

# A Statistical Detection of an Anomaly from a Few Noisy Tomographic Projections

**Lionel Fillatre**

*ISTIT, FRE CNRS 2732, Université de Technologie de Troyes, 12 rue Marie Curie, BP 2060, 10010 Troyes Cedex, France  
Email: fillatrl@utt.fr*

**Igor Nikiforov**

*ISTIT, FRE CNRS 2732, Université de Technologie de Troyes, 12 rue Marie Curie, BP 2060, 10010 Troyes Cedex, France  
Email: nikiforov@utt.fr*

*Received 1 January 2004; Revised 19 November 2004*

The problem of detecting an anomaly/target from a very limited number of noisy tomographic projections is addressed from the statistical point of view. The imaged object is composed of an environment, considered as a nuisance parameter, with a possibly hidden anomaly/target. The GLR test is used to solve the problem. When the projection linearly depends on the nuisance parameters, the GLR test coincides with an optimal statistical invariant test.

**Keywords and phrases:** statistical hypotheses testing, (non)linear parametric model, nuisance parameter, invariant tests, missing observations, computerized tomography.

## 1. INTRODUCTION

Computerized tomography (CT) is a technique for reconstructing an object from its projections that are essentially the collections of line integrals of the attenuation scalar field at some set of orientations. The noninvasive nature of tomography has made it very useful for a variety of applications, including medical imaging, quantitative non-destructive testing, object recognition, and (biomedical) system monitoring, among others [1, 2, 3]. In certain practical applications, like baggage X-ray scanning or nondestructive testing, only a few projections are available and, hence, a perfect reconstruction of the scene is impossible [2].

The detection of an anomaly/target from projections has been already studied in [4], where the authors estimate the unknown location of a parameterized object by maximum likelihood (ML) estimation. Recently in [5], it is proposed to detect and localize an elementary geometrical shape on a rectangular grid but such a problem involves hypotheses testing over hypotheses spaces of extremely large cardinality. The computational complexity is then reduced by a multi-scale approach. Otherwise, such a problem can be viewed as an object recognition problem [6]. In contrast with [4, 5, 6], a more realistic problem statement has been discussed in [7, 8], where the observed scene is composed of an unknown deterministic environment, which is considered as a nuisance parameter, with a possibly hidden anomaly/target (detection of welding defects in radiographic images). A key issue

in such a detection problem is to state the significance of the observed deviation (due to the anomaly/target) with respect to a random noise and nuisance parameter. Because an anomaly/target is superimposed on the environment (see details in [7, 8]), it can be acquired by subtracting the environment from the original image.

The goal of this paper is to propose an optimal statistical tool to detect an anomaly/target directly from a limited number of noisy tomographic projections and/or with limited angles of view, without reconstructing the scene, that is, without solving the inverse problem, which may be ill-posed.

The contribution of this paper consists in the following developments.

- (1) A parametric-based approach, which describes a partly unknown environment by using a (non-) linear parametric *parsimonious* model, is proposed. Unlike the traditional “pixel-”based reconstruction methods, this approach permits to avoid the difficulties of the tomographic reconstruction problem that may be ill-posed in the case of a limited number of projections and/or angles of view. The hypothesis that the environment (nuisance parameter) is unknown but non-random practically means that it can be intentionally chosen to mask the anomaly/target, for example, by an adversary.
- (2) By using the proposed parametric approach, the anomaly/target detection problem is formalized as a

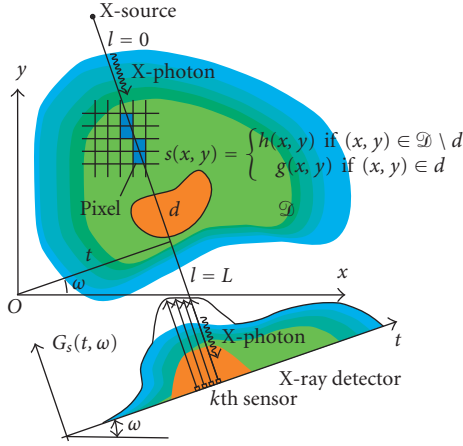


FIGURE 1: Detection of an anomaly from tomographic projections.

binary statistical decision with a nuisance parameter. The main issue of such a decision problem is its ability to detect an anomaly/target while being insensitive to the nuisance parameters and to deal with composite hypotheses.

- (3) To solve the statistical hypotheses testing problem, the general Wald's theory of uniformly best constant power tests, combined with the theory of invariant tests, is used. This theory has been adapted to the linear stochastic model describing an anomaly/target superimposed on the partly unknown deterministic environment. The proposed test coincides with the generalized likelihood ratio (GLR) test. This allows us to extend the developed theory to the nonlinear parsimonious parametric model and to warrant an  $\varepsilon$ -optimality of this extension.

The paper is organized as follows. In Section 2, a short introduction to the tomographic projection processing is proposed and the anomaly detection problem is stated. In Section 3, the GLR test is derived and it is shown that this test coincides with the optimal invariant decision rule designed for detecting the anomaly when the unknown environment is linearly parameterized. The mathematical details of nuisance parameter rejection and the optimal test design can be found in Appendices A and B. Section 4 is devoted to the nonlinear nuisance parameter case. The subtlety of nonlinear GLR test is discussed here. To illustrate the relevance of the developed tools, a practical example related to the problem of solid rocket motor examination is presented in Section 5. Finally, Section 6 contains some concluding remarks.

## 2. PROBLEM STATEMENT

The physical background is briefly presented in this section. Next, the environment and anomaly models are formalized and the problem of detecting an anomaly from a few noisy tomographic projections is briefly introduced.

### 2.1. Physical background

We are dealing with a specific inverse problem consisting in the detection of an anomaly from a few noisy tomographic projections. Some tutorial introduction to the computerized tomography (CT) can be found in [4, 9]. To simplify the presentation, only the two-dimensional (2D) case is discussed here but the 3D case can be easily derived from the following results. We define the X-ray attenuation coefficient  $s$  of the scene as a function of two variables  $x$  and  $y$ :  $(x, y) \mapsto s(x, y)$ . This situation is depicted in Figure 1. It is assumed that  $s$  is an element of the space  $\mathcal{L}_2(\mathcal{D})$  of square-integrable real-valued functions supported on a compact set  $\mathcal{D}$  in the plane. A projection  $G_s$  of the scene  $s$  at a particular angle  $\omega$  is a real-valued function defined on the real line by

$$G_s(t, \omega) = \int_0^L s(l) dl = \int_0^L s(t \cos \omega - l \sin \omega, t \sin \omega + l \cos \omega) dl, \quad (1)$$

where  $L$  is defined by the acquisition system geometry, and  $s(l) = s(t \cos \omega - l \sin \omega, t \sin \omega + l \cos \omega)$  corresponds to the X-ray attenuation coefficient of the scene  $s$  along the straight line parameterized by the angle  $\omega$  and the radius  $t$  in polar coordinates. This line corresponds to the path of the X-ray flux emitted by the X-source. In the case of CT, the projections are obtained by using a linear numerical detector composed of  $n$  elementary sensors counting the total amount of X-photons passing through the object (see Figure 1). We define a family  $\tau = \{t_1, t_2, \dots, t_n\}$  of  $n$  points of interest regularly spaced on the linear detector. To simplify the presentation, it is assumed that the X-source is infinitely far from the detector. Hence, a parallel-beam line-integral projection taken at the angle  $\omega$  is related to the original object  $s$  by the discrete Radon transform  $R_{\omega, \tau} : \mathcal{L}_2(\mathcal{D}) \mapsto \mathbb{R}^n$  defined by

$$R_{\omega, \tau}(s) = (G_s(t_1, \omega), G_s(t_2, \omega), \dots, G_s(t_n, \omega))^T. \quad (2)$$

### 2.2. Environment model

The anomaly-free environment (called *background* in [4, 5]) corresponds to the original scene without any anomaly. Sometimes such an environment is assumed to be a known function or a zero-mean Gaussian random field [4, 5]. However, in many practical applications (see, e.g., [7, 8]), such a hypothesis is not verified and the environment must be considered as an unknown deterministic function, which is called nuisance parameter in the rest of the paper. From the practical point of view, the environment is naturally space limited to a compact set  $\mathcal{D} \subset \mathbb{R}^2$ . Physically, the value  $s(x, y) = h(x, y)$  corresponds to the X-ray attenuation coefficient of the anomaly-free environment defined at the point  $(x, y) \in \mathcal{D}$  (see Figure 1). We stress the peculiarity of our parametric CT approach. A traditional CT method consists in reconstructing the original scene, in our case the function  $(x, y) \mapsto h(x, y)$ , with a finite number of variables by decomposing the compact  $\mathcal{D}$  into nonoverlapping regions, "pixels"  $\{V_i\}_{i \in \mathbb{N}}$  (or "voxels" in the 3D case), which are naturally ordered over  $\mathcal{D}$ , in the lexicographical sense, by the set of natural integers  $\mathbb{N}$ . Different decompositions well adapted to the

tomographic problem have already been studied [10, 11]. We consider a parallel-beam line-integral projection, taken at the angle  $\omega$ , and define a set of pixel's indicators:

$$\phi_i(x, y) = \begin{cases} 1 & \text{if } (x, y) \text{ is within the } i\text{th} \\ & \text{partition : } (x, y) \in V_i, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

