Statistical Detection of Defects in Radiographic Images Using an Adaptive Parametric Model

Rémi Cogranne∗, and Florent Retraint∗∗
ICD - LM2S - Université de Technologie de Troyes - UMR STMR CNRS
12, rue Marie Curie - CS 42060 - 10004 Troyes cedex - France

Abstract
In this paper, a new methodology is presented for detecting anomalies from radiographic images. This methodology exploits a statistical model adapted to the content of radiographic images together with the hypothesis testing theory. The main contributions are the following. First, by using a generic model of radiographies based on the acquisition pipeline, the whole non-destructive testing process is entirely automated and does not require any prior information on the inspected object. Second, by casting the problem of defects detection within the framework of testing theory, the statistical properties of the proposed test are analytically established. This particularly permits the guaranteeing of a prescribed false-alarm probability and allows us to show that the proposed test has a bounded loss of power compared to the optimal test which knows the content of inspected object. Experimental results show the sharpness of the established results and the relevance of the proposed approach.

Keywords: Parametric radiographies model, Hypothesis testing theory, Optimal detection, Non-destructive testing, Nuisance parameter, Background subtraction.

1. Introduction
During the past decades X-ray or gamma radiography inspection have been widely used in non-destructive testing processes. It is based on the fact that high energies photons have the ability to pass through inspected object which allows the imaging of its internal structure. In most of the Quantitative Non-Destructive Testing (QNDT) applications, a voxel-by-voxel reconstruction in not necessary because the goal of the inspection is to detect an anomaly or a defect. Usually, the detection process based on radiographies relies on human experts to perform manual interpretation of acquired images. However, this process might be subjective, labour intensive, and sometimes biased. Therefore there is a wide need for computer-aided systems or for fully-automatic methods. However, depending on the inspected object geometry and its internal structure the detection may be made difficult due to the non-anomalous background.

1.1. State of the Art
The prior methods for detection of internal defects using radiographies can be divided into three categories [21, 41]: 1) generic methods that do not require any priori knowledge, 2) methods that require a ground truth or a reference, and 3) computerized methods based on statistical signal and image processing. The methodology proposed in the present paper belongs to this last group.
The first category includes defect detection methods which do not require any prior model of object structure. Such approaches typically involve image processing tools for enhancement (field flattening, contrast enhancement, edge detection, etc. . .) see [9, 28, 39] and the references therein, together with pattern recognition [6, 36] methods. More recently, state-of-the-art image processing methods, such as multi-resolution representation [54], sparse dictionary learning [37] and variational methods [32], have been applied for automatic defect detection. Similarly, classification methods have been used for automatic recognition of defect, mainly with the help of supervised machine learning [11, 58, 61]. These methods rely on the assumption that common properties can define all kinds of anomalies and distinguish them from any non-anomalous background [26, 40, 41]. The existence of such properties is doubtful in practice and these methods are thus often sensitive to the object and anomaly geometry and to the presence of noise.

The methods from the second group are based on an available ground truth: a reference radiography used as a model [41]. If a sufficient difference is measured by comparing this ground truth with radiographies, it is then assumed that the inspected object is defective [12, 53]. This approach is efficient but is very sensitive to experimental conditions, such as object position and projection angles. Moreover a ground truth is not always available in practice.

Finally, most of the computerized methods rely on the reconstruction of the inspected object from a (large) set of radiographies. However, when a limited number of radiographies is available for inspection, a full voxel-by-voxel reconstruction is impossible, see [3, Chap. 15], and [48], and the use of prior information on the non-anomalous background is necessary. Two two main approaches have been proposed to introduce statistical prior knowledge: Bayesian and non-Bayesian approaches. For a more detailed review on methods for automatic defect detection, the reader is referred to [13, 35, 41]

1.2. Motivation of This Work

The methodology proposed in the present paper belongs to the category of non-Bayesian statistical methods for anomaly detection. Let us briefly recall the relative advantages and drawbacks of Bayesian and non-Bayesian approaches in order to emphasize the contribution of the proposed methodology. The most commonly found methods in the literature belongs to the Bayesian approach [24]. In fact, Bayesian statistical approach offers an efficient (with minimal expected risk) and rather simple solution for the detection problem. However it assumes that 1) occurrence of an anomaly is a random event with known prior probability and, 2) the non-anomalous background and the anomaly are random and drawn from apriori known distributions. These apriori knowledges on the inspected object and potential anomalies are not always available; this may compromise the application of Bayesian approaches.

In such a situation, a more convenient modelling of radiographies can be obtained by representing the expected non-anomalous background as a linear combination of basis functions. This modelling particularly makes possible the use of non-Bayesian hypothesis testing theory. In fact, such hypothesis testing methods allow the introducing of “nuisance parameters” which have no interest for the considered QNDT but which must be taken into account because their impact may be non negligible. In the considered anomaly detection problem, this is especially relevant because the non-anomalous background has no interest while it may hide the anomaly and, hence, may prevent their detection.

To remove, or reject, the nuisance parameters, the main idea is to split the space of observations into two subspaces: one containing all the nuisance parameters and the other one completely free from these nuisance parameters. Of course, the main difficulty here is to have a model which is accurate enough to represent the content of any radiography, while leaving the anomalies into the background-free subspace, to preserve a high detection performance. Examples of such a non-anomalous background representation and rejection can be found in [30, 58] but this approach has seldom been used together hypothesis testing theory [21]. Some other approaches have been proposed to allow the subtraction of background such as the simple field-flattening operation [9] and the image processing methods for denoising [35, 46]. The main drawback of these methods is that they do no allow the establishing of neither the decision threshold which guarantee a given false-alarm probability nor the missed-detection probability (or detection power).

Similarly, several Bayesian approaches have been proposed. Those methods require that the apriori distribution of Background is known, so the problem of its rejection is trivial. Hence, the problem reduces to test the presence of signal in noise in the framework of binary [49] or multiple hypothesis testing theory [24].

On the contrary, the use of a linear model of non-anomalous background have been used to provide a test with known statistical properties in [21]. Unfortunately, this test is based on a geometrical model, or on a Computer-Aided Design (CAD) model, of the inspected object which may not always be available. Moreover, a precise calibration process is required to match the observed radiography with the geometrical model. Hence, this method is compromised when the inspected object geometry or acquisition conditions may slightly change.

In our previous work, this methodology, using hypothesis testing theory with the use of a linear model of background, allows us to address the problem of hidden data detection [14, 15]. It has also been shown, for that detection problem, that the use of a precise model [16, 17] allows the obtaining of a statistical test with better performance. The main difference is that, is the context of anomaly detection, no information is available of the potential anomaly (shape, size, position, etc.) while, for hidden data detection, the alternative hypothesis is well described statistically.

1.3. Contribution of This Paper

This paper proposes a novel methodology for automatic detection of anomalies from a few radiographies. An original model of radiographies is proposed based on a study of the whole radiography acquisition process [47, 57]. This model
is used for background subtraction/rejection. Then, a statistical test with analytically established properties is proposed. The main contributions are the following:

1. By modelling the whole acquisition process, the proposed methodology does not require any prior knowledge on the inspected object and, hence, is entirely automated; this also allows the application of the proposed methodology to a wide range of inspected objects.

2. Thanks to an original linearisation of the proposed radiography model, the whole detection process is computationally efficient.

3. The statistical properties of the proposed test are established. This permits us to calculate the decision threshold that guarantees a prescribed false-alarm probability.

4. It is proved that the proposed test is almost optimal, i.e. the loss of performance, compared to the optimal statistical test, is bounded.

Numerical results emphasize the accuracy of the proposed model and highlight the sharpness of theoretical findings.

1.4. Organization of This Paper

The paper is organized as follows. Section 2 states the problem of anomaly detection within the framework of hypothesis testing theory. When the non-anomalous radiography background is known, Section 3 presents the optimal test and establishes its statistical performance. Section 4 presents the proposed piecewise non-linear model of radiographies. Section 5 presents the proposed statistical test, which exploits this model; the statistical properties of the proposed test are also analytically established. Section 6 presents numerical results and Section 7 concludes the paper.

2. Statistical Detection of Anomalies: Problem Statement

Let $Z = \{z_{m,n}\}, (m, n) \in Z$ be a a noisy radiography represented as an image of $M \times N$ pixels, hence $Z = \{1, \ldots, M\} \times \{1, \ldots, N\}$. For the sake of clarity the quantization of pixels’ value is not considered, hence, each $z_{m,n}$ belongs to $\mathbb{R}_+$. When the inspected object is free from anomaly, the value of each pixel can decomposed as:

$$z_{m,n} = \mu_{m,n} + \xi_{m,n},$$

where $\mu_{m,n}$ is a deterministic value related to the amount of gamma or X-ray photons passing through the object and reaching sensor at location $(m, n)$; in practice, $\mu_{m,n}$ corresponds to the expectation of pixel $z_{m,n}$. On the contrary, the $\xi_{m,n}$’s are the realisation of statistically independent Gaussian random variables $\xi_{m,n} \sim \mathcal{N}(0, \sigma^2_{m,n})$, whose variances $\sigma^2_{m,n}$ > 0 change from pixel to pixel, mainly due to the photo-counting Poissonian process [22, 29].

