging. For this particular hypothesis test, we have determined, among the class of admissible experiments yielding the data required for the hypothesis test, that particular experiment which optimizes the detection rate for a given false alarm rate. In the simulation examples, we have seen that optimizing the experiment is certainly useful in practice since, for a fixed false alarm rate, the difference between the optimal detection rate and the worst case detection rates can be significant.

Even though these have been only applied to one particular hypothesis testing problem in this paper, the concept of optimal detection experiment introduced in this paper is much more general and can be applied to other problems. In particular, in a future work, we will treat the problem of detecting with the least disturbing experiment on the closed-loop plant whether an observed performance drop is due to a change in dynamics of the plant or due to any other causes (temporary disturbances, actuator or sensor failures, etc.).

REFERENCES

A.C. Atkinson and V. Fedorov. The design of experiments for discriminating between two rival models. *Biometrika*, 62(1):57–70, 1975.

M. Basseville. Information criteria for residual generation and fault detection and isolation. *Automatica*, 33(5): 783–803, 1997.

M. Basseville. On board component fault detection and isolation using the statistical local approach. *Automatica*, 34(11):1391–1415, 1998.

T. Bohlin and L. Rewo. Experiment design for maximum-power model validation. *Automatica*, 16:405–408, 1980.

V. Fedorov and V. Khabarov. Duality of optimal designs for model discrimination and parameter estimation. *Biometrika*, 73(1):183–190, 1986.

K.J. Friston. Statistical parametric mapping and other analysis of functional imaging data. In *Brain Mapping: The Methods*, pages 363–385. Academic Press, 1996.

B. Huang and E. Tamayo. Model validation for industrial model predictive control systems. *Chemical Engineering Science*, 55:2315–2327, 2000.

S.M. Kay. *Fundamentals of statistical signal processing: Detection theory.* Prentice-Hall, 1998.

F. Kerestecioglu and M. Zarrop. Input design for detection of abrupt changes in dynamical systems. *International Journal of Control*, 59:1063–1084, 1994.

Y. Liu, T.F. Wong, and W.W. Hager. Training signal design for estimation of correlated MIMO channels with colored interference. *IEEE Transactions on Signal Processing*, 55(4):1486–1497, 2007.

L. Ljung. *System Identification: Theory for the User.* Prentice-Hall, Englewood Cliffs, NJ, 2nd edition edition, 1999.

R. Nikoukhah. Guaranteed active failure detection and isolation for linear dynamical systems. *Automatica*, pages 1345–1358, 1998.

K. Uosaki, I. Tanaka, and H. Sugiyama. Optimal input design for autoregressive model discrimination with constrained output variance. *IEEE Transactions on Automatic Control*, 29(4):348–350, 1984.