Hence a line integral projection  $G_h(t, \omega)$  of the function  $h$  can be approximated in the following manner:

$$\begin{aligned} G_h(t, \omega) &= \int_0^L h(t \cos \omega - l \sin \omega, t \sin \omega + l \cos \omega) dl \\ &= \sum_{i \in \mathbb{N}_{t, \omega}} \int_0^L h(t \cos \omega - l \sin \omega, t \sin \omega + l \cos \omega) \\ &\quad \times \phi_i(t \cos \omega - l \sin \omega, t \sin \omega + l \cos \omega) dl \\ &\simeq \sum_{i \in \mathbb{N}_{t, \omega}} h_i \int_0^L \phi_i(t \cos \omega - l \sin \omega, t \sin \omega + l \cos \omega) dl \\ &= \sum_{i \in \mathbb{N}_{t, \omega}} h_i l_i, \end{aligned} \quad (4)$$

where  $\mathbb{N}_{t, \omega} = \mathbb{N}(t, \omega)$  is the subset of lexicographically ordered pixels over the path from the source to the X-ray detector taken at the angle  $\omega$  and the parameter  $t$ ,  $h_i$  is the “mean” value of the function  $(x, y) \mapsto h(x, y)$  over the pixel  $V_i$ , and  $l_i$  is the length of the X-ray beam restricted to the pixel  $V_i$  (see Figure 1). It appears from the literature that no general rules can be designed for choosing a set of decompositions. Moreover, in our case, another problem arises due to a limited number of projections. To detect a hidden anomaly, it is necessary to obtain a good approximation of the unknown environment  $(x, y) \mapsto h(x, y)$ . Generally, to calculate such an approximation, the total number of pixels or unknown variables  $h_i$  should be large. This leads to a contradiction with a limited number of tomographic projections and/or with limited view angles. The resulting inverse problem is ill-posed. In other words, if the decomposition of  $\mathcal{D}$  produces a good approximation of the unknown environment  $(x, y) \mapsto h(x, y)$  at the expense of an important number of unknown variables, then this decomposition is not adapted to the considered problem. In contrast to the above pixel decomposition, a linearly parameterized *parsimonious* decomposition of the environment  $h$  is used in Section 3:

$$h_\mu(x, y) = \sum_{k=1}^m \mu_k h_k(x, y) \quad \forall (x, y) \in \mathcal{D}, \quad (5)$$

where  $\{h_1, \dots, h_m\}$  is a known family of basic functions in  $\mathcal{L}_2(\mathcal{D})$  and  $\mu = (\mu_1, \dots, \mu_m)^T$  is a real-valued vector of unknown parameters. We assume that the unknown environment  $h = h_\mu$  is composed of a finite number of a priori known basic elements  $h_k$ , representing the X-ray attenuation

coefficients, with unknown scaling coefficients  $\mu_k$ . Here, the line integral projection is given by

$$\begin{aligned} G_{h_\mu}(t, \omega) &= \int_0^L \sum_{k=1}^m \mu_k h_k(t \cos \omega - l \sin \omega, t \sin \omega + l \cos \omega) dl \\ &= \sum_{k=1}^m \mu_k \int_0^L h_k(t \cos \omega - l \sin \omega, t \sin \omega + l \cos \omega) dl \\ &= \sum_{k=1}^m \mu_k G_{h_k}(t, \omega). \end{aligned} \quad (6)$$

### 2.3. Anomaly model

It is desirable to detect an anomaly without any a priori information about it. For example, the support of the anomaly can be a nonconnected domain  $d = \bigcup_{i=1}^N d_i$ ,  $d_i \cap d_j = \emptyset$ , in  $\mathbb{R}^2$  with  $N \in \mathbb{N}^*$ . The goal is to detect any significant deviations from the environment observed in the original scene. We define the function  $(x, y) \mapsto f_{g, \mu}(x, y) \in \mathcal{L}_2(\mathcal{D})$  representing a local variation of the attenuation coefficient in the original scene due to the presence of the anomaly:

$$f_{g, \mu}(x, y) = \begin{cases} g(x, y) - h_\mu(x, y) & \text{if } (x, y) \in d, \\ 0 & \text{if } (x, y) \in \mathcal{D} \setminus d, \end{cases} \quad (7)$$

where  $(x, y) \mapsto g(x, y)$ ,  $(x, y) \in d$ , is the anomaly attenuation coefficient. Practically, the quantity  $f_{g, \mu}$  represents a *contrast* between the environment  $h$  and the true anomaly  $g$ . In the rest of the paper, it is assumed that the anomaly is a *contrast* with respect to the environment. This requirement can be expressed in the following manner:

$$\iint_d (f_{g, \mu}(x, y))^2 dx dy > 0 \quad \forall \mu \in \mathbb{R}^m. \quad (8)$$

### 2.4. Measurement model

We define now the model of X-ray attenuation coefficient  $(x, y) \mapsto s(x, y)$  that describes the two possible situations,  $\mathcal{H}_0$ : {environment is anomaly-free} and  $\mathcal{H}_1$ : {the anomaly is superimposed on the environment}, in the following manner:

$$s(x, y) = \begin{cases} h_\mu(x, y), & (x, y) \in \mathcal{D} & \text{under } \mathcal{H}_0, \\ f_{g, \mu}(x, y) + h_\mu(x, y), & (x, y) \in \mathcal{D} & \text{under } \mathcal{H}_1. \end{cases} \quad (9)$$

Putting together (2) and (9), we obtain the following measurement model for a particular view angle  $\omega$ :

$$Y_\omega = R_{\omega, \tau}(s) + \xi, \quad (10)$$

where  $Y_\omega \in \mathbb{R}^n$  is the vector of observations,  $\xi \in \mathbb{R}^n$  is a zero-mean gaussian noise,  $\xi \sim \mathcal{N}(0, \sigma^2 I_n)$ , corresponding to errors introduced by the acquisition system. The variance  $\sigma^2 > 0$  is assumed to be known. Due to the linearity of the

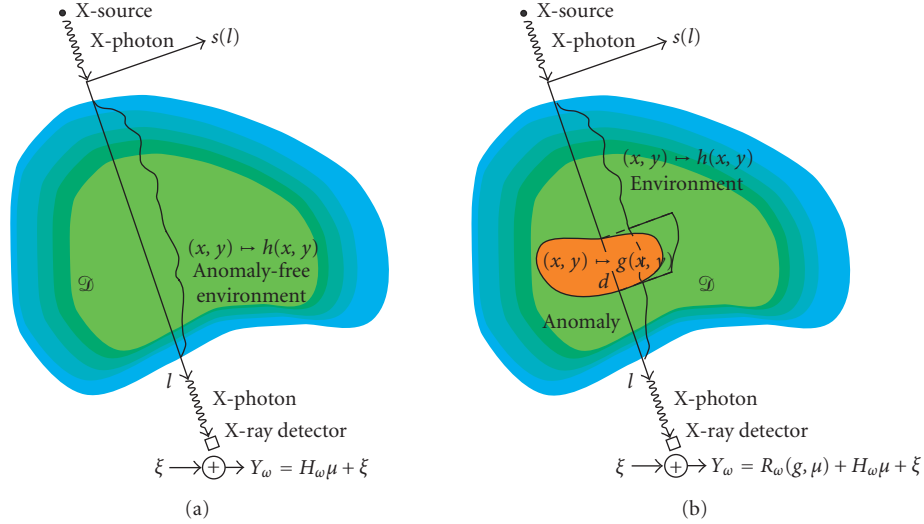


FIGURE 2: The “source-to-sensor” direction attenuation coefficient: (a) hypothesis  $\mathcal{H}_0$ : {environment is anomaly-free} and (b) hypothesis  $\mathcal{H}_1$ : {the anomaly is superimposed on the environment}.

operator  $R_{\omega, \tau}$ , the above measurement model can be rewritten as

$$Y_{\omega} = \begin{cases} H_{\omega} \mu + \xi & \text{under } \mathcal{H}_0, \\ R_{\omega}(g, \mu) + H_{\omega} \mu + \xi & \text{under } \mathcal{H}_1, \end{cases} \quad (11)$$

where  $R_{\omega}(g, \mu) = R_{\omega, \tau}(f_{g, \mu})$  and  $H_{\omega} = (H_{\omega}^1, \dots, H_{\omega}^m)$  with  $H_{\omega}^k = R_{\omega, \tau}(h_k) \in \mathbb{R}^n$  for  $k = 1, \dots, m$ . It is supposed that  $m < n$ . This situation is depicted in Figure 2. If  $P$  projections are available, the vector  $Y_{\omega_1, \dots, \omega_P}$ , the function  $R_{\omega_1, \dots, \omega_P}(g, \mu)$ , and the matrix  $H_{\omega_1, \dots, \omega_P}$  can be easily designed from “elementary” components  $Y_{\omega_i}$ ,  $R_{\omega_i}(g, \mu)$ , and  $H_{\omega_i}$  by concatenation. In the rest of the paper, the subscript  $\omega$  will be omitted to simplify the notations.

### 2.5. Anomaly detection problem

The detection problem consists in deciding which hypothesis is the true one,  $\mathcal{H}_0$ : {environment is anomaly-free} or  $\mathcal{H}_1$ : {anomaly is superimposed on the environment}, while considering the parameterized environment  $H\mu$  as a *nuisance parameter*. It is assumed that the anomaly projection  $R = R(g, \mu)$  is unknown in the rest of the paper. Nevertheless, sometimes it is possible to include some a priori information on the anomaly in the model by assuming that  $R = M\theta$ , where  $M$  is a full-rank matrix of size  $(n \times p)$  and the parameter vector  $\theta \in \mathbb{R}^p$  describes the anomaly projection. It is assumed in the rest of the paper that  $q + p < n$ , where  $\text{rank } H = q$ . We rewrite (11) to obtain the linear measurement model:

$$Y = M\theta + H\mu + \xi. \quad (12)$$

The anomaly detection problem consists in testing between  $\mathcal{H}_0$  and  $\mathcal{H}_1$ :

$$\mathcal{H}_0 = \{\theta = 0, \mu \in \mathbb{R}^m\}, \quad \mathcal{H}_1 = \{\theta \neq 0, \mu \in \mathbb{R}^m\}, \quad (13)$$

while considering  $\mu$  as an unknown vector.

## 3. STATISTICAL HYPOTHESES TESTING: LINEAR CASE

The optimal anomaly detection test is discussed in this section. It is assumed that the linear parsimonious decomposition perfectly fits the unknown environment:  $h_{\mu}(x, y) = h(x, y)$ . The decomposition  $h_{\mu}$  is linear according to the nuisance parameter  $\mu$ . First, a short introduction to the composite hypotheses testing is proposed. The mathematical details of nuisance parameter rejection and the optimal test design can be found in Appendices A and B. Next, it is shown that the GLR test coincides with this optimal invariant decision rule. Finally, the role of the matrix  $M$  and the statistical performances of the test are discussed.

### 3.1. Testing between two hypotheses

Let  $Y$  be a random variable with distribution  $P_{\theta, \mu}$  belonging to the parametric family  $\mathcal{P} = \{P_{\theta, \mu}\}$ , where  $\theta \in \Theta \subseteq \mathbb{R}^p$  is the informative parameter and  $\mu \in \mathbb{R}^m$  is the nuisance one. A *composite hypothesis* refers to a set of parameters  $\mathcal{H}_i : \{\theta \in \Theta_i\}$  with  $\Theta_i \subset \Theta$ ,  $i = 0, 1$ . It is assumed that  $\Theta_0 \cap \Theta_1 = \emptyset$ . The hypotheses testing problem (12) and (13) consists in deciding which hypotheses  $\mathcal{H}_i$  is the true one while considering  $\mu$  as an unknown vector. The quality of a statistical test is defined with the probability of false alarm and the power of the test. First of all, we define the class  $\mathcal{K}_{\alpha} = \{\delta : \sup_{\mu \in \mathbb{R}^m} \Pr_{\theta=0, \mu}(\delta = \mathcal{H}_1) \leq \alpha\}$  of tests with upper-bounded maximum false alarm probability, where the probability  $\Pr_{\theta, \mu}$  stands for the vector of observations  $Y$  being generated by distribution  $P_{\theta, \mu}$  and  $\alpha$  is the prescribed probability of false alarm. The power function  $\beta_{\delta}$  is defined with the probability of detection:  $\beta_{\delta}(\theta; \mu) = \Pr_{\theta, \mu}(\delta = \mathcal{H}_1)$ ,  $\theta \in \Theta_1$ , while considering  $\mu$  as an unknown vector. The pair  $(\alpha, \beta)$  is then a sufficient performance index. Obviously,  $\alpha$  should be as small as possible and  $\beta_{\delta}(\theta; \mu)$  should be large for every  $\theta \in \Theta_1$  and  $\mu \in \mathbb{R}^m$ . Without nuisance parameters, the ideal solution is a *uniformly most powerful* (UMP) test.