If follows from Equation (1) that for a non-anomalous object, each $z_{m,n}$ follows a Gaussian distribution:

$$z_{m,n} \sim \mathcal{N}(\mu_{m,n}, \sigma^2_{m,n}).$$

When an anomaly is present in the inspected object, the value of each pixel can decomposed as:

$$z_{m,n} = \mu_{m,n} + a_{m,n} + \xi_{m,n},$$

where $a_{m,n}$ represents the variation of expected pixels’ value due to the presence of the anomaly. Usually the anomaly is of limited size, hence, $a_{m,n}$ equals zero except for the relatively few locations $(m, n)$ where the anomaly occurs. It follows from Equation (3) that when an anomaly is present, each $z_{m,n}$ follows a Gaussian distribution:

$$z_{m,n} \sim \mathcal{N}(\mu_{m,n} + a_{m,n}, \sigma^2_{m,n}).$$

When inspecting an object from the (noisy) radiography $Z$, the goal of anomaly detection is to decide between the two following hypotheses:

$$\{\mathcal{H}_0 = \{z_{m,n} \sim \mathcal{N}(\mu_{m,n}, \sigma^2_{m,n}), \forall (m, n) \in Z\}$$

$$\mathcal{H}_1 = \{z_{m,n} \sim \mathcal{N}(\mu_{m,n} + a_{m,n}, \sigma^2_{m,n}), \forall (m, n) \in Z\}.$$ (5)

This formalisation of the problem of defect detection is referred to as an “ideal observer model” in the field of medical imaging, see [3, Chap. 13] and [5, Chap. 9].

Formally, a statistical test is a mapping $\delta : \mathbb{R}^{M \times N} \rightarrow \{\mathcal{H}_0; \mathcal{H}_1\}$ such that hypothesis $\mathcal{H}_1$ is accepted if $\delta(Z) = \mathcal{H}_1$ (see [34] for details about statistical hypothesis testing).

As discussed in the introductory Section 1, the main goal of this paper is to design a statistical test with known performance, in order to guarantee a false-alarm probability. Hence, let:

$$\mathcal{K}_{\alpha_0} = \{\delta : P_{\mathcal{H}_0}(\delta(Z) = \mathcal{H}_1) \leq \alpha_0\},$$ (6)

be the class of all the tests whose false alarm probability are upper-bounded by $\alpha_0$. Here $P_{\mathcal{H}_i}(A)$ stands for the probability of event $A$ under hypothesis $\mathcal{H}_i, i \in \{0, 1\}$. Among all the tests in $\mathcal{K}_{\alpha_0}$, it is aimed at finding a test $\delta$ which maximizes the power function, defined by the probability of correct detection:

$$\beta_0 = P_{\mathcal{H}_1}(\delta(Z) = \mathcal{H}_1).$$ (7)

Alternatively, $\beta_0$ can also be defined as $1 - \alpha_1$ where $\alpha_1 = P_{\mathcal{H}_0}(\delta(Z) = \mathcal{H}_0)$ represents the probability of missed-detection. The hypothesis testing problem as described in (5) - (7) emphasizes the major difficulties of automatic anomaly detection. The first difficulty is that, in practice, the expectation of pixels is unknown. Hence it is necessary to design an accurate model of radiography content in order to estimate the expectation of all the pixels. However, characterizing accurately the content of any radiography is very difficult due to high diversity of inspected objects geometry or structure. Moreover, the estimation of all nuisance parameters $\mu_{m,n}$ and $\sigma^2_{m,n}$ from projection $Z$ remain an open problem in the field of image processing. The second difficulty is that, usually, no prior knowledge on the potential anomalies is available. In fact, even though a specific type of anomaly might be suspected, a bone break or a tumor, for instance, in medical imaging [2, 10] or a particular welding defects in NDT applications [58], the location, size and shape of
the anomaly are unknown, which remains a difficult problem; As a consequence, it is impossible to model \( \sigma_{m,n} \) for all pixels. Moreover, the nuisance parameters may “hide” the potential anomalies and, hence, may prevent their detection. It is thus necessary to design a test which explicitly takes into account an accurate model of radiography content in order to subtract the non-anomalous background / nuisance parameter. Note that it is also important to warrant that the anomaly is not subtracted together with the non-anomalous background; the model must distinguish the non-anomalous background from the anomaly.

3. Optimal Detection with a Linear Model of Non-Anomalous Nuisance background

When \( \mu_{m,n} \) are unknown, the detection problem (5) is not solvable. To overcome this difficulty, it is proposed in this paper to form \( I \) vectors of pixels by grouping small sets of \( J \) adjacent and non-overlapping pixels. By denoting the vector \( z_i = (z_{1,i}, z_{2,i}, \ldots, z_{J,i}) \), it follows from Equation (2) that for a non-anomalous object, the distribution of \( z_i \) is given by:

\[
\mathbf{z}_i = \mathbf{\mu}_i + \mathbf{\xi}_i \Leftrightarrow \mathbf{z}_i \sim \mathcal{N}(\mathbf{\mu}_i, \sigma^2 I),
\]

where \( \mathbf{\mu}_i = (\mu_{1,i}, \mu_{2,i}, \ldots, \mu_{J,i})^T \) is the expectation of vector \( \mathbf{z}_i \), \( \sigma^2 \) is its variance, assumed constant over the \( J \) pixels, and \( I \) is the identity matrix of size \( J \times J \). Here, it should be noted the assumption of independent pixels with the same variance is adopted for clarity but not required in practice. In fact, for any covariance matrix \( \Sigma \), the change of variable \( g(\mathbf{z}_i) = \Sigma^{-1/2} \mathbf{z}_i \), where \( \Sigma^{-1/2} \) is the symmetric matrix satisfying \( \Sigma^{-1} = \Sigma^{-1/2} \Sigma^{-1/2} \), permits the obtaining of model (8).

Similarly, with the previously introduced notation of vectors, it follows from Equation (4) that, when an anomaly is present, each vector \( \mathbf{z}_i \) follows a Gaussian distribution given by:

\[
\mathbf{z}_i \sim \mathcal{N}(\mathbf{\mu}_i + \mathbf{a}_i, \sigma^2 I),
\]

where \( \mathbf{a}_i = (a_{1,i}, a_{2,i}, \ldots, a_{J,i})^T \) is the vector representing the variation of \( i \)-th vector expectation due to the presence of anomaly. As briefly described in the introduction, it is proposed in the present paper to model the expectation of each vector as a linear combination of basis functions. This yields the following linear model:

\[
\mathbf{\mu}_i = \mathbf{Hs}_i.
\]

Here, \( \mathbf{H} \) is known full-column rank matrix of size \( J \times p \) whose columns are the basis functions used to represent \( \mathbf{\mu}_i \), and \( s_i \in \mathbb{R}^p \) is a set of \( p < J \) nuisance parameters which entirely defines \( \mathbf{\mu}_i \). It follows from Equations (8), (9) and (10) that using the vectors of pixels, anomaly detection problem can be written as a choice between the hypotheses defined by:

\[
\begin{align*}
\mathcal{H}_0 &= \{ \mathbf{z}_i \sim \mathcal{N}(\mathbf{Hs}_i, \sigma^2 I), \quad \forall i = \{1, \ldots, I\} \} \\
\mathcal{H}_1 &= \{ \mathbf{z}_i \sim \mathcal{N}(\mathbf{Hs}_i + \mathbf{a}_i, \sigma^2 I), \quad \forall i = \{1, \ldots, I\} \}.
\end{align*}
\]

Ideally one wishes to design a test which is Uniformly Most Powerful (UMP) in the class \( \mathcal{K}_m \) i.e. a test which maximizes the detection power whatever the anomaly \( \mathbf{a}_i, i = \{1, \ldots, I\} \) might be. Unfortunately, such a test does not exists in general, see [7, 23, 34, 51] for further details. To solve the hypothesis testing problem of unknown anomaly detection in the presence of nuisance parameters, the theory developed by Wald [60] and the theory of invariant tests are useful. One of the key idea of Wald is to replace the UMP criterion of optimality by the Uniformly Best Constant Power (UBPC) criterion. The reason behind this choice is that the power of a test depends both on the nuisance parameters and on the anomaly. In fact, the anomaly might be hidden by the nuisance parameters; for instance if there exists \( \mathbf{c} \) such that \( \mathbf{a}_i = \mathbf{Hc} \) it is obvious that the anomaly will be undetectable using the model \( \mathbf{H} \).

3.1. Criterion of Optimality

The idea of invariant theory for hypothesis testing is to exploits the geometrical properties of the observations. It requires that the detection problem posses a natural geometrical invariance with respect to some group of transformations. First, let us note that the above mentioned hypothesis testing problem, given by (11), remains invariant under the group of translations \( \mathcal{G} = \{ g : g(\mathbf{z}_i) = \mathbf{z}_i + \mathbf{Hx}, \quad \mathbf{x} \in \mathbb{R}^p \} \). To apply the theory of invariance, see [34, chap.6], let us define \( \mathcal{R}(\mathbf{H}) \subseteq \mathbb{R}^p \) the column space spanned by \( \mathbf{H} \). The usual application of invariance theory consists in projecting each observation vector \( \mathbf{z}_i \) onto the orthogonal complement \( \mathcal{R}(\mathbf{H})^\perp \) of the column space \( \mathcal{R}(\mathbf{H}) \).

The \( J - p \)-dimensional subspace \( \mathcal{R}(\mathbf{H})^\perp \) is well-known under the name “parity space” in the literature of automatic control, see [23, 25, 51].

The main idea is that behind the use of parity space is to decompose each observation \( \mathbf{z}_i \) into an information-part, i.e. \( \mathbf{\mu}_i = \mathbf{Hs}_i \), and a noisy part \( \mathbf{z}_i - \mathbf{\mu}_i = \mathbf{\xi}_i \). It is expected that an anomaly does not lies in the same subspace that the
information-part, in other words it can be distinguished from the background. Hence, when an anomaly is present, (part of) the anomaly should be present in the noisy-part. The problem thus reduces to detect a signal in noise. This is very similar to the well-known matched subspace detector \([51]\), except that this detector assumes that a specific subspace in which lies the anomaly is known. In this paper, without any information on the potential anomaly the whole parity space is exploited.

To obtain this decomposition, each observation \(z_i\) is projected onto the \(J - p\) dimensional subspace \(R(H^+)\) as follows: \(\tilde{z}_i = Wz_i\). Here, the matrix \(W^T = (w_1, \ldots, w_{J-p})\) is composed of the eigenvectors \(w_1, \ldots, w_{J-p}\) of the projection matrix \(P_{H} = I_J - H(H^TH)^{-1}H^T\) corresponding to eigenvalue 1. The matrix \(W\) verifies, among others, the following useful properties:

\[
WH = 0, \quad W^T W = P_{H}, \quad WW^T = I_{J-p}.
\] (12)

Hence, by projecting the radiographic observation \(z_i\) onto the parity subspace \(R(H^+)\) yields:

\[
\tilde{z}_i = Wz_i = W\xi_i [ + Wa_i],
\] (13)

which shows that the projection onto the parity space \(R(H^+)\) provides an easy method for algebraic subtraction of background or nuisance parameters. The results \(\tilde{z}_i\) of observations \(z_i\), projected onto the parity subspace \(R(H^+)\), is sometimes referred to as “residuals” because it free from non-anomalous background and hence should only contains the residual noises. It is shown in \([23]\) that the use of the “residuals” \(\tilde{z}_i\) allows the design of a test which has the UBPC property over the family of surface:

\[
C_{H} = \left\{ C_{\xi} : \sum_{i=1}^{I} \frac{1}{\sigma^2_i} ||W_{a_i}||_2^2 = \varrho^2 \right\}.
\] (14)

In the present paper \(C_{\xi}\), as defined in Equation (14), is referred to as “Anomaly-to-Noise Ratio” (ANR) and plays an important role to define the detectability of an anomaly.

These concepts of algebraic background subtraction and UBPC test are illustrated in the Figure 1 and are exemplified in the Example 1 below. For a more detailed review on nuisance parameters rejection and on testing composite hypothesis, the reader is referred to \([4]\).

**Example 1.** For simplicity and clarity, this example is presented without noise. Let us consider the following matrix:

\[
\begin{pmatrix}
4 & 5 & 6 \\
9 & 9 & 9 \\
14 & 13 & 12 \\
19 & 17 & 15 \\
24 & 21 & 18 \\
\end{pmatrix}
= \begin{pmatrix}
1 & 0 \\
1 & 1 \\
1 & 2 \\
1 & 3 \\
1 & 4 \\
\end{pmatrix} \times \begin{pmatrix}
4 & 5 & 6 \\
5 & 4 & 3 \\
\end{pmatrix}.
\]

Again, the columns of the matrix \(H\) span the \(\mathbb{R}^3\) space in which the observations are expected to lie. The set of parameters \([s_1, s_2, s_3]\) entirely defines the expectation of the observations, that is \([\mu_1, \mu_2, \mu_3]\).

In this example, it is assumed that the two last observations are corrupted by the following anomalies \(a_2 = (1\ 1\ 3\ 1 - 1)^T\) and \(a_3 = (0\ 2\ 2\ 0)^T\). It follows that the observations are:

\[
\tilde{z}_1 = \begin{pmatrix}
4 \\
9 \\
14 \\
19 \\
24 \\
\end{pmatrix}, \quad \tilde{z}_2 = \begin{pmatrix}
6 \\
10 \\
16 \\
18 \\
20 \\
\end{pmatrix}, \quad \text{and} \quad \tilde{z}_3 = \begin{pmatrix}
6 \\
11 \\
17 \\
18 \\
\end{pmatrix}.
\]

In order to detect the presence of anomalies, we have to consider only the part of observations that lie in the parity by rejecting the nuisance parameter (that is the part of observations that lie in \(\mathbb{R}^2\) space spanned by the columns of \(H\)). A direct calculation shows that the projector onto the parity space \(P_{H}\) and the projection matrix \(W\) are given by:

\[
P_{H} = \begin{pmatrix}
4 & -4 & -2 & 0 & 2 \\
-4 & 7 & -2 & -1 & 0 \\
-2 & -2 & 8 & -2 & -2 \\
0 & -1 & -2 & 7 & -4 \\
2 & 0 & -2 & -4 & 4 \\
\end{pmatrix},
\]

\[
W = \begin{pmatrix}
\frac{2}{\sqrt{10}} & \frac{-1}{\sqrt{10}} & \frac{-2}{\sqrt{10}} & \frac{-2}{\sqrt{10}} & \frac{-2}{\sqrt{10}} \\
\frac{1}{\sqrt{10}} & \frac{2}{\sqrt{10}} & \frac{0}{\sqrt{10}} & \frac{-1}{\sqrt{10}} & \frac{2}{\sqrt{10}} \\
\frac{2}{\sqrt{10}} & \frac{2}{\sqrt{10}} & \frac{0}{\sqrt{10}} & \frac{0}{\sqrt{10}} & \frac{2}{\sqrt{10}} \\
\end{pmatrix}.
\]

Now a direct calculation shows that:

\[
||W_{z_{1}}||_2^2 = 0, \quad ||W_{z_{2}}||_2^2 = 6.4 \text{ and } ||W_{z_{3}}||_2^2 = 11.2.
\]

This shows that the anomaly are easily detected in the parity space. In addition, this simple example also shows that the detectability of an anomaly depends on its norm into the parity space. The anomalies \(a_1\) and \(a_2\) have the following norms \(||a_1||_2^2 = 13\) and \(||a_2||_2^2 = 12\). However, while the anomaly \(a_2\) is almost orthogonal to each column of \(H\), it is not the case for \(a_1\). This is why the anomaly \(a_2\) is easier to detect into the parity space. This is also explains the necessity to define the UBPC criterion which relies on the surface \(C_{H}\) see Equation (14), that, roughly speaking, defines anomalies whose projections onto the parity space have the same norm.

### 3.2. Design and Performance of the Optimal UBPC Invariant Test

Finally, from the theory of invariance, the anomaly detection problem (11) is strictly equivalent to the problem of choosing between the following hypotheses:

\[
\begin{align*}
\mathcal{H}_0 & = \{ \tilde{z}_i \sim N(0, \sigma_i^2 I_{J-p}) \}, \quad \forall i = \{1, \ldots, I\} \\
\mathcal{H}_1 & = \{ \tilde{z}_i \sim N(W_{a_i}, \sigma_i^2 I_{J-p}) \}, \quad \forall a_i \in \mathbb{R}^J, \quad \forall i = \{1, \ldots, I\}.
\end{align*}
\] (15)
To solve testing problems such as (15) it is shown in [23] that the test defined by:

\[
\delta^*(Z) = \begin{cases} 
\mathcal{H}_0 & \text{if } \Lambda^*(Z) = \sum_{i=1}^I \frac{\|z_i\|^2}{\sigma_i^2} = \sum_{i=1}^I \frac{\|Wz_i\|^2}{\sigma_i^2} < \tau_{\alpha_0}^* \\
\mathcal{H}_1 & \text{if } \Lambda^*(Z) = \sum_{i=1}^I \frac{\|z_i\|^2}{\sigma_i^2} = \sum_{i=1}^I \frac{\|Wz_i\|^2}{\sigma_i^2} \geq \tau_{\alpha_0}^* 
\end{cases}
\]

is UBPC over the class \(\mathcal{K}_{\alpha_0}\) provided that the threshold \(\tau_{\alpha_0}^*\) is the solution of equation \(\mathbb{P}_{\mathcal{K}_{\alpha_0}}[\Lambda^*(Z) \geq \tau_{\alpha_0}^*] = \alpha_0\). It worth noting that the norm of the “residuals” \(\|Wz_i\|^2\) can also be obtained using the matrix \(P_H^2\) because, from the properties (12), a straightforward calculation shows that

\[
\|P_H^2z\|^2 = z_i^2W^3WW^3Wz = \|Wz\|^2.
\]

For clarity, in the present paper the matrix \(W\) is used in all calculus while matrix \(P_H\) is used for illustrations and figures, to keep the same number of observations. Now that an optimal test has been defined, it is crucial to establish its statistical performance. From the properties of Gaussian random variables, it is straightforward to verify that:

\[
\sum_{i=1}^I \frac{1}{\sigma_i^2} \|z_i\|^2 \sim \begin{cases} \chi^2(0) & \text{under } \mathcal{H}_0 \\
\chi^2(\rho^2) & \text{under } \mathcal{H}_1. 
\end{cases}
\]

Where \(\chi^2(\rho^2)\) represents the non-central chi-squared distribution with \(N = (J - p)\) degree of freedom and non-centrality parameter \(\rho^2\), defined in Equation (14).

The Equation (17) can be interpreted as follows: when no defect is present, the non-anomalous background is correctly rejected, or subtracted, hence the residuals follow a Gaussian distribution with zero mean and their norms follow a \(\chi^2\) distribution. On the contrary, when a defect is present, the residuals also contain the defect projected onto the parity space; they thus follow a Gaussian distribution but with a non-zero mean due to the defect, hence, their norm follow a non-central \(\chi^2\) distribution.

From the definition of the UBPC test \(\delta^*\) (16) and from Equation (17), a short algebra permits the establishing of UBPC test properties which are given in the following Theorem 1.