A test  $\delta^*$  is said to be UMP in the class  $\mathcal{K}_\alpha$  if for all  $\delta \in \mathcal{K}_\alpha$ , for all  $\theta \in \Theta_1$ ,  $\beta_{\delta^*}(\theta) \geq \beta_\delta(\theta)$ . Unfortunately, UMP tests scarcely exist, except that when the parameter  $\theta$  is scalar, the family of distributions  $\mathcal{P} = \{P_\theta\}$  has a monotone likelihood ratio and the test is one sided, namely,  $\mathcal{H}_0 : \{\theta \leq \theta_0\}$  and  $\mathcal{H}_1 : \{\theta > \theta_1\}$ , with  $\theta_1 \geq \theta_0$  [12, 13]. In the case of a vector parameter  $\theta$ , the crucial issue is to find an optimal solution over a set of alternatives which is too rich. To overcome this difficulty in the case of the parametric family  $\mathcal{P} = \{P_\theta\}$ , Wald [14] proposes to impose an additional constraint on the class of considered tests, namely, a *constant power function* over a family of surfaces  $\mathcal{S} = \{S_\lambda, \lambda > 0\}$  defined on the parameter space  $\Theta_1$ , in order to avoid the existence of UMP tests over a subspace  $\bar{\Theta}$  of  $\Theta_1$  which are very inefficient over  $\Theta_1 \setminus \bar{\Theta}$ . A test  $\delta^* \in \mathcal{K}_\alpha$  is said to have uniformly best constant power (UBCP) on the family of surfaces  $\mathcal{S}$ , if the following conditions are fulfilled [14]:

- (1) for any pair of points  $\theta_1$  and  $\theta_2$  which lies on the same surface  $S_\lambda \in \mathcal{S}$ ,  $\beta_{\delta^*}(\theta_1) = \beta_{\delta^*}(\theta_2)$ , where  $\beta_\delta(\theta) = \Pr_\theta(\delta = \mathcal{H}_1)$  is the power function of the test  $\delta$ ,
- (2) for another test  $\delta \in \mathcal{K}_\alpha$ , which satisfies the previous condition, we have  $\beta_{\delta^*}(\theta) \geq \beta_\delta(\theta)$ .

It is worth emphasizing that the choice of the family of surfaces is naturally imposed by the statistical nature of the treated problem. For example, in the case of linear Gaussian models, the family  $\mathcal{S}$  is composed of ellipsoids in the informative parameter space  $\mathbb{R}^p$ . The presence of nuisance parameters transforms this family of ellipsoids into a family of elliptic cylinders in the observation space  $\mathbb{R}^n$  (see details in [15]). To simplify the presentation, the mathematical details of the nuisance parameter rejection and Wald theory, applied to the hypotheses testing problem (12) and (13), are moved to Appendices A and B.

### 3.2. Generalized likelihood ratio test

The GLR test is based on the following ratio:

$$\hat{\delta}(Y) = \begin{cases} \mathcal{H}_0 & \text{if } \hat{\Lambda}(Y) = 2 \log \frac{\sup_{\theta \in \mathbb{R}^p, \mu \in \mathbb{R}^m} \varphi_{\theta, \mu}(Y)}{\sup_{\mu \in \mathbb{R}^m} \varphi_{\theta=0, \mu}(Y)} < \gamma_\alpha, \\ \mathcal{H}_1 & \text{else,} \end{cases} \quad (14)$$

where

$$\varphi_{\theta, \mu}(Y) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left\{ -\frac{1}{2\sigma^2} \|Y - M\theta - H\mu\|_2^2 \right\} \quad (15)$$

is the pdf of  $Y$ ,  $\|X\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$  is the norm of  $X$ , and the threshold  $\gamma_\alpha$  is chosen to satisfy the definition of the class  $\mathcal{K}_\alpha$ . When this ratio is relatively large, the probability associated to the parameter of interest  $\theta$  being in the set defined by  $\mathcal{H}_1 : \{\theta \neq 0, \mu \in \mathbb{R}^m\}$  (the anomaly is superimposed on the environment) is relatively high and when the ratio is relatively small, the probability that the environment is anomaly-free (null hypothesis  $\mathcal{H}_0 : \{\theta = 0, \mu \in \mathbb{R}^m\}$  is true) is relatively high. The precise optimal properties of the GLR tests in the general case are unknown, but for some special

cases, the GLR test is optimal. We show now that the GLR test given by (14) coincides with the invariant UBCP test  $\Lambda(Y)$  defined in Appendices A and B. We assume that  $\tilde{H} = (M \ H)$  is a full-rank matrix. A simple algebra shows that

$$\hat{\Lambda}(Y) = -\frac{1}{\sigma^2} \|Y - M\hat{\theta}_1 - H\hat{\mu}_1\|_2^2 + \frac{1}{\sigma^2} \|Y - H\hat{\mu}_0\|_2^2 \quad (16)$$

$$= \frac{1}{\sigma^2} Y^T (P_H - P_{M,H}) Y \quad (17)$$

$$= \frac{1}{\sigma^2} Y^T P_H M (M^T P_H M)^{-1} M^T P_H Y = \Lambda(Y), \quad (18)$$

where the ML estimates  $\hat{\mu}_0$ ,  $\hat{\mu}_1$ , and  $\hat{\theta}_1$  are defined by using the least-squares (LS) algorithm:

$$\{\hat{\mu}_0\} = \arg \min_{\mu \in \mathbb{R}^m} \|Y - H\mu\|_2^2, \quad (19)$$

$$\{\hat{\theta}_1, \hat{\mu}_1\} = \arg \min_{\theta \in \mathbb{R}^p, \mu \in \mathbb{R}^m} \|Y - M\theta - H\mu\|_2^2,$$

$P_H = I_n - H(H^T H)^{-1} H^T$ ,  $P_{M,H} = I_n - \tilde{H}(\tilde{H}^T \tilde{H})^{-1} \tilde{H}^T$ , and  $A^{-1}$  is the inverse matrix of  $A$ . By comparing (18) and (B.3), it can be concluded that the GLR test is an invariant UBCP test. More details can be found in [16].

### 3.3. A priori information on the anomaly model

We briefly discuss the role of the matrix  $M$ . The  $N = \mathbb{E}(Y)$  parameter space  $\mathbb{R}^n$  is composed of two orthogonal linear subspaces:  $R(H)$  and  $R(H)^\perp$  such that  $\dim R(H) = \text{rank } H = q$ . In the general case, the anomaly partly belongs to the subspace  $R(H)$  (this part of the anomaly is undetectable) and to the subspace  $R(H)^\perp$  (this part is detectable). This subspace of detectable anomalies  $R(H)^\perp$  is spanned by the columns of the matrix  $W^T$  defined by (A.4) in Appendix A. If no a priori information on the anomaly is available, it is natural to put  $M = W^T$  and  $\theta \in \mathbb{R}^{n-q}$ , that is, the parameter  $\theta$  describes only the detectable part of the anomaly. The definition of the decision function  $\hat{\Lambda}(Y) = (1/\sigma^2) Y^T P_H Y$ , and the associated family of surfaces  $\mathcal{S} = \{S_\lambda : (1/\sigma^2) \|\theta\|_2^2 = \lambda, \lambda \in ]0; +\infty[ \}$  are especially simple in this case. As follows from Appendix B (see (B.5) and (B.6)), the properties of the test  $\hat{\delta}$  depend on the matrix  $M$ . Hence, if some a priori information on the anomaly is available, a parameterization  $M\theta$  with a specially chosen matrix  $M$  can be proposed. For example, the matrix  $M$  can serve to define the a priori known location of the anomaly on the projection  $Y$  or to reduce the uncertainty of the anomaly description.

### 3.4. Numerical study of statistical performances of the test

In some practical applications (e.g., detecting the welding defects in radiographic images [7, 8]), the anomaly-free environment is approximated by a polynomial function. We examine the statistical properties of the proposed test by using the following numerical example. The anomaly-free environment is defined by  $h(x, y) = 1 + 2x + 5y - 30x^2 + 20y^2 - 50xy$ , where  $(x, y) \in \mathcal{D} = \{(x, y) \in \mathbb{R}^2 \mid -10 \text{ cm} \leq x, y \leq 10 \text{ cm}\}$ .

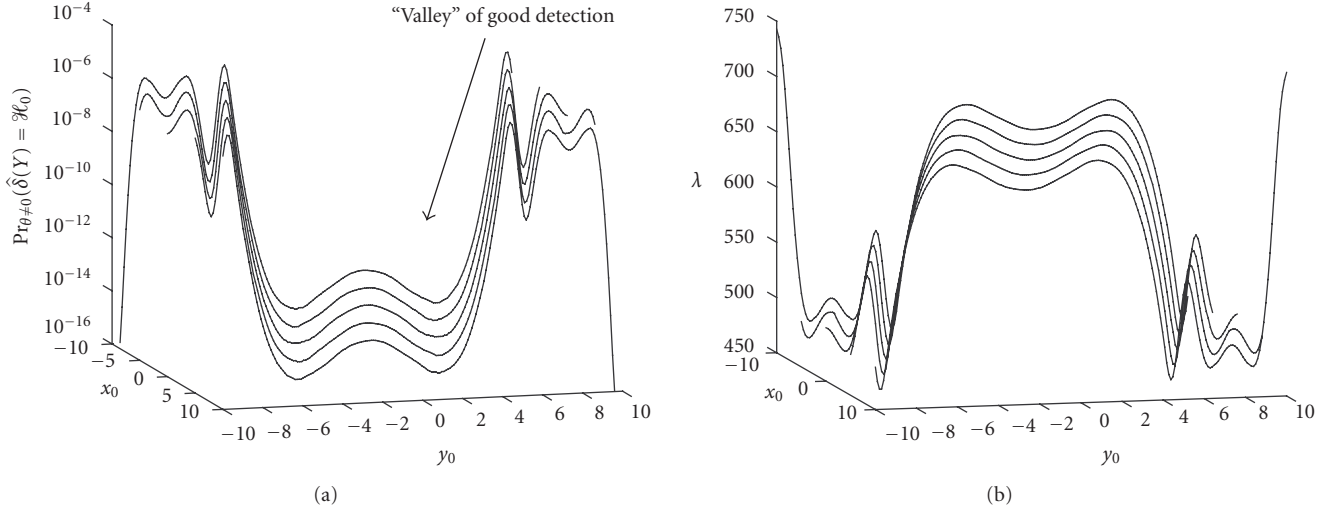


FIGURE 3: (a) The probability of nondetection  $\Pr_{\theta \neq 0}(\hat{\delta}(Y) = \mathcal{H}_0)$  as a function of the anomaly location. (b) The noncentrality parameter  $\lambda$  of  $\chi^2_{n-m,\lambda}$  as a function of the anomaly location.