**Theorem 1.** For testing the hypotheses defined in Equation (15), it follows from distribution (17) that for any \(\alpha_0 \in (0; 1)\) the decision threshold:

\[
\tau_{\alpha_0}^* = F_{\chi^2_{\rho^2}}^{-1}(1 - \alpha_0; 0),
\]

guarantees that \(\mathbb{P}_{\mathcal{K}_{\alpha_0}}[\Lambda^*(Z) \geq \tau_{\alpha_0}^*] = \alpha_0\) so that the UBPC test \(\delta^*\) (16) is in the class \(\mathcal{K}_{\alpha_0}\). Here, \(F_{\chi^2_{\rho^2}}(x; \rho^2)\), and \(F_{\chi^2_{\rho^2}}^{-1}(p; \rho^2)\), represents the cumulative distribution function of the non-central chi-squared, and its inverse, at value \(x\), or at probability \(p\), with \(\chi\) degree of freedom and a non-centrality parameter \(\rho^2\).

Choosing the threshold \(\tau_{\alpha_0}^*\) as defined in (18), it follows from distribution (17) that the power function associated with the UBPC test \(\delta^*\) (16) is given by:

\[
\beta^*(\rho, \alpha_0) = 1 - F_{\chi^2_{\rho^2}}^{-1}(1 - \alpha_0; 0; \rho^2) = 1 - F_{\chi^2_{\rho^2}}^{-1}(1 - \alpha_0; 0; \rho^2) = 1 - F_{\chi^2_{\rho^2}}^{-1}(1 - \alpha_0; 0; \rho^2).
\]

**Proof.** of the Theorem 1 is given in the Appendix A.

Theorem 1 emphasizes that, provided that the model \(H(10)\) is chosen correctly, then the decision threshold only depends on \(\alpha_0\) and \(\mathcal{T}\). Hence, it is straightforward to guarantee a prescribed false-alarm probability whatever the inspected object might be. Moreover, Theorem 1 also provides an upper bound on the detection power one can expect from any test that aims at detecting an anomaly; this power function is denoted \(\beta^*(\rho, \alpha_0)\) (19) to highlight that it mainly depends on the ANR \(\alpha\) (14) and on the prescribed false-alarm probability \(\alpha_0\).

### 4. Statistical Model of Radiographic Images

#### 4.1. Modelisation of Observed Material

An object is entirely characterized by its absorption function \(a(x, y, z, \nu)\), where \(\nu\) is the photons energies and \((x, y, z)\) are the continuous coordinates. Without loss of generality, the coordinates \((x, y)\) verify \((x, y) \in \mathcal{D} \subset \mathbb{R}^2\) where \(\mathcal{D}\) is the photo-sensor domain; the sensor lies in the plane \(z = 0\) and the Gamma or X-ray source is in the plane \(z = z_m\). It follows from Beer-Lambert law [9], that the “ideal” mean number of photons passing through the inspected object and reaching the sensor at location \((x, y)\) is given by [47]:

\[
\omega(x, y) = N_0 A(x, y),
\]

where \(N_0\) represents the mean number of Gamma or X-ray photons emitted by the source together with quantum efficiency and amplification factor. Here, \(A(x, y)\) represents the global absorption function over the whole emission spectrum and between the Gamma or X-rays source and the sensor:

\[
A(x, y) = \int_{\mathbb{R}} \phi(\nu) \exp \left( - \int_{0}^{z_m} a(l(z), z, \nu) dz \right) d\nu.
\]

where \(\phi(\nu)\) represents the energy distribution of emitted photons, or emission spectrum, satisfying \(\int_{\mathbb{R}} \phi(\nu) d\nu = 1\) and \(l : \mathbb{R} \to \mathbb{R}^2\) represents the path of photons, from the source to the sensor, considered as a straight line in this paper.

However, the mean measured number of incident photons \(\mu(x, y)\) widely differs from \(\omega(x, y)\) because deterministic degradations affect the acquisition process. On the one hand, the scattered beam, due to photons interactions within the inspected object, is recorded along with the radiography [3, Chap. 16] and [47]. From a practical point of view, the scattering degradation is modelled by a convolution product. Actually, in most of the NDT applications, the distance between the photo-sensor and the inspected object is sufficiently important to approximate accurately the scattered radiation by a polynomial [59]:

\[
\mu_{\text{scat}}(x, y) \approx \sum_{i=0}^{p_i} \sum_{j=0}^{p_j} b_{ij} x^i y^j,
\]

where the \(b_{ij}\)'s are the bivariate polynomial coefficients with degrees \(p_i\) and \(p_j\).

On the other hand, the non-scattered incident radiations are also affected by others physical phenomena, such as diffraction,
conversion screen, cone beam angle, etc... see [3, Chap. 16] and therein references. Taking into account these deterministic phenomena, the mean number of incident photons, denoted $\mu(x,y)$, is given by [47]:

$$\mu(x,y) = \mu_{\text{inc}}(x,y) + \int_{\mathbb{R}^2} a(x,y) h_{(x,y)}(u-x, v-y) \, dx \, dy,$$

(22)

where (22) represents a two-dimensional convolution product with kernel $h_{(x,y)}$ at location $(x,y)$; $h_{(x,y)}$ is sometimes referred to as the Point Spread Function (PSF) [27]. It should be noted that without loss of generality, it is assumed that the convolution kernel depends on the location $(x,y)$, to emphasize this dependence, the convolution kernel is denoted $h_{(x,y)}$. It also worth noting that scattering and others degradations are usually modelled separately because the scattered radiations are much more low pass filtered, see Section 4.2.

The model of absorption function of the inspected object $a(x,y,z,v)$ used in this paper is inspired from [8, 42, 44]. Roughly speaking, this model is based on the idea that the inspected object is composed of different materials separated by abrupt edges. Each material exhibits a specific absorption function $a$, therefore, the global absorption $A$ is model as a juxtaposition of continuous functions defined over $N_A$ areas which are separated by abrupt edges. From the physical properties of photons emission and absorption, one can expect that the following properties hold.

1. Each area is associated with a domain $D_i$ over which the absorption varies smoothly.
2. The absorption function $A$ is discontinuous across (most of) the domains boundaries.
3. The material composing the inspected object are of regular shape, i.e. domains boundaries are regular curves.

More formally, it follows from the previous properties that a general model of the absorption function $A(x,y)$ is given by:

$$A(x,y) = \sum_{i=1}^{N_A} A_i(x,y) \mathbf{1}_{D_i}(x,y),$$

(23)

where $A_i(x,y)$ are real functions continuous over $D_i$, $\mathbf{1}_{D_i}$ is the indicator function of the set $D_i$ and $\{D_i\}$ forms a partition of $D$. From the regularity of boundaries separating the different materials, the discontinuities of $A(x,y)$ can be considered as a set of parametric curves. The model described by (23) is sufficiently general to describe the absorption function of a large class of objects.

### 4.2. A Parametric Model of Radiographies

The PSF of an imaging system depends on many elements [27], from both optical and sensor system, and the literature proposes a wide range of mathematical models. The methodology presented in this paper can formally be applied for any isotropic PSF. This is especially relevant as in practice all the blurring processes of a radiography acquisition system can be approximated as isotropic [9, 47]. However, for the sake of definition, the calculations presented in this paper are restricted to a Gaussian PSF, though an extension to another model is possible at the cost of computation. Hence, let the imaging system be locally modelled using the following PSF:

$$h(x,y; \varsigma_{\text{psf}}) = \frac{1}{\varsigma_{\text{psf}}^2} \varphi \left( \frac{\sqrt{x^2 + y^2}}{\varsigma_{\text{psf}}} \right),$$

(24)

where $\varphi(u) = (2\pi)^{-1} \exp(-u^2/2)$ and $\varsigma_{\text{psf}} = \varsigma_{\text{psf}}(x,y) > 0$ is the local PSF parameter which represents the non-stationary nature of the scattering process over the whole image plane (due to aberrations, materials properties, etc...).

As explained in Section 4.1, the sensor domain $D$ is split into $N_A$ areas, denoted $D_i$, which generally exhibit a complex geometry. The design of a two-dimensional (2D) parametric models of such regions is a difficult problem which is crucial in this paper to detect anomalies / defects with a high confidence level. Moreover, a large computational time is usually required to correctly manipulate 2D models; this is also a major drawback to inspect a large set of radiographies or a large number of objects. Hence, to overcome these difficulties, it is proposed to exploit the image model in one-dimension (1D). To this end, the scene domain $D$ is decomposed into $I$ non-overlapping domains $L_i, i = 1, \ldots, I$, referred to as segments due to their small width. For clarity reasons, in this paper each segment is extracted from an horizontal (or vertical) line of the sensor plane which corresponds in practice to $J$ consecutive pixels.

Therefore, let $A_i(x) = A(x,y), (x,y) \in L_i$ be the absorption function over the $i$-th segment. The small width of segment $L_i$ allows the approximation of coordinates in one dimensional. From the scene model (22) and (23), $A_i$ is a piecewise continuous function (continuous except across discontinuities between materials) and thus admits the following decomposition [8]:

$$\forall x \in L_i, \ A_i(x) = A_i^{(c)}(x) + A_i^{(s)}(x),$$

where $A_i^{(c)}$ is uniformly continuous and $A_i^{(s)}$ is the singular part. The function $A_i$ is piecewise constant and hence can be written:

$$\forall x \in L_i, \ A_i^{(s)}(x) = \sum_{d=1}^{r_i} u_{i,d} \mathbf{1}_{L_i}(x - t_{i,d}),$$

(25)

with $r_i$ the number of discontinuities in $L_i$ and $\mathbf{1}_{L_i}$ the indicator function of the set $\mathbb{R}^+$. The parameters $u_{i,d}$ and $t_{i,d}$ characterize, respectively, the intensity and the location of the $d$-th discontinuity in $L_i$. From the linearity of Equation (22), the degraded
(blurred or scattered) mean number of incident photons reaching the sensor over \(i\)-th segment \(L_i\) is given by:
\[
\mu_i(x) = \mu_i^{(c)}(x) + \mu_i^{(s)}(x),
\]
where \(\mu_i^{(c)}\) and \(\mu_i^{(s)}\) respectively correspond to the convolved continuous and singular part of “ideal” unaltered flow \(\omega_i\). The following Theorem 2 gives a parametric model of the singular part \(\mu_i^{(s)}\). In the following, \(\Phi(u)\) denotes the Gaussian integral function
\[
\Phi(u) = \int_{-\infty}^{x} \exp\left(-\frac{x^2}{2}\right) dx.
\]
with \(\phi(u) = (2\pi)^{-1/2}\exp(-u^2/2)\).

**Theorem 2.** Assuming that the absorption function is described by (23) and the PSF by (24), then, the function \(\mu_i^{(s)}(x)\) is given as:
\[
\mu_i^{(s)}(x) = \sum_{d=1}^{r_i} u_{i,d} \Phi\left(\frac{x - t_{i,d}}{\varsigma_{i,d}}\right),
\]
where \(\varsigma_{i,d} = \varsigma_{pdf}/(\cos(\psi))\), and \(\psi \in [-\frac{\pi}{2}, \frac{\pi}{2}]\) is the angle between the local normal to the discontinuity and the analysed segment.

**Proof.** of the Theorem 2 is given in the Appendix B. \(\square\)

On the opposite, the function \(\mu_i^{(c)}\) results from the convolution between \(A_i^{(c)}\) and \(h\), hence, it is assumed to be very smooth or low frequency. Consequently, it can be modelled accurately by an algebraic polynomial of degree \(p - 1\). Such piecewise polynomial model of smooth areas has been used in [1, 31, 52] to design image coding or compression methods and in [16, 17, 18] for detection of data hidden in images. Using the piecewise polynomial model of the continuous part \(\mu_i^{(c)}\) and the Equation (27) to model blurred discontinuities \(\mu_i^{(s)}\), it follows from Equation (26) that:
\[
\mu_i(x) = \sum_{q=0}^{p-1} s_{i,q} x^q + \sum_{d=0}^{r_i} u_{i,d} \Phi(x; \eta_{i,d}),
\]
where \(s_{i,q}\) and \(\eta_{i,d}\) are the polynomial coefficients representing the continuous part, \(\eta_{i,d} = (t_{i,d}, \varsigma_{i,d})\) represents discontinuity parameters and \(\Phi(x; \eta_{i,d}) = \Phi(x - t_{i,d}/\varsigma_{i,d})\). It worth noting that, in the present paper, the beam scattering, as modelled in Equation (21), it taken into account within the continuous part \(\mu_i(x)\). In fact, as discussed in Section 4.1, it can be modelled as a bi-variate polynomial and, hence, along a line it only impacts the one-dimensional polynomial coefficients \(s_{i,q}\).

The number of incident photons is finally integrated over each photo-cell to produce a digital image of the acquired radiography. Let us denote \(\mu_i = (\mu_{i,1}, \ldots, \mu_{i,J})\) the vector with \(J\) measurements or pixels after sampling:
\[
\mu_{i,j} = \int_{x_j}^{x_{j+\Delta}} \mu_i(x) dx,
\]
where \(\Delta\) is the size of sensor photo-cells and \(x_j\) represents the relatives coordinates of sample points. From the Equation (28), the measured intensity of pixels on the \(i\)-th segment is given by:
\[
\mu_i = \mathbf{H}_i \mathbf{F}(\eta_i) \mathbf{u}_i = \mathbf{G}(\eta_i) \mathbf{v}_i.
\]

Here, \(\mathbf{H}\) is the matrix of size \(L \times p\) whose element \((i,q)\) is \(x_j^q\) and \(\mathbf{F}(\eta_i)\) is the matrix of size \(L \times r_i\) whose element \((i,d)\) is \(\Phi(x; \eta_{i,d})\). It is obvious that \(\mathbf{G}(\eta_i) = (\mathbf{H}(\mathbf{F}(\eta_i)))\) and \(\mathbf{v}_i = (\mathbf{s}_i, \mathbf{u}_i)^T\). It should be noted that matrix \(\mathbf{H}\) remains the same for all \(i\) because all the vectors, \(\mu_i\), have the same number of components and are modelled using polynomial of the same degree.

### 5. Almost Optimal Detection Using Proposed Non-Linear Model of Nuisance Parameters

#### 5.1. Discussion on Non-Linear GLR test

When the discontinuity parameters \(\eta_i\) are known, it is obvious that the proposed locally adapted model (29) is linear. Hence, the application of the invariance theory, as described in Section 3, is straightforward.

Unfortunately, the discontinuity parameters \(\eta_i\) are unknown in practice. In such a situation, a usual solution is to replace the unknown parameters by their Maximum Likelihood (ML) estimation to design a Generalized Likelihood Ratio Test (GLRT).

But for the considered anomaly detection problem, the design of a GLRT suffers from several drawbacks.

On a theoretical point of view, the non-asymptotic optimality of the GLRT remains an open problem. Under some regularity conditions and when the number of observations growth to infinity, optimal asymptotic properties of the GLRT can be shown [50, 60]. However, this remains questionable when the number of observations is limited or when the required regularity conditions are not satisfied, see [45, 62] for details.

On a more practical point of view, the main difficulty of the proposed model (29) is due to the fact that expectations \(\mu_i\) are non-linear with respect to discontinuity parameters \(\eta_i\). Though accurate estimations can be obtained numerically, using usual iterative algorithms optimization [43], it is proposed in this paper to avoid this approach. In fact, estimations obtained using such algorithms are sensitive to initial conditions and often computationally expensive.

Moreover, the statistical properties of such estimated cannot be calculated which prevents us to establish the statistical performance of the GLRT.

In the present paper, these difficulties are overcome by using an original linearisation of the locally adapted model (29).

#### 5.2. Estimation of Observations Expectation and Variance

Using the proposed locally adapted model (LAM) given in Equation (29) each vector \(\mathbf{z}_i\) can be modelled as:
\[
\mathbf{z}_i \sim \mathcal{N}(\mu_i, \sigma^2 \mathbf{I}_L), \text{ with } \mu_i = \mathbf{G}(\eta_i) \mathbf{v}_i.
\]

To estimate the expectation \(\mu_i\) from the observations \(\mathbf{z}_i\), the main difficulty is due to the discontinuities, which are non-linear with respect to the parameter \(\eta_i\). To tackle this difficulty, it is proposed in this paper to follow the general methodology proposed in [50]; this approach, depicted in the Figure 3 and exemplified in the Example 2, is based on a Taylor series expansion of the non-linearity which allows the writing of:
\[
\mu_i = \mathbf{H}_i + \mathbf{F}(\hat{\eta}_i) + \alpha_i \mathbf{F}(\hat{\eta}_i)(\eta_i - \hat{\eta}_i) + \varepsilon_i + \alpha_i||\eta_i - \hat{\eta}_i||^2,
\]
where \(\mathbf{F}(\hat{\eta}_i) = \left(\frac{\partial \mathbf{F}(x; \hat{\eta}_i)}{\partial \hat{\eta}_{i,1}}, \ldots, \frac{\partial \mathbf{F}(x; \hat{\eta}_i)}{\partial \hat{\eta}_{i,r}}\right)\).
is the Jacobian matrix of $F(\hat{\eta})$ with size $N \times 2r_i$, $\hat{\eta}$ is an estimation of discontinuity parameter $\eta$, and $\epsilon_i$ is the (deterministic) error due to the negligence of second order term in (31). This yields the locally-adapted linear model:

$$\mu_i = \hat{G}(\hat{\eta})v_i + \epsilon_i + o(||\eta - \hat{\eta}||^2_2),$$

(31)

with $\hat{G}(\hat{\eta}) = (H | F(\hat{\eta}) | \hat{F}(\hat{\eta}))$ and

$$v_i = \begin{pmatrix} s_i \\ u_i \\ u_i(\eta - \hat{\eta}) \end{pmatrix}.$$

From (30) - (31), the background, or nuisance parameter $\mu_i$, subtraction can be achieved using the following property:

$$W_{G(\hat{\eta})}x_i = W_{G(\hat{\eta})}x_i + W_{G(\hat{\eta})}x_i + W_{G(\hat{\eta})}x_i + W_{G(\hat{\eta})}x_i = 0.$$

The relation (32) is obtained using the algebraic background subtraction $W_{G(\hat{\eta})}G(\hat{\eta})v_i = 0$.

**Example 2.** Let us consider the following non-linear model

$$\mu(x; \eta) = s + u \exp(-\eta x^2).$$

After sampling, along the variable $x$, the observation is given by:

$$z = \mu + \xi = Hs + uF(\eta) + \xi.$$

with $H = \begin{pmatrix} \vdots \\ 1 \end{pmatrix}$ and $F(\eta) = \begin{pmatrix} \exp(-\eta x^2) \\ \exp(-x(\eta + 1)/\eta) \end{pmatrix}$.