Hence,  $m = 6$ . The anomaly is a small constant attenuation coefficient disk of radius  $r = 0.5$  cm located at the position  $(x_0, y_0)$ . The considered tomographic projections  $Y$  are taken at the angle of view  $\omega = 75^\circ$ . The number of observations  $n$  is equal to 200, which corresponds to a constant sampling step of 0.05 cm. The  $(200 \times 6)$  matrix  $H$  is full column rank,  $M = W^T$ , and  $\theta \in \mathbb{R}^{194}$ . The probability of false alarm  $\alpha = \Pr_0(\hat{\Lambda}(Y) \geq \gamma_\alpha)$  is equal to  $10^{-2}$ . The variance  $\sigma^2$  is equal to 179. First, we compute the probability of nondetection  $1 - \beta_{\hat{\delta}}$  as a function of the anomaly location  $(x_0, y_0) : (x_0, y_0) \mapsto 1 - \beta_{\hat{\delta}}(x_0, y_0)$ . It is assumed that  $x_0$  and  $y_0$  vary from  $-10$  cm to  $10$  cm. The numerical results are shown in Figure 3a by using a base-10 logarithmic scale to attenuate their variations. A “valley” of good detection clearly appears on the surface. It corresponds to the disk locations  $(x_0, y_0)$  where the environment is rejected with a great efficiency. In contrast to this “valley,” the edges of the surface  $(x_0, y_0) \mapsto 1 - \beta_{\hat{\delta}}(x_0, y_0)$  correspond to locations where the anomaly is well hidden by the environment. To explain this phenomenon, we consider the variations of the noncentrality parameter  $\lambda = (1/\sigma^2)\theta^T M^T P_H M \theta$  of the  $\chi^2_{n-m,\lambda}$  distribution of the statistics  $\hat{\Lambda}(Y)$  under the hypothesis  $\mathcal{H}_1$  (see Figure 3b). The noncentrality parameter can be rewritten as  $\lambda = (1/\sigma^2)\tilde{M}^T(I_n - \tilde{H})\tilde{M}$ , where  $\tilde{M} = M\theta$  and  $\tilde{H} = H(H^T H)^{-1}H^T$ . Assuming that the anomaly  $\tilde{M}$  differs from zero for only  $k$  adjacent positions (the anomaly is “compact”), it can be written as  $\tilde{M} = (0, \dots, 0, \tilde{m}_p, \dots, \tilde{m}_{p+k-1}, 0, \dots, 0)^T$ , where  $p$  is the subscript of the first nonzero component of  $\tilde{M}$ . It follows immediately that

$$\lambda = A(\tilde{M}, \sigma^2) - E(\tilde{M}, \tilde{H}, \sigma^2, p), \quad (20)$$

where  $A(\tilde{M}, \sigma^2) = (1/\sigma^2) \sum_{i=p}^{p+k-1} \tilde{m}_i^2$ ,  $E(\tilde{M}, \tilde{H}, \sigma^2, p) = (1/\sigma^2) \sum_{i=p}^{p+k-1} \sum_{j=p}^{p+k-1} \tilde{m}_i \tilde{m}_j \tilde{h}_{i,j}$ , and  $\tilde{h}_{i,j}$  is the element  $(i, j)$

of the matrix  $\tilde{H}$ . For a given angle of view  $\omega$ , the anomaly projection has a constant geometrical shape for all possible locations of the anomaly. Hence, the finite set  $\tilde{m}_p, \tilde{m}_{p+1}, \dots, \tilde{m}_{p+k-1}$  is independent of the first subscript  $p$ , and the positive term  $A(\tilde{M}, \sigma^2)$  represents the constant anomaly impact on the parameter  $\lambda$ . In contrast to this first term, the second nonnegative term  $E(\tilde{M}, \tilde{H}, \sigma^2, p)$  (the matrix  $\tilde{H}$  is positive semidefinite) clearly depends on the subscripts  $p, p+1, \dots, p+k-1$  through the elements  $\tilde{h}_{i,j}$  of the matrix  $\tilde{H}$ . Hence, a translation of the anomaly leads to a variation of the second term  $E(\tilde{M}, \tilde{H}, \sigma^2, p)$ . Therefore, a change of the anomaly location can imply a drastic variation of the probability of nondetection for the test  $\hat{\delta}$ . In particular, according to the choice of the angle of view and the physical nature of the environment, some “valleys” of good (or “plateaus” of bad) detection (see Figure 3a) can appear.

Second, we compute the probability of nondetection  $1 - \beta_{\hat{\delta}}$  of the disk as a function of its radius  $r$  and constant contrast (i.e. local variation of the anomaly attenuation coefficient)  $f_{g,\mu}(x, y) = f_d$  with respect to the environment. The numerical results are shown in Figure 4 by using a base 10 logarithmic scale. The radius  $r$  varies from 0.1 cm to 0.45 cm and the contrast varies from  $10^2$  to  $10^4$  for the X-ray attenuation coefficient varying from  $-6 \times 10^5$  to  $4 \times 10^5$ . Obviously, the probability of nondetection  $1 - \beta_{\hat{\delta}}(r, f_d)$  decreases with the increasing radius  $r$  and/or contrast  $f_d$  of the disk.

### 3.5. Computational cost of the anomaly detection algorithm

To evaluate the computational cost of the decision procedure, the number of floating-point operations (flops) of different steps of the anomaly detection algorithm (14), (16), (17), and (18) is summarized in Table 1. The singular value decomposition (SVD) is the most costly computational operation.

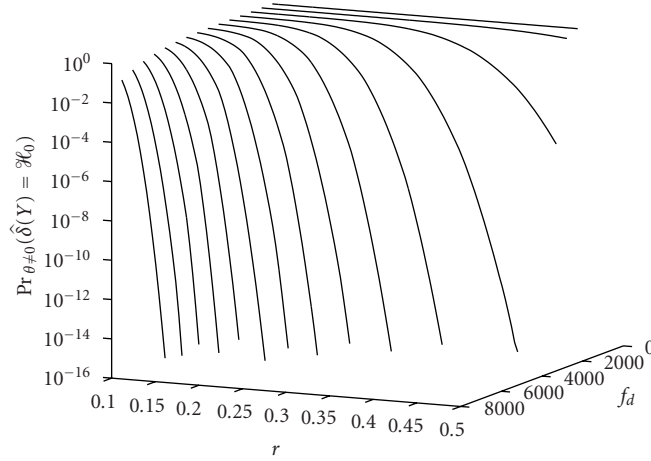


FIGURE 4: The probability of nondetection  $\Pr_{\theta \neq 0}(\hat{\delta}(Y) = \mathcal{H}_0)$  as a function of the radius  $r$  and the contrast  $f_d$  characterizing the anomaly.

For example, if the full-column-rank matrix  $H$  is chosen so that  $n = 200$  and  $m = 6$ , the decision function calculation takes about 0.01 second on a personal computer with a 2 GHz processor and 256 MB RAM by using the Matlab computing software.

#### 4. STATISTICAL HYPOTHESES TESTING: NONLINEAR CASE

We consider now a more realistic experimental context when the geometry of the acquisition system and/or imaged object is partially unknown. First, it is assumed that the nonlinear parsimonious decomposition  $h_\mu$  perfectly fits the unknown environment:  $h_\mu(x, y) = h(x, y)$ . The nonlinear GLR test is introduced and the difficulties induced by the nonlinear decomposition  $h_\mu$  are discussed. Finally, some remarks of the applicability and limits of the nonlinear parsimonious decomposition are discussed.

##### 4.1. Nonlinear nuisance parameters and GLR test

In the previous sections, we have considered a linear model of observations with respect to the nuisance parameter given by (5). From the practical point of view, this means that the geometry of the acquisition system and/or original scene is perfectly known. A more realistic experimental context corresponds to a partially unknown geometry. Hence, we consider the nonlinear parameterized parsimonious model

$$Y = M\theta + H(\mu) + \xi, \quad (21)$$

where  $\mu \mapsto H(\mu)$  is a *known* vector function of the *unknown* parameter  $\mu \in \mathbb{R}^m$ . The anomaly detection problem consists in testing between  $\mathcal{H}_0$  and  $\mathcal{H}_1$  defined by (13) while considering  $\mu$  as an unknown vector. In the case of nonlinear model, a simple algebra shows that the statistics  $\hat{\Lambda}$  of the GLR test

(14) can be rewritten as follows:

$$\begin{aligned} \hat{\Lambda}(Y) &= 2 \log \frac{\sup_{\theta \in \mathbb{R}^p, \mu \in \mathbb{R}^m} \varphi_{\theta, \mu}(Y)}{\sup_{\mu \in \mathbb{R}^m} \varphi_{\theta=0, \mu}(Y)} \\ &= -\frac{1}{\sigma^2} \|Y - M\hat{\theta}_1 - H(\hat{\mu}_1)\|_2^2 + \frac{1}{\sigma^2} \|Y - H(\hat{\mu}_0)\|_2^2, \end{aligned} \quad (22)$$

where

$$\begin{aligned} \{\hat{\mu}_0\} &= \arg \min_{\mu \in \mathbb{R}^m} \|Y - H(\mu)\|_2^2, \\ \{\hat{\theta}_1, \hat{\mu}_1\} &= \arg \min_{\theta \in \mathbb{R}^p, \mu \in \mathbb{R}^m} \|Y - M\theta - H(\mu)\|_2^2. \end{aligned} \quad (23)$$

As it follows from (22) and (23), the calculation of the statistics  $\hat{\Lambda}(Y)$  is based on two nonlinear LS algorithms. Under some regularity conditions (see [17, 18] for details), the estimates  $\hat{\mu}_0$ ,  $\hat{\theta}_1$ , and  $\hat{\mu}_1$  exist and are unique. In the case of a small anomaly  $\theta$  detection and/or slightly nonlinear model, the minimization methods based on the model linearization are usually used. The impact of the term  $M\theta$  on the linearization working point is very limited and, exactly as in the linear case, the statistics  $\hat{\Lambda}(Y)$  is computed by using a single minimization problem  $\arg \min_{\mu \in \mathbb{R}^m} \|Y - H(\mu)\|_2^2$  and a standard Gauss-Newton algorithm [18, page 619]. It is assumed that the starting point  $\mu_0$  initializing the algorithm is good enough to ensure the convergence of this method. These two conditions are verified in the practical example of Section 5.