In this example, the non-linear parameter $\eta$ and the linear parameters $s$ and $u$ are unknown. The proposed methodology relies on an estimation $\hat{\eta}$ of the parameter $\eta$ in order to design the linearised model $G(\hat{\eta})$. Here, a direct calculation allows the writing of:

$$\frac{\partial}{\partial \eta} \exp(-\eta x^2) = \frac{2x^2}{\eta} \exp(-\eta x^2).$$

Hence the proposed linearisation method consist in using $F(\eta)$ and its first derivative both evaluated around the estimation $\hat{\eta}$. Using the previous relation and a first-order Taylor expansion allow the writing of:

$$\mu(x; \eta) = \mu(x; \hat{\eta}) + 2u(\eta - \hat{\eta}) \frac{x^2}{\eta} \exp(-\eta x^2) + o(\eta - \hat{\eta}).$$

which, after sampling, yields the following linear model:

$$z = \mu + \xi = Hs + uF(\hat{\eta}) + 2u(\eta - \hat{\eta})\hat{F}(\hat{\eta}) + \epsilon + \xi$$

with $\hat{F}(\hat{\eta}) = \begin{pmatrix} \frac{x^2}{\hat{\eta}} \exp(-\hat{\eta} x^2) \\ \frac{1}{\hat{\eta}} \exp(-x(\eta + 1)/\eta) \end{pmatrix}$.

Finally, defining the matrix $\hat{G}(\hat{\eta}) = (H | F(\hat{\eta}) | \hat{F}(\hat{\eta}))$, the proposed methodology allows the linearisation of the model:

$$\mu = \hat{G}(\hat{\eta})v \quad \text{with} \quad v = \begin{pmatrix} s \\ u \\ 2u(\eta - \hat{\eta}) \end{pmatrix}.$$

As described in Section 5.3, this linearisation methodology allows us to estimate easily the unknown parameters and also permits us to calculate the consequence on the statistical performance of the proposed detector.

Before describing and analysing the proposed test, based on background subtraction (32), it is necessary to address the problem of estimating noise variance. The noise can accurately be modelled as the sum of a Poissonian random variable and a Gaussian random variable. The former models the shot noise, due to the photo-counting process, while the later models the electronic and read-out noises. In the field of radiography inspection, the Poissonian model is usually used when the mean number of photons reaching the sensor is small (for instance in astrophysics or in low-dose medical imaging). On the contrary, the Gaussian model is used when the mean number of photons is high such as in most NDT applications.

In the present paper, it is proposed to model all the noises as a sum of a Poissonian random variable and a Gaussian random variable. Using the Gaussian approximation of Poisson random variable, yields a Gaussian heteroscedastic model [22, 29, 55, 56] in which, the variance of each observation depends on the observation expectation. This relation is given by $\sigma_i^2 = a \mu_i + b$ where $a$ and $b$ only depend on the imaging system and, hence, remain constant for a set of radiographies.

For a given vector of observations $z_i$ the ML estimation of the variance is given by:

$$\sigma_i^2 = \frac{1}{J - p - 3r_i} \|W_{G(\hat{\eta})}^2 \|^2_2.$$

(33)

However, it is useless to exploit directly this estimation because putting it into the decision statistics yields

$$\frac{1}{\sigma_i^2} \|W_{G(\hat{\eta})} z_i \|^2_2 = J - p - 3r_i$$

which obviously makes no sense because it completely prevents anomaly detection.

To overcome this problem, it is proposed in the present paper...
to estimate the parameters $a$ and $b$ which entirely characterize the noise variance. Those parameters are estimated with the weighted least-square (WLS) approach by gathering observations from all the radiographies, see details in [55]. To this end, the ML estimation of expectation of each vector $z_i$ is calculated as follows:

$$
\hat{\mu}_i = P_{G_i|\hat{\eta}_i} z_i. \quad (34)
$$

Here $P_{G_i|\hat{\eta}_i} = G(\hat{\eta}_i) (G(\hat{\eta}_i)^T G(\hat{\eta}_i))^{-1} G(\hat{\eta}_i)^T$ and $P_{G_i|\hat{\eta}_i} = I - P_{G_i|\hat{\eta}_i}$ are respectively the projection matrices onto subspace spanned by $G(\hat{\eta}_i)$ and its orthogonal complement, the parity space, see Section 3.1. All, the observations which approximatively have the same estimated expectation $\hat{\mu}_i$, are grouped together and the variance of this set of observations is estimated using the Maximum Likelihood (ML) method. The couples of estimated expectations and associated estimated variances finally allow estimating the linear parameters $a$ and $b$, which are then used to obtain variances as follows $\hat{\sigma}^2 = \hat{\alpha} \hat{\mu} + \hat{b}$. This allows for the normalising of “residuals” as follows:

$$
\frac{1}{\hat{\sigma}^2} P_{G_i|\hat{\eta}_i} z_i = \frac{1}{\hat{\alpha} \hat{\mu} + \hat{b}} P_{G_i|\hat{\eta}_i} z_i. \quad (35)
$$

This estimation is asymptotically equivalent to the ML estimation of parameters $a$ and $b$, and, especially allows the establishment of statistical properties of estimated variances. Because this aspect does not lies at the core of this paper, the reader interested can see details in [22] and particularly in [55] about the properties of WLS estimates.

5.3. Design and Performance of the Proposed Invariant Test

Using the linearisation approach (31) and ensuing algebraic background subtraction, or nuisance parameter rejection (32), allows the design of the proposed linearised GLRT:

$$
\hat{\delta}(\bar{Z}) = \left\{ \begin{array}{ll}
\mathcal{H}_0 & \text{if } \Lambda(\bar{Z}) = \sum_{i=1}^{l} \frac{1}{\hat{\sigma}^2_i} \|W_{G_i|\hat{\eta}_i} z_i\|_2^2 < \tau_{m_0} \\
\mathcal{H}_1 & \text{if } \Lambda(\bar{Z}) = \sum_{i=1}^{l} \frac{1}{\hat{\sigma}^2_i} \|W_{G_i|\hat{\eta}_i} z_i\|_2^2 \geq \tau_{m_0}. 
\end{array} \right. \quad (36)
$$

To establish the properties of the GLR test $\delta$ (36), the main difficulty is to analyse the error term, $W_{G_i|\hat{\eta}_i} z_i$, see (32), due to estimation error $\hat{\mu}_i \neq \bar{\mu}_i$. The impact of this term on GLR $\hat{\delta}$ can not be calculated but can only be bounded. To this end, the non-linear parameters must be bounded [33]. Hence, it is assumed that the error on the estimation of non-linear parameters is bounded by a (small) constant $||\eta_i - \hat{\eta}_i|| \leq \zeta$. The literature proposes methods which give such estimates, see [8, 38].

For the sake of clarity, it is assumed that one discontinuity (at most) is present in each vector $z_i$; the extension to the case of multiple discontinuities is straightforward at the cost of complicated notations. As shown in the Appendix C, assuming that the estimation error $||\eta_i - \hat{\eta}_i|| \leq \zeta$ is bounded by $\zeta$, permits the writing of:

$$
\frac{\|W_{G_i|\hat{\eta}_i} z_i\|_2^2}{\hat{\sigma}^2} \leq \frac{\hat{\mu}^2}{4\pi\sigma^4} \zeta^2,
$$

which immediately allows us to obtain:

$$
\sum_{i=1}^{l} \frac{1}{\hat{\sigma}^2_i} \|W_{G_i|\hat{\eta}_i} z_i\|_2^2 \leq \sum_{i=1}^{l} \frac{\hat{\mu}^2}{4\pi\sigma^4} \zeta^2 = b_{\max}. \quad (37)
$$

It follows from Equation (37), and properties of Gaussian random variables, that under hypothesis $\mathcal{H}_0$ one has:

$$
\Lambda \sim \chi^2(2; \lambda^2), \quad (38)
$$

where $\hat{\delta}$ the number of degree of freedom is given $\hat{\delta} = \sum_{i=1}^{l} J - p - 3r_i$ and the unknown non-centrality parameters $\lambda^2$ satisfies $\lambda^2 \leq b_{\max}$.

Similarly, under hypothesis $\mathcal{H}_1$, the Equation (37) and properties of Gaussian random variables permit the establishment of GLR $\hat{\delta}$ statistical distribution:

$$
\Lambda \sim \chi^2(2; \hat{\lambda}^2 + \lambda^2) \quad \text{with } \hat{\lambda}^2 = \sum_{i=1}^{l} \frac{1}{\hat{\sigma}^2_i} \|W_{G_i|\hat{\eta}_i} a_i\|_2^2. \quad (39)
$$

Here, $\hat{\lambda}^2$ represents the estimated “Anomaly-to-Noise Ratio” (ANR).

From the distribution of GLR $\hat{\delta}$ given in Equations (38) - (39), a short algebra permits the establishing of proposed GLRT test $\hat{\delta}$ (36) properties in the following Theorem 1.

**Theorem 3.** Assuming model (29) holds and that $\hat{\eta}_i$ is an unbiased estimator of $\eta_i$, satisfying $||\eta_i - \hat{\eta}_i|| \leq \zeta$, then it follows from (38) that for any $\alpha_0 \in (0; 1)$ the decision threshold:

$$
\tau_{m_0} = F^{-1}_{\chi^2(2; \lambda^2)}(1 - \alpha_0; b_{\max}) \quad (40)
$$

guarantees that $\mathbb{P}_{\mathcal{H}_0}[\Lambda(\bar{Z}) \geq \tau_{m_0}] \leq \alpha_0$ so that the GLR test $\hat{\delta}$ (36) is in the class $\mathcal{K}_{m_0}$.