##### 4.2. Difficulties induced by a nonlinear nuisance parameter: biased residuals

In contrast to the linear case, where the ML estimation of the nuisance parameter  $\mu$  is unbiased under the hypothesis  $\mathcal{H}_0$ , the nonlinear parameterization leads to a biased estimation of  $\mu$  and, hence, to a positive noncentrality parameter of the  $\chi^2$  law of the statistics  $\hat{\Lambda}(Y)$  under the hypothesis  $\mathcal{H}_0$ . The same difficulty takes place under the hypothesis  $\mathcal{H}_1$ . These biases could considerably modify the statistical properties of the test. For example, it may increase the probabilities of false alarm or nondetection. As it follows from [17], [18, page 127], if the number  $n$  of sensors by projection becomes large then the bias  $\hat{b} = \mathbb{E}(\hat{\varepsilon})$  of the residuals,  $\hat{\varepsilon} = Y - M\hat{\theta} - H(\hat{\mu})$  can be estimated. To quantify the nonlinearity (curvature) of the regression model, two different measures are usually used [18, page 127]: the *intrinsic curvature* and the *parameter-effect curvature*, which describes the degree of curvature induced by the choice of the parameterization  $\mu$ . In particular, the bias  $\hat{b}$  is essentially connected to the intrinsic curvature. When the number  $n$  becomes large, a practical method of the bias  $\hat{b}$  estimation consists in a second-order expansion of  $\hat{b}$ , namely, it is possible to estimate  $\hat{b}$  by considering second-order derivatives of  $\mu \mapsto H(\mu)$ . More details can be found in [17, page 116].

As it follows from [19], if the nuisance parameter  $\mu$  belongs to a compact set, then the impact of the nonlinearity

TABLE 1: Computational cost of the anomaly detection algorithm.

Step	Action	Number of flops
(1)	Initialize: – $H(n \times m \text{ matrix}), \sigma \text{ (real)}, \gamma_\alpha \text{ (real)}$	Negligible
(2)	Compute the decision function $\hat{\Lambda}(Y) = \frac{Y^T P_H Y}{\sigma^2}$ : – compute the SVD of $H : H = UDV^T$ – design $W$ by extracting $n - m$ columns of $V$ associated to singular value 0, – compute $Z = WY$ – compute $\hat{\Lambda}(Y) = \frac{Z^T Z}{\sigma^2}$	$4n^2m + 22m^3$ Negligible $2n(n - m)$ $2(n - m) + 2$
(3)	Check: – if $\hat{\Lambda}(Y) \geq \gamma_\alpha$ , then declare anomaly	Negligible

$\mu \mapsto H(\mu)$  can be estimated by an additional bias of residuals. Therefore, the noncentrality parameter  $\lambda$  of the statistics  $\hat{\Lambda}(Y)$  is equal to the sum of two components:

$$\lambda = \lambda(\theta, \mu) = \lambda_r(\theta) + \Delta\lambda(\theta, \mu), \quad (24)$$

where  $\lambda_r(\theta)$  is the regular component due to the linearization of the parsimonious model around a “working point” and  $\Delta\lambda(\theta, \mu)$  is the component due to the nonlinearity of the model. In the case of a linear parsimonious model (see Section 3), the GLR test is optimal (invariant UBCP). We consider the potential loss of optimality of the nonlinear GLR test given by (22) and (23). The upper and lower bounds of the power function  $\beta(\theta)$  of the nonlinear GLR test have been established in [19]. Therefore, the potential loss of optimality can be bounded that leads to the so-called  $\varepsilon$ -optimality of the proposed solution. In Section 5, such an approach is applied to the problem of solid rocket motor examination in order to check that an additional bias  $\Delta\lambda(\theta, \mu)$  due to the nonlinearity of the parsimonious model is negligible.

#### 4.3. Parameterized parsimonious approximation: applicability and limits

We finish the theoretical discussion of the proposed method with some remarks on the applicability of the (non-) linear parameterized parsimonious approximation  $(x, y) \mapsto h_\mu(x, y)$  of the unknown environment  $(x, y) \mapsto h(x, y)$ . First, in the case of linear parsimonious approximation of the environment  $h$  (see (5)), the choice of basic functions  $\{h_1, \dots, h_m\}$  is crucially important but there is no general way to make this choice. In practical cases, a tradeoff between the quality of approximation and its complexity should be established: some guiding principles can be found in [20]. For example, it is natural to approximate a rotation invariant scene by radial functions.

Second, we assume that the environment is approximated with a certain error  $(x, y) \mapsto \psi(x, y)$ , that is,  $h(x, y) = h_\mu(x, y) + \psi(x, y)$ . Because of mismatches between  $h_\mu$  and  $h$ , the line integral projection of  $\psi(x, y)$  leads to an additional

term in the measurement model given by (11), namely,

$$Y_\omega = \begin{cases} H_\omega(\mu) + R_\omega(\psi) + \xi & \text{under } \mathcal{H}_0, \\ R_\omega(g, \mu) + H_\omega(\mu) + R_\omega(\psi) + \xi & \text{under } \mathcal{H}_1, \end{cases} \quad (25)$$

where  $R_\omega(\psi) = R_{\omega, \tau}(\psi)$  is the discrete Radon transform of the approximation error  $\psi$ . It follows from (25) that the nonlinear parameterized parsimonious approximation (21) has to be rewritten in the following way:

$$Y = M\theta + H(\mu) + \Psi + \xi, \quad (26)$$

where the vector  $\Psi \in \mathbb{R}^n$  represents the discrete Radon transform of approximation errors. To simplify the presentation, we assume that  $H(\mu) = H\mu$ . By rejecting the nuisance (see Appendices A and B) in the above model, we get  $Z = WY = WM\theta + W\Psi + \zeta$  (compare with (B.1)). The mismatches between  $h_\mu$  and  $h$  lead to the noncentrality parameters  $\lambda_0 = (1/\sigma^2)\Psi^T P_H \Psi$  under  $\mathcal{H}_0$  and  $\lambda_1 = (1/\sigma^2)[M\theta + \Psi]^T P_H [M\theta + \Psi]$  under  $\mathcal{H}_1$ . Therefore, the negative impact of mismatches between  $h_\mu$  and  $h$  on the statistical performances of the GLR test consists in increasing the probability of false alarms (with respect to a prescribed level  $\alpha$ ) and in the reduction of the test power  $\beta(\theta)$ . Putting together the above-mentioned expressions for the noncentrality parameters and the approximation error projection  $\Psi = \Psi(\psi(x, y))$  provides a means to estimate the applicability of the proposed detection method and to define its limits. If in some practical cases, the unknown environment  $h$  can be perfectly defined by a parameterized function, which is not necessarily a *parsimonious* decomposition, or it can be accessible via a discrete sample on a compact support, then a linear approximation with a limited impact on the probabilities of false alarm and nondetection can be found (see details in [19]).

## 5. PRACTICAL EXAMPLES: SOLID ROCKET MOTOR EXAMINATION

To illustrate the developed theory, the problem of solid rocket motor examination is discussed now. First, the geometry



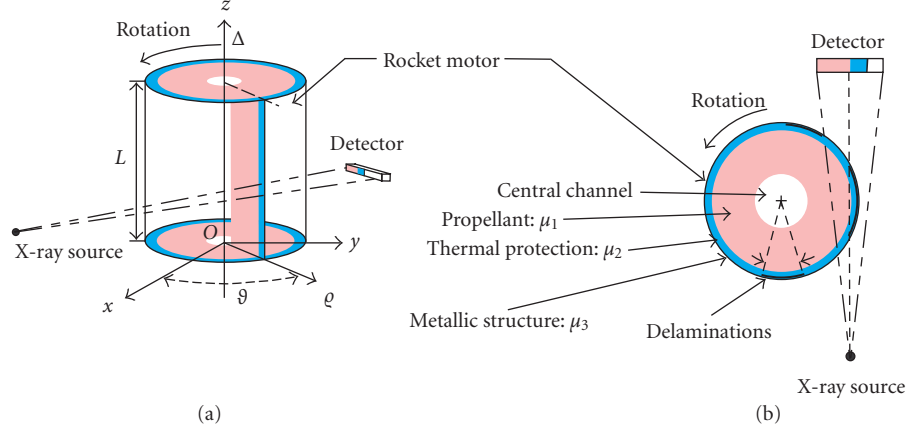


FIGURE 5: (a) Tomosynthesis tangential geometry. (b) Horizontal slice of the rocket motor.

of the acquisition system/object is supposed to be perfectly known. The linear measurement model and the GLR test are used. Next, the geometry of the acquisition system/object is assumed to be partially unknown. The nonlinear GLR test is used here and the impact of nonlinearity is shown to be negligible.

### 5.1. Description of the practical problem

A typical scenario of the solid rocket motor examination can be described as follows (see details in [21]). During the monitoring process, the rocket motor turns around its vertical axis \$\Delta\$ and it is imaged slice by slice along the vertical axis. The projection image is the radiograph of a tangential area of the cylindrical object (see Figure 5a). Hence, the tomographic system provides the researchers only with “local” projections of a tangential part of the motor. The revolution axis of the motor is outside of the field of view. Such a procedure is also known as a tangential tomosynthesis. Using cylindrical coordinates denoted by \$(\varrho, \vartheta, z)\$ (see Figure 5a), it is natural to parameterize the motor, namely \$h(\varrho, \vartheta, z)\$, as follows:

$$h(\varrho, \vartheta, z) = \begin{cases} 0 & \text{if } 0 \leq \varrho < r_1, \\ \mu_1 & \text{if } r_1 \leq \varrho < r_2, \\ \mu_2 & \text{if } r_2 \leq \varrho < r_3, \\ \mu_3 & \text{if } r_3 \leq \varrho < r_4, \\ 0 & \text{if } \varrho \geq r_4, \end{cases} \quad z \in [0; L], \vartheta \in [0, 2\pi[, \quad (27)$$

where \$\mu\_1\$, \$\mu\_2\$, and \$\mu\_3\$ are the attenuation coefficients of the propellant, thermal protection, and metallic structure, respectively (see Figures 6a and 6b). The radiuses \$r\_1 = 50\$ cm, \$r\_2 = 146\$ cm, \$r\_3 = 149\$ cm, and \$r\_4 = 150\$ cm define the cylindrical boundaries of each homogeneous part of the rocket motor (see Figure 6). The X-source is assumed to be infinitely far from the rocket motor and, hence, the tangential radiographies can be considered as parallel-beam line-integral X-ray projections (defined by (2)).