Choosing the threshold $\tau_{m_0}$ as defined in (40), it follows from distribution (38) that the power function associated with the GLR test $\hat{\delta}$ (36) is bounded by:

$$
1 - F_{\chi^2(2; \lambda^2)}(\tau_{m_0}; \hat{\lambda}^2 + b_{\max}) \leq \hat{\beta}(\alpha_0) \leq 1 - F_{\chi^2(2; \tau_{m_0}; \hat{\lambda}^2)}(\tau_{m_0}; \hat{\lambda}^2). \quad (41)
$$

**Proof.** Of the Theorem 3 is given in the Appendix C.

The Theorem 3 shows that the loss of power of the proposed GLRT $\hat{\delta}$, compared to the optimal UBPC test $\delta^*$, is due to the two following reasons. First, the linearisation (31) involves using matrix $G(\hat{\eta}_i)$ instead of matrix $G(\eta_i)$. As the matrix $G(\eta_i)$, used in practice, has more columns that $G(\eta_i)$, there is a smaller number of “degree of freedom” in the parity subspace: $\hat{\delta} \leq \delta$. Second, the linearisation (31) is not perfect. There is an error of approximation due to estimation error $||\eta_i - \hat{\eta}_i|| \leq \zeta$ and neglected higher order terms which cause the bias $\lambda_i^2$ in the calculation of GLR $\hat{\delta}$. Fortunately, provided that the number of discontinuities $r$ and error on estimation of discontinuity parameters $\zeta$ are sufficiently small, the GLR test $\hat{\delta}$ performs almost as well as the optimal UBPC test $\delta^*$.

**Remark 1.** It should be noted that the application of the proposed test to the anomaly detection / localisation problem is
straightforward. In fact, by calculating the proposed GLRT by columns and by rows, it is possible to localise the anomaly, see an example in the Figure 8. This only requires a sufficient number of pixels compared to the ANR, exactly as the proposed GLRT applied for the whole radiographic image.

5.4. Main Strengths and Limits of the Proposed Methodology

Before discussing the numerical results presented in Section 6 let us briefly summarize the limits and the main strengths of the proposed methodology.

- The main strength of the proposed methodology is that its application do not require any prior information, because all the unknown parameters are estimated (opposed to [21]). In fact, the matrices \( G(\hat{\eta}) \) and \( W_{G(\hat{\eta})} \) allow the designing of a parametric linear model automatically adapted to the structured content of inspected objects.

- In order to apply the proposed method to a wide range of applications, the noise is modelled as a sum of Poissonian and Gaussian random variables. This includes the two main trends of radiographies noise modelling, for small and high mean number of incident photons. In the proposed methodology only parameters \( \alpha \) and \( \beta \), from the heteroscedastic noise model (33) - (35), are modified according to the mean number of incident photons.

- Finally, the proposed methodology allows the establishing of GLRT statistical properties. In particular, this permits us to guarantee a prescribed false-alarm probability; the expression of proposed GLRT power function also provides an insight on the detectability of a given anomaly.

Besides of these main strengths, the proposed methodology has relative limitations which are summarized below.

- The adaptation of parametric linear model of observations has the main drawback that it can not guarantee the detectability of an anomaly. In fact, the ANR \( \hat{\varrho} \) (39) might be negligible if the anomaly is modelled as part of the non-anomalous background. However, in any NDT applications, the need of an automatic detection scheme is crucial to detect small defects, which are not modelled as background: the detection of defects of large intensity is straightforward from a human point of view.

- The proposed methodology adapts the matrix \( G(\hat{\eta}) \) by adding basis vectors to model blurred edges. Thus, for very textured areas, this might decreases the performance of the proposed test. However, the results presented in Section 6 show that this phenomenon does not occur for inspection of usual radiography. In addition, very textured areas are well-known to be difficult to model and, hence, it is usually admitted that defects in such regions are very difficult to detect.

It also worth noting that the proposed methodology is based on general statistical concepts. Hence, it can be extended to other NDT technologies, provided that a rigorous model of acquisition process is available; this is outside the scope of this paper.

6. Numerical Results

6.1. Verification of Theoretical Results on Simulated Data

On the main goal of the proposed methodology is to address the problem of anomaly detection by providing a test statistically with analytically established properties. Particularly, it is aimed to establish the decision threshold which guarantees a prescribed false-alarm probability and to establish the detection power of the proposed test as the function of Anomaly-to-Noise Ratio, the ANR \( \varrho \).

To verify numerically the relevance and the sharpness of theoretically established results, it is proposed to extract a segment, from a radiography of a nuclear fuel rod, and to perform a Monte-Carlo simulation based on these data. The segments used for this experimentation are shown by the blue line in Figures 6a and 6c. A Gaussian noise, which follows the heteroscedastic model (33) - (35), was added to the extracted line. This line is analysed 50 000 times and modelled by two segments of 75 samples, or pixels, described by a polynomial of order \( p = 1 = 3 \). Two segments, with and without anomaly, are used to obtain results under both hypotheses \( H_0 \) and \( H_1 \), see also data presented in Figures 5a and 5b.

The comparison between the empirical results, obtained from
Monte-Carlo simulation, and the theoretically established ones are presented in the Figure 4a. This figure shows the results as ROC curves; that is, the detection powers, $\beta^\ast$ (19) and $\hat{\beta}$ (41), are plotted as functions of false-alarm probability $\alpha_0$ for a fixed ANR.

Similarly, Figure 4b presents a comparison between theoretical and empirical detection powers as a function of ANR and for a fixed false-alarm probability $\alpha_0 = 0.05$. Those results are obtained using a Monte-Carlo simulation with the same settings. In order to vary the ANR, the experimentation is performed on simulated data which are based on the observed segment, with and without anomaly.

The results presented in Figures 4a and 4b show the relevance of the proposed methodology and the sharpness of the theoretically established results. In addition, these results also emphasize that the detection power $\hat{\beta}$ cannot be calculated but is bounded, because the error on discontinuity parameters remains unknown in practice, see Equation (41).

6.2. Numerical Results on Real Radiographic Images

To complete the previous numerical results, it is proposed to illustrate the results of background subtraction on segments used in the Figure 4. Those segments are shown by the blue line in Figures 6a and 6c. The observed data from a few line of the noisy radiography are shown in the Figure 6a and 6b. Those figures also show the results from background subtraction, or nuisance parameter rejection. As described above, those line are modelled by two segments of $J = 75$ samples, or pixels, modelled by a polynomial of order $p - 1 = 3$ (for readability, only a portion of those lines are drawn in the Figure 5).

The Figures 5a and 5b respectively show the data obtained from lines without and with anomaly. Obviously the proposed methodology allows the rejection of nuisance parameter while preserving the presence of the anomaly. It should be noted that obviously, the first part of the line can not be modelled using a mere algebraic polynomial of degree 3. However, thanks to the proposed locally adapted model of scattered discontinuities, the residuals calculated from background subtraction are free from
inaccurate regression artefacts. In addition, it should be mentioned that the position of anomaly is rather close to the position of discontinuity. In such a situation, it is crucial to model accurately the discontinuity to avoid a potential obliteration of the anomaly signature by the background subtraction, which would prevents its detection.

The inspection of a nuclear fuel rod is presented through two radiographies, shown in the Figure 6. The Figures 6a and 6d show the two radiographies. Is should be noted that among all the data, those two observations are chosen because they show the cases of a radiography which contains almost invisible anomalies (slight voids, see [58] for a detailed classification of defects), Figure 6a, and a radiography with contains an obvious anomaly, Figure 6d. The residuals obtained from background or nuisance parameter subtraction as follows \( z_i - P_{G(\hat{\eta})}^c z_i = P_{G(\hat{\eta})}^c z_i \) are shown in Figures 6b and 6e. Those results particularly emphasize the relevance of heteroscedastic noise model; the variance of the central area, more exposed, is obviously much more important than the variance of residuals at the border, where the mean number of incident photons is much smaller. Finally, the Figures 6c and 6f show the “normalised” residuals obtained from background subtraction where each sample or pixel has been normalised by its estimated variance, as given in Equation (35). These two last figures particularly emphasize the accuracy of proposed samples variance estimation to obtained “normalised” or “whitened” residuals which exhibits the same variance. As previously discussed in Section 3, this property is particularly important to allows the application of the proposed statistical hypothesis test.

In order to show potential applications of the proposed model, and ensuing statistical test, to a wide range of anomalies detection problems, the Figure 7 presents results obtained from medical images. Similarly to Figure 6, the results presented in the Figure 7 show the application of proposed model to detect a incomplete transverse fracture. The Figure 7a shows the original longitudinal radiograph of a forearm, focussing on the radius area which exhibits a slight fracture. The Figure 7b shows the noise residuals obtained using the background or nuisance parameter subtraction \( z_i - P_{G(\hat{\eta})} z_i = P_{G(\hat{\eta})} z_i \). Figure 7c shows the normalised or whitened residuals obtained using the heteroscedastic noise property to estimated the variance of each pixel, as described in Section 5.2. Note that the radiograph is analysed line by line; each line is made of 248 pixels, which are divided into four segments of 62 observations whose smooth part \( \mu_i^{(c)} \) are modelled by a polynomial of degree \( p - 1 = 3 \). Finally, Figures 7d and 7e shows the residuals obtained from the well-known wavelets approximation. To emphasize the interest of the proposed locally adapted model, the Figure 7d shows the residuals obtained from a “fine-scale” wavelet decomposition, with the 6-th Daubechies wavelet, level of decomposition and soft-thresholding of wavelet coefficients set for each level to half the minimax threshold, see [20, 19] for details. Similarly, Figure 7e shows the residuals obtained from a “coarse-scale” wavelet decomposition, with the same wavelet, level of decomposition and a threshold set, for each level, to 1.5 time the minimax threshold. These two figures highlight the difficulty of distinguishing the non-anomalous background from anomaly when using a model which is not adapted to the content of radiographs. Particularly, it can be seen that a precise approximation lead to consider anomaly as part of the non-anomalous background while, on the contrary, a coarse approximation creates artefacts with similar impact on the test than the anomaly. In addition, it should be noted that wavelets, as most of other image models, are global which prevents the estimations of each pixel variance.