### 5.2. Anomaly detection with a known motor/acquisition system geometry

For this simulation, it is assumed that the center of revolution coordinates \$(x\_0, y\_0)\$, the radiuses \$r\_1, r\_2, r\_3, r\_4\$, and the detector position are perfectly known but the attenuation coefficients \$\mu\_1, \mu\_2\$, and \$\mu\_3\$ are unknown for the algorithm. We define the function \$h\_i\$, for \$1 \leq i \leq m = 3\$, as follows:

$$h_i(\varrho, \vartheta, z) = \begin{cases} 1 & \text{if } r_i \leq \varrho < r_{i+1}, 0 \leq z \leq L, \\ 0 & \text{else.} \end{cases} \quad (28)$$

Hence, the background can be linearly parameterized by using the decomposition:

$$h(\varrho, \vartheta, z) = \sum_{i=1}^3 \mu_i h_i(\varrho, \vartheta, z). \quad (29)$$

First, the efficiency of the proposed statistical tools is studied in the case of a single projection (linear X-ray detector) by using numerical simulations and, next, in the case of several projections.

#### 5.2.1. Detection with a single projection

The goal is to detect a “delamination,” that is, a lack of material into the thermal protection (the delamination attenuation coefficient is \$\mu = 0\$), between the metallic structure and the peripheral thermal protection of the motor (see Figure 6b). The delamination is assimilated to a limited portion (\$\simeq 5^\circ\$) of circle centered at the revolution axis. Its radial size is about 0.3 mm. It is assumed that the decision is directly based on a single projection (see Figure 6b). The linear X-ray detector is composed of \$n = 200\$ elementary sensors. The known geometry allows us to use the linear GLR test (14), (16), (17), and (18) developed in Section 3.

The typical behavior of a single tangential projection (without and with delamination) and its “parity vector” \$Z = WY\$ is depicted in Figure 7 to show the capacity of the proposed test to eliminate the unknown background corresponding to the metallic structure, thermal protection, and

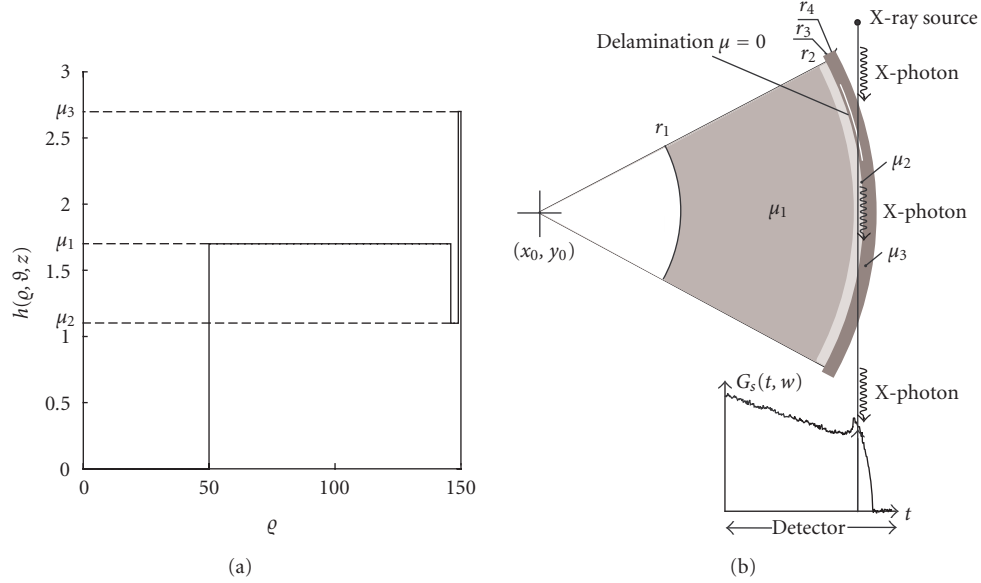


FIGURE 6: (a) Attenuation coefficient  $h(\rho, \theta, z)$  as a function of the radius  $\rho$  for  $z \in [0, L]$  and  $\theta \in [0, 2\pi[$ . (b) Typical tangential projection of rocket motor.

propellant. The threshold  $\gamma_\alpha$  of the GLR test (14), (16), (17), and (18) is chosen to satisfy a preassigned level of false alarms  $\alpha = 0.01$ . We define the signal-to-noise ratio (SNR) in the following manner (typical for CT):  $\text{SNR} = \|Y^t\|_2 / (\sigma\sqrt{n})$ , where  $Y^t$  is the theoretical projection of the scene without any noise. It is assumed that the variance  $\sigma^2$  is preliminarily estimated with a high precision during the calibration procedure of the acquisition system: test object is imaged and the variance is calculated from repetitive resulting radiographies. The power  $\beta_\delta$  of the GLR test (14), (16), (17), and (18) as a function of the SNR is depicted in Figure 8 (see the line  $K = 1$ ).

### 5.2.2. Detection with several projections

We suppose now that several projections  $Y_1, \dots, Y_K$  for  $K$  different view angles  $\omega_i$  or horizontal motor slices are available. To assure the comparison between the results of detection with different numbers  $K$ , it is assumed that the theoretical profile  $Y^t$  holds constant for all tangential projections. As it follows from Section 2.4, it is possible to design a “cumulated” projection  $Y_1, \dots, Y_K$ . The power  $\beta_\delta$  of the GLR test (14), (16), (17), and (18) as a function of the SNR for different numbers  $K$  of “cumulated” projections is depicted in Figure 8. It is easy to see that the accumulation of several projections considerably increases the power  $\beta_\delta$  of the test.

### 5.3. Anomaly detection with an unknown motor/acquisition system geometry

We assume now that the center of revolution coordinates  $(x_0, y_0)$  and the detector position are perfectly known, but the radii  $r_1, r_2, r_3$ , and  $r_4$  are only approximately known. The measurement model is then nonlinear. The efficiency of the statistical tools proposed in Section 4 is examined now in

the case of planar detector. Due to the quality of the motor production process, the radii  $r_i$  vary into confidence intervals  $I_i = [\bar{r}_i - \varepsilon_i; \bar{r}_i + \varepsilon_i]$ , where  $\bar{r}_i$  is the “prescribed” radius and  $\varepsilon_i$  is a small positive constant provided by the designer of rocket motors. Let  $X = (\mu_1, \mu_2, \mu_3, r_2, r_3, r_4)^T$  be the vector of unknown nuisance parameters (the radius  $r_1$  is out of the field of the projection). The nonlinear measurement model is defined as follows:

$$Y = M\theta + H(X) + \xi, \quad (30)$$

where the projection  $H(X) = R_{\omega, \tau}(h_X)$  is defined for a particular angle of view  $\omega$  and a family of sampling coordinates  $\tau$  (see (2)),  $h_X$  is a nonlinear model of the motor, parameterized by  $X$ ,  $\theta$  is a vector corresponding to a delamination, and  $\xi$  is a zero-mean Gaussian white noise,  $\xi \sim \mathcal{N}(0, \sigma^2 I_n)$  with  $\sigma^2$  known. It is assumed that  $M = W_0^T$  and the matrix  $W_0 = W(J_H(X_0))$  satisfies the conditions given by (A.4), where  $H$  is replaced by the Jacobian matrix of size  $(n \times m)$   $J_H(X_0) = \partial H / \partial X|_{X=X_0}$  calculated at the working point  $X_0$ .

Assuming that the height of the planar detector is small enough relatively to the height of the motor, it is assumed that all the lines of the planar detector correspond to the same radii  $r_2, r_3$ , and  $r_4$ , which vary into confidence intervals  $I_2, I_3$ , and  $I_4$ , respectively. To warrant the existence of a nonlinear detection problem solution [17, 18], the nonlinear LS estimate  $\hat{X}$  of the parameter  $X$  is calculated from the vector of observations  $Y = (G_s(t_1, \omega), G_s(t_2, \omega), \dots, G_s(t_n, \omega))^T$  (see (2)) limited to the components  $G_s(t_i, \omega)$  such that  $t_i \in \hat{I}$  with  $\hat{I} = [\underline{q}; \bar{q}] \setminus (\cup_{i=2}^4 I_i)$ , where  $\underline{q}$  and  $\bar{q}$  are the minimal and maximal radii corresponding to the X-ray detector. The capacity of the proposed method to estimate the nonlinear model and to eliminate the unknown background corresponding to the metallic structure, thermal protection, and

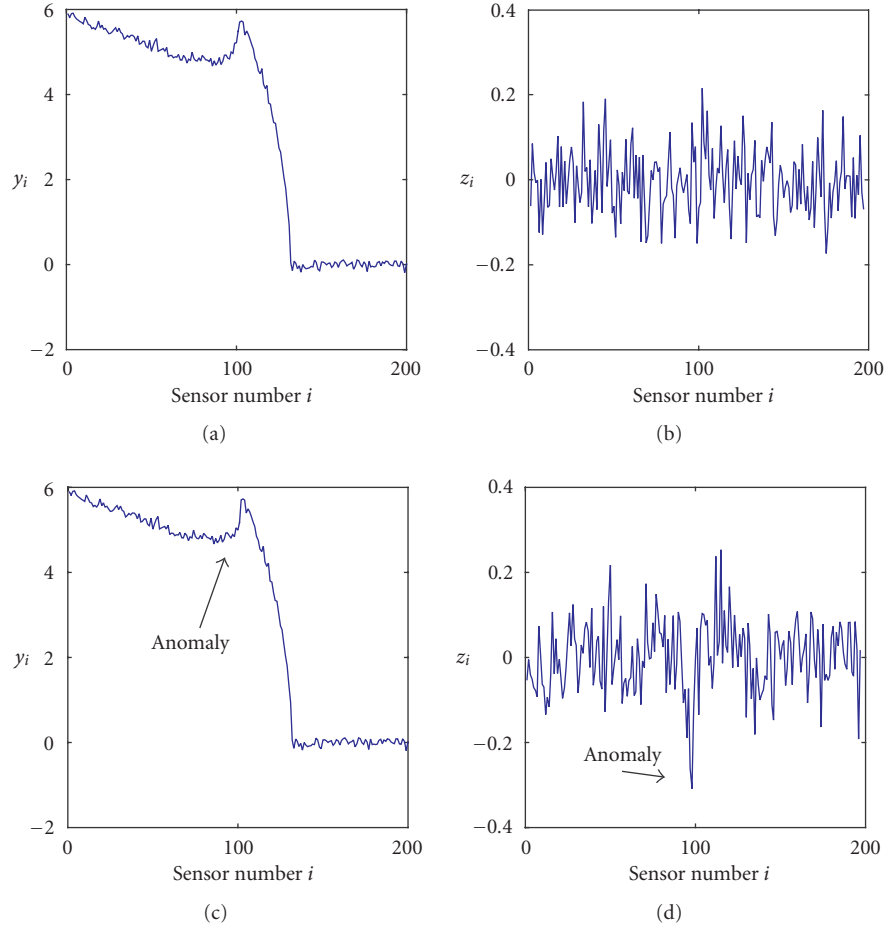


FIGURE 7: Detection with a single projection: (a) a tangential projection  $Y$  without delamination, (b) the parity vector  $Z = WY$  calculated from (a) ( $\mathcal{H}_0$  is accepted:  $\hat{\Lambda}(Y) = 173.63 < \gamma_\alpha$ ), (c) a tangential projection  $Y$  with a delamination, and (d) the parity vector  $Z$  calculated from (c) ( $\mathcal{H}_1$  is accepted:  $\hat{\Lambda}(Y) = 210.8 \geq \gamma_\alpha$ ).

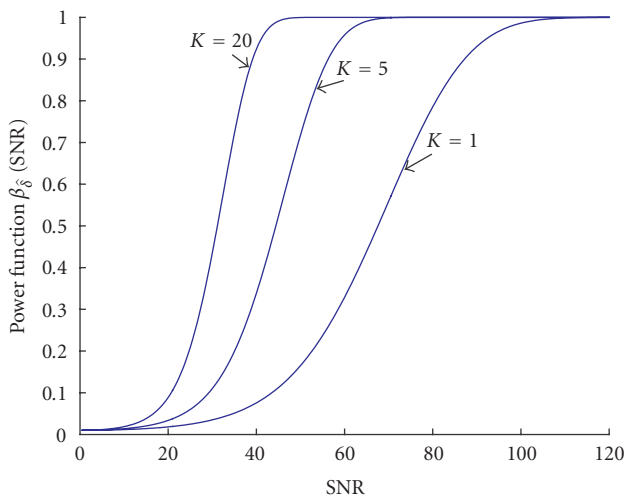


FIGURE 8: The power of test  $\beta_\delta(\text{SNR})$  (probability of the delamination detection) as a function of the SNR for different numbers  $K$  of “cumulated” projections.

propellant by using a planar X-ray detector is depicted in Figures 9 and 10. Here the simulation results for a planar detector of  $40 \text{ cm} \times 40 \text{ cm}$  composed of  $400 \times 400$  elementary sensors are presented. The delamination’s width and height are chosen  $390 \times 60 \text{ mm}$ , respectively. For the illustration purpose, to make the delamination more visible by eyes on the projection (see Figure 10), its radial size has been chosen of  $3 \text{ mm}$ . But the proposed algorithm is able to stably detect a delamination of  $0.3 \div 1 \text{ mm}$ . The threshold  $\gamma_\alpha$  of the nonlinear GLR test (14), (22), and (23) is chosen to satisfy a preassigned level of false alarms  $\alpha = 0.01$ .

To examine the impact of nonlinearity, the mean bias  $\hat{b}$  of the residual vector  $\hat{\epsilon} = Y - H(\hat{\mu})$  has been computed. The decision function  $\hat{\Lambda}(Y)$  (22) asymptotically follows the non-central  $\chi^2$  law with  $n - m$  degrees of freedom ( $n = 400^2$  and  $m = 6$ ) and the noncentrality parameter  $\lambda$ . The component of this parameter  $\Delta\lambda$  due to the nonlinearity of the parsimonious model (see (24)) constitutes the main degradation of the optimality properties of the nonlinear GLR test with respect to the linear one. A numerical study shows that the

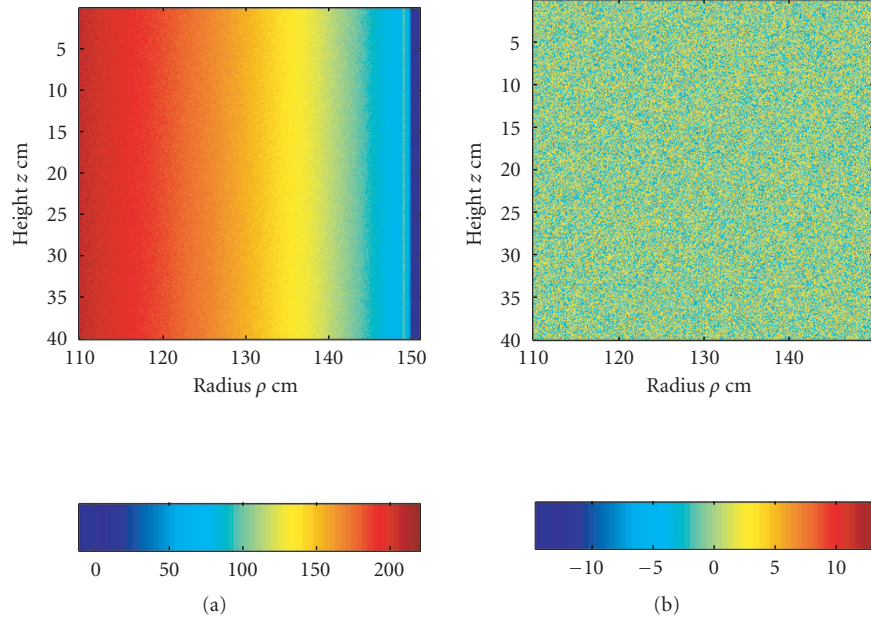


FIGURE 9: Detection with a planar detector of 40 cm  $\times$  40 cm: (a) typical projection of a tangential part of the motor without delamination, (b) parity vectors calculated from (a) ( $\mathcal{H}_0$  is accepted:  $\hat{\Lambda}(Y) = 1.6 \cdot 10^5 < \gamma_\alpha (= 1.657 \cdot 10^5)$ ).

ratio  $\Delta\lambda/\lambda$  is very small ( $\approx 7.1 \cdot 10^{-5}$ ) under  $\mathcal{H}_1$ . Hence, the variation of the power function of the test is negligible. Moreover, under  $\mathcal{H}_0$ , the parameter  $\Delta\lambda$  is also very small (less than 0.01) and, therefore, the probability of false alarm practically does not change.

## 6. CONCLUSION

The problem of anomaly detection in a parameterized parsimonious environment has been stated as a composite hypotheses testing problem with nuisance parameters. It has been shown that the GLR test coincides with an optimal UBCP invariant test in the linear case. The robustness of the GLR test for a nonlinearly parameterized environment has been numerically studied when the dimension of the projection is large. It has been shown that the impact of the nonlinearity on the detector optimality can be estimated by using the variation of the  $\chi^2$  noncentrality parameter with respect to its regular value due to the linearized model. The same noncentrality parameter has been used to estimate the impact of the approximation errors when the (non-) linear parsimonious decomposition does not perfectly fit the unknown environment. The relevance of the proposed method has been illustrated with the numerical simulation of the solid rocket motor examination.

## APPENDICES

### A. NUISANCE PARAMETERS REJECTION

The presence of the nuisance parameters certainly complicates the statistical decision problem. As it follows from

(12), the vector of observations  $Y$  is an additive sum of the nuisance  $H\mu$  and a possibly hidden anomaly  $M\theta$ . In the case of anomaly detection, we need to detect the anomaly  $\theta$  while considering the “environment”  $\mu$  as an unknown vector (or nuisance parameter). Since the nuisance parameter  $\mu \in \mathbb{R}^m$  is completely unknown, the test  $\delta$  should be independent of its value. To obtain a statistics which is independent of the nuisance parameter, the theory of invariance is used [13]. We consider now the hypotheses testing problem (12) and (13). We define two parameter sets

$$\begin{aligned}\Omega_0 &= \{N = H\mu + M\theta, \theta = 0, \mu \in \mathbb{R}^m\}, \\ \Omega_1 &= \{N = H\mu + M\theta, \mu \in \mathbb{R}^m, \theta \neq 0\}\end{aligned}\quad (\text{A.1})$$

in the parameter (expectation)  $N = \mathbb{E}(Y)$  space  $\mathbb{R}^n$  corresponding to the hypotheses  $\mathcal{H}_0$  and  $\mathcal{H}_1$ , respectively. First of all, we note that the family of distributions  $Y \sim \mathcal{N}(M\theta + H\mu, \sigma^2 I_n)$  remains invariant (see [13] for details and definitions) under the group of translations

$$G = \{g : g(Y) = Y + HC, C \in \mathbb{R}^m\}, \quad (\text{A.2})$$

which induces the group

$$\bar{G} = \{\bar{g} : \bar{g}(N) = N + HB, B \in \mathbb{R}^m\}, \quad (\text{A.3})$$

in  $\mathbb{R}^n$ , where  $N = \mathbb{E}(Y)$  is the expectation that preserves both  $\Omega_0$  and  $\Omega_1$ , that is,  $\bar{g}\Omega_0 = \Omega_0$  and  $\bar{g}\Omega_1 = \Omega_1$ . Hence, the hypotheses testing problem (12) and (13) remains invariant



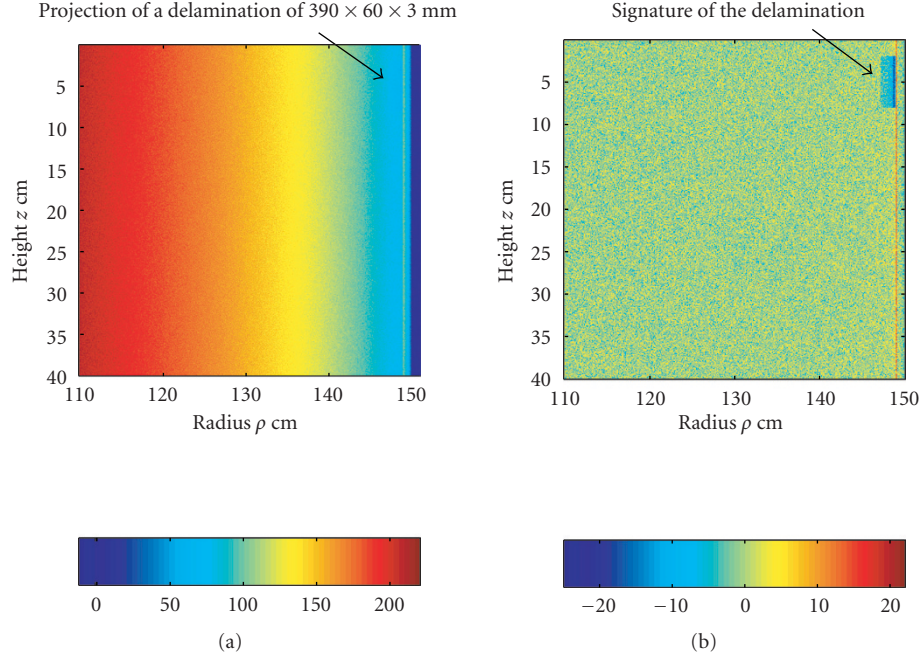


FIGURE 10: Detection with a planar detector of 40 cm  $\times$  40 cm: (a) a tangential projection with a delamination of 390  $\times$  60  $\times$  3 mm, (b) the parity vectors calculated from (a) with the signature of the delamination ( $\mathcal{H}_1$  is accepted:  $\hat{\Lambda}(Y) = 1.702 \cdot 10^5 > \gamma_\alpha (= 1.657 \cdot 10^5)$ ).

under  $G$ . The optimal invariant tests are based on the maximal invariants (principle of invariance) [13]. We define the column space  $R(H)$  of the matrix  $H$  with rank  $H = q$ . First, it is supposed that  $q = m$  ( $H$  is a full-rank matrix). The standard solution is the projection of  $Y$  on the orthogonal complement  $R(H)^\perp$  of the column space  $R(H)$ . The space  $R(H)^\perp$  is well known under the name “parity space” in the analytical redundancy literature (see details in [15, 16]). The parity vector  $Z = WY$  is the transformation of the measured output  $Y$  into a set of  $n - m$  linearly independent variables by projection onto the left null space of the matrix  $H$ . The matrix  $W^T = (w_1, \dots, w_{n-m})$  of size  $n \times (n - m)$  is composed of the eigenvectors  $w_1, \dots, w_{n-m}$  of the projection matrix  $P_H = I_n - H(H^T H)^{-1} H^T$  corresponding to eigenvalue 1. The matrix  $W$  satisfies the following conditions:

$$WH = 0, \quad W^T W = P_H, \quad W W^T = I_{n-m}. \quad (\text{A.4})$$

It follows from the first of the above conditions that the transformation by  $W$  completely removes the interference of the nuisance parameter  $\mu$ . It can be shown [15, 22] that the function  $M(Y) = Z = WY$  is maximal invariant to the group of translations  $G = \{g : g(Y) = Y + H\mu\}$ . For this reason, all invariant tests depend on  $Y$  only via the vector  $Z = WY$ . Second, we assume that rank  $H = q < m$  ( $H$  is not a full-rank matrix). The matrix  $W^T = (w_1, \dots, w_{n-q})$  of size  $n \times (n - q)$  is then composed of the eigenvectors  $w_1, \dots, w_{n-q}$  of the projection matrix  $P_H = I_n - H(H^T H)^- H^T$ , where  $A^-$  is a generalized inverse of  $A$  [23, Chapter 1.5.3], corresponding to eigenvalue 1.

## B. OPTIMAL INVARIANT TEST

As it follows from Appendix A, the measurement model  $Y = M\theta + H\mu + \xi$  can be rewritten in the following manner:

$$Z = WY = WM\theta + \zeta, \quad (\text{B.1})$$

where  $\zeta \sim \mathcal{N}(0, \sigma^2 I_{n-q})$ ,  $\sigma^2 > 0$ . As it follows from [15, 16, 22], the test  $\delta^*$  given by

$$\delta^*(Y) = \begin{cases} \mathcal{H}_0 & \text{if } \Lambda(Y) < \gamma_\alpha, \\ \mathcal{H}_1 & \text{else,} \end{cases} \quad (\text{B.2})$$

where the decision function is (under assumption that the matrix  $M^T P_H M$  is invertible)

$$\Lambda(Y) = \frac{1}{\sigma^2} Y^T P_H M (M^T P_H M)^{-1} M^T P_H Y, \quad (\text{B.3})$$

and the threshold  $\gamma_\alpha$  is chosen to satisfy the definition of the class  $\mathcal{K}_\alpha$ :

$$\sup_{\mu \in \mathbb{R}^m} \Pr_{0,\mu}(\delta^*(Y) = \mathcal{H}_1) = \Pr_0(\Lambda(Y) \geq \gamma_\alpha) = \alpha, \quad (\text{B.4})$$

is invariant and UBCP over the family of surfaces

$$\mathcal{S} = \left\{ S_\lambda : \frac{1}{\sigma^2} \theta^T M^T P_H M \theta = \lambda, \lambda \in ]0; +\infty[ \right\}. \quad (\text{B.5})$$

The statistics  $\Lambda$  is distributed according to the  $\chi^2$  law with  $n - q$  degrees of freedom. This law is central under  $\mathcal{H}_0$  and noncentral under  $\mathcal{H}_1$  with the noncentrality parameter  $\lambda = (1/\sigma^2)\theta^T M^T P_H M \theta$ . Hence, the power function of  $\delta^*$  is directly given by

$$\beta_{\delta^*}(\lambda) = \Pr_{\lambda}(\Lambda(Y) \geq \gamma_{\alpha}) = \int_{\gamma_{\alpha}}^{+\infty} \varphi_{\lambda}(x) dx, \quad (\text{B.6})$$

where  $\varphi_{\lambda}(x)$  is the pdf of the noncentral distribution  $\chi^2$  with  $n - q$  degrees of freedom and the noncentrality parameter  $\lambda$ .

## ACKNOWLEDGMENT

The material of this paper was presented in part at the CT-IP'03, Berlin, Germany, June 2003, at the GRETSP'03, Paris, France, September 2003, and at the ISPA'03, Rome, Italy, September 2003.

## REFERENCES

- [1] Y. Bresler and A. Macovski, "Three-dimensional reconstruction from projections with incomplete and noisy data by object estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 8, pp. 1139–1151, 1987.
- [2] J. Idier, Ed., *Approche bayésienne pour les problèmes inverses*, Traitement du signal et de l'image, Hermès, Lavoisier (Traité IC2), Paris, France, 2001.
- [3] R. Palenichka, A. Alekseichuk, and U. Zscherpel, "Flaw detection in radiographic images by structure-adaptive binary segmentation," in *Proc. International Symposium on Computerized Tomography for Industrial Applications and Image Processing in Radiology*, pp. 221–232, DGZfP, Berlin, Germany, March 1999.
- [4] D. J. Rossi and A. S. Willsky, "Reconstruction from projections based on detection and estimation of objects—Parts I and II: performance analysis and robustness analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 4, pp. 886–906, 1984.
- [5] A. Frakt, W. Karl, and A. Willsky, "A multiscale hypothesis testing approach to anomaly detection and localization from noisy tomographic data," *IEEE Trans. Image Processing*, vol. 7, no. 6, pp. 825–837, 1998.
- [6] M. D. Wheeler and I. Katsushi, "Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 3, pp. 252–265, 1995.
- [7] I. Kazantsev, I. Lemahieu, G. Salov, and R. Denys, "Statistical detection of defects in radiographic images in nondestructive testing," *Signal Processing*, vol. 82, no. 5, pp. 791–801, 2002.
- [8] G. Wang and T. W. Liao, "Automatic identification of different types of welding defects in radiographic images," *NDT & E International*, vol. 35, no. 8, pp. 519–528, 2002.
- [9] F. Natterer, *The Mathematics of Computerized Tomography*, John Wiley & Sons, New York, NY, USA, 1986, Editions Wiley.
- [10] M. H. Buonocore, W. R. Brody, and A. Macovski, "A natural pixel decomposition for two-dimensional image reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 28, no. 2, pp. 69–77, 1981.
- [11] R. Lewitt, "Alternatives to voxels for image representation in iterative reconstruction algorithms," *Physics in Medicine and Biology*, vol. 37, no. 3, pp. 705–716, 1992.
- [12] A. A. Borovkov, *Mathematical Statistics*, Gordon and Breach Sciences Publishers, Amsterdam, the Netherlands, 1998.
- [13] E. L. Lehman, *Testing Statistical Hypotheses*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1986.
- [14] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Transactions of the American Mathematical Society*, vol. 54, pp. 426–482, 1943.
- [15] M. Fouladirad and I. Nikiforov, "Optimal statistical fault detection with nuisance parameters," *Automatica*, vol. 41, no. 7, pp. 1157–1171, 2005.
- [16] I. Nikiforov, "Éléments de théorie de la décision statistique II: compléments, Chap. 2," in *Décision et reconnaissance des formes en signal*, R. Lengellé, Ed., pp. 43–84, Hermes Science Publications (Traité IC2), Paris, France, 2002.
- [17] A. Antoniadis, J. Berruyer, and R. Carmona, *Régression non linéaire et applications*, Économie et statistiques avancées, Economica, Paris, France, 1992.
- [18] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*, John Wiley & Sons, New York, NY, USA, 1989.
- [19] H. Lacresse, A. Grall, and I. V. Nikiforov, "Statistical fault detection in a certain class of non-linear systems with nuisance parameters," in *Proc. American Control Conference*, vol. 4, pp. 3003–3008, Denver, Colo, USA, June 2003.
- [20] K. M. Hanson and G. W. Wecksung, "Local basis-function approach to computed tomography," *Applied Optics*, vol. 24, no. 23, pp. 4028–4039, 1985.
- [21] M. Antonakios, P. Rizo, and P. Lamarque, "Real time digital tomosynthesis system dedicated to industrial solid rocket motor examination," in *Proc. 15th World Conference on Non Destructive Testing (WCNDT '00)*, Rome, Italy, October 2000.
- [22] M. Fouladirad and I. Nikiforov, "Optimal statistical fault detection with nuisance parameters," in *Proc. American Control Conference*, vol. 4, pp. 2997–3002, Denver, Colo, USA, June 2003.
- [23] K.-R. Koch, *Parameter Estimation and Hypothesis Testing in Linear Models*, Springer-Verlag, New York, NY, USA, 1999.

**Lionel Fillatre** received the M.S. degree in decision and information engineering and the Ph.D. degree in systems optimization from the University of Technology of Troyes (UTT), France, in 2001 and 2004, respectively. He is currently an Assistant Professor at the Institute of Computer Sciences and Engineering of Troyes (ISTIT), FRE CNRS 2732. His current research interests include decision theory, signal and image processing, and nondestructive testing.



**Igor Nikiforov** received the M.S. degree in automatic control from the Moscow Physical-Technical Institute in 1974, and the Ph.D. degree in automatic control from the Institute of Control Sciences, Moscow, in 1981. He joined the University of Technology of Troyes, France, in 1995, where he is a Professor and Head of the Institute of Computer Sciences and Engineering of Troyes, FRE CNRS 2732. His scientific interests include statistical decision theory, fault detection, signal processing, and navigation.