Finally, Figure 8 show the extension of the proposed methodology to address the anomaly detection and localisation problem. To this end, the GLR \( \Lambda(36) \) is calculated line by line and column by column to estimate the location of potentially detected anomalies. Figure 8 shows normalised residuals from Figure 6f together with the GLR \( \Lambda(36) \) calculated line by line (on the right hand side) and column by column (on the bottom). Note that, for a better readability, the theoretical mean of the GLR \( \Lambda(36) \) under \( H_0 \) is shown by light blue dashed lines.

Figure 8 obviously shows that the main anomalies can clearly be localised from the GLR \( \Lambda(36) \) calculated line by line and column by column. Moreover, it can be observed from Figure 8 that the discontinuities in the inspected radiography do not create artefacts that largely impact the GLR.
Appendix A. UBCP Test: Linear Model

Let us recall that each segment $z_i$ from the inspected radiographic image $Z$ is described by the following model, $\forall i \in \{1, \ldots, I\}$:

$$
\begin{align*}
\mathcal{H}_0 &= \{ z_i = Hs_i + \xi_i, s_i \in \mathbb{R}^p \}, \\
\mathcal{H}_1 &= \{ z_i = Hs_i + a_i + \xi_i, s_i \in \mathbb{R}^p, a_i \in \mathbb{R}^L \},
\end{align*}
$$

By using the projection matrix, denoted $W$, onto the orthogonal complement of $H$, the parity vector defined by:

$$
\tilde{z}_i = Wz_i,
$$

verifies

$$
\begin{align*}
\tilde{z}_i &= \left\{ \begin{array}{ll}
W\xi_i & \sim \mathcal{N}(0, \sigma_i^2 I_{p-p}), \\
Wa_i + W\xi_i & \sim \mathcal{N}(0, \sigma_i^2 I_{p-p}),
\end{array} \right.
\end{align*}
$$

and that the UBCP test is defined as:

$$
\delta^*(Z) = \begin{cases}
\mathcal{H}_0 & \text{if } \Lambda(Z) < \tau^*_0, \\
\mathcal{H}_1 & \text{if } \Lambda(Z) \geq \tau^*_0,
\end{cases}
$$

where the decision statistics $\Lambda(Z)$ is given by:

$$
\Lambda(Z) = \sum_{i=1}^I \Lambda(z_i) = \sum_{i=1}^I \frac{||\tilde{z}_i||^2}{\sigma_i^2} = \sum_{i=1}^I \frac{||Wz_i||^2}{\sigma_i^2}.
$$

From the statistical properties of Gaussian random variables, it is well known that under hypothesis $\mathcal{H}_0$:

$$
\Lambda(Z) \sim \chi^2_p, \quad (A.1)
$$

where $\chi^2_p$ represents the central $\chi^2$-squared distribution with $p$ degree of freedom. On the opposite, under hypothesis $\mathcal{H}_1$:

$$
\Lambda(Z) \sim \chi^2_p(\varrho^2), \quad (A.2)
$$

where the non-centrality parameter $\varrho$ is given by:

$$
\varrho^2 = \sum_{i=1}^I \frac{||Wa_i||^2}{\sigma_i^2} \quad (A.3)
$$

Hence, for any decision threshold $\tau^*_0$, it follows from (A.1) that the false-alarm probability $\alpha_0$ is given by:

$$
\alpha_0 = \mathbb{P}_{\mathcal{H}_0} \left[ \Lambda(Z) \geq \tau^*_0 \right] = 1 - F_{\chi^2_p} \left( \tau^*_0, 0 \right).
$$

Where $F_{\chi^2_p}(\cdot)$ represents the cumulative distribution function of $\chi^2$ distribution with $p$ degrees of freedom and $F_{\chi^2_p}^{-1}$ its inverse. Therefore, for any prescribed false-alarm probability $\alpha_0$, the associate decision threshold $\tau^*_0$ is given by:

$$
\alpha_0 = 1 - F_{\chi^2_p} \left( \tau^*_0, 0 \right) \Leftrightarrow \tau^*_0 = F_{\chi^2_p}^{-1} (1 - \alpha_0, 0). \quad (A.4)
$$
Similarly, it follows from the distribution of $\Lambda(Z)$ under hypothesis $H_1$, see Equation (A.2), that for any decision threshold $\tau_{a_0}$ the power of the UBCP test is given by:

$$\beta^*(\varrho, \tau_{a_0}^*) = \mathbb{P}_{H_1}\left[\Lambda(Z) \geq \tau_{a_0}^*\right] = 1 - F_{\tilde{\chi}_1}^{-1}\left(1 - \alpha_0; \varrho\right).$$

Substituting the threshold $\tau_{a_0}$ by the value given in Equation (A.4) yields:

$$\beta^*(\varrho, 0) = 1 - F_{\tilde{\chi}_1}^{-1}\left(1 - \alpha_0; \varrho\right). \quad (A.5)$$

Appendix B. Demonstration of Theorem 2

Let $(x_{d, i}, y_{i, d})$ be the coordinates of the $d$-th discontinuity in the $k$-th segment and $\varphi$ the angle between the local normal to the discontinuity and the analysed segment. Using the appropriate coordinate transformation, $\Psi : (\bar{x}, \bar{y}) \to (x, y)$ defined as $\Psi : (\bar{x}, \bar{y}) = (x_k + \bar{x} \cos(\psi_k) + \bar{y} \sin(\psi_k), y_k + \bar{x} \sin(\psi_k) - \bar{y} \cos(\psi_k))$ a short calculation shows that:

$$f_k^{(3)}(\bar{x}, \bar{y}) = h * (u_{k, d} \delta(\bar{x})) = u_{k, d} \int_{-\infty}^{\infty} h(u, v) \, du \, dv$$

$$= \frac{u_{k, d}}{s_{p/d} \sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2}\right) \, du = u_{k, d} \Phi\left(\frac{\bar{x}}{s_{p/d}}\right).$$

Using the inverse coordinate transformation $\Psi^{-1}$ provide the results given in Equation (27) with $s_{k, d} = s_{p/d} / \cos(\psi)$.

Appendix C. GLR test with Proposed Locally Adapted Model

Let us recall that, for the sake of clarity, only one discontinuity is present (at most) in each segment and that the mean of each segments is given from Equation (29) by:

$$\mu_i = H\mathbf{s}_i + u_i \mathbf{F}(\eta) = G(\eta_i)\mathbf{v}_i.$$ 

The methodology proposed in Section 5.1 is relies on estimations of $\eta_i$ around which the model (29) is linearised as follows:

$$f(x; \eta_i) = f(x; \bar{\eta}_i) + \bar{\mathbf{F}}(x; \bar{\eta}_i)(\eta_i - \bar{\eta}_i)$$

$$+ \frac{1}{2}(\eta_i - \bar{\eta}_i)^T \bar{\mathbf{F}}(x; \bar{\eta}_i)(\eta_i - \bar{\eta}_i) + o(||\eta_i - \bar{\eta}_i||^2_2),$$

where $\bar{\mathbf{F}}$ and $\bar{\mathbf{F}}$ respectively represents the the Jacobian matrix and the Hessian matrix of function $f(x; \eta_i)$. Dropping the second order term yields the model (31):

$$\mu_i = H\mathbf{s}_i + u_i \mathbf{F}(\bar{\eta}_i) + u_i \mathbf{F}(\eta_i)(\eta_i - \bar{\eta}_i) + \varepsilon_i + o(||\eta_i - \bar{\eta}_i||^2_2),$$

$$= G(\bar{\eta}_i)\mathbf{v}_i + \varepsilon_i + o(||\eta_i - \bar{\eta}_i||^2_2),$$

with $\mathbf{G}(\bar{\eta}_i) = \begin{bmatrix} \mathbf{H} & \mathbf{F}(\bar{\eta}_i) & \bar{\mathbf{F}}(\bar{\eta}_i) \end{bmatrix}$ and $\mathbf{v}_i = \begin{bmatrix} \mathbf{s}_i & u_i & (\eta_i - \bar{\eta}_i) \end{bmatrix}$.

As discussed in Sections 5.2 - 5.3 it is crucial to calculate the the (deterministic) error $\varepsilon_i$ due to the negligence of second order term. For the sake of clarity, the index $i$ of segments are omitted from the analytic calculus below. In the following, the goal is to study the norm of the residuals errors. To this end, it is necessary to study the derivatives and to calculate the normal and tangential curvatures due to the second order derivatives, see [50]. Let us define $f(x; \eta) = \Phi(\bar{\mathbf{x}})$. A straightforward differentiation calculus gives

$$\hat{f}_i = \frac{\partial f(x; \eta)}{\partial t} = -\frac{1}{\varsigma \sqrt{2\pi}} \exp\left(-\frac{(x-t)^2}{2\varsigma^2}\right) = -\phi\left(\frac{x-t}{\varsigma}\right),$$

$$\hat{f}_c = \frac{\partial f(x; \eta)}{\partial c} = -\frac{(x-t)}{\varsigma^2 \sqrt{2\pi}} \exp\left(-\frac{(x-t)^2}{2\varsigma^2}\right) = -\frac{x-t}{\varsigma} \phi\left(\frac{x-t}{\varsigma}\right)$$

where $\phi(u) = (2\pi)^{-1/2} \exp(-u^2/2)$. This immediately permits the writing of the norms of these functions:

$$\|\hat{f}_i\|^2 = \int_R \hat{f}_i(x; \eta)^2 \, dx = \frac{1}{2\varsigma \sqrt{\pi}}, \quad \|\hat{f}_c\|^2 = \int_R \hat{f}_c(x; \eta)^2 \, dx = \frac{1}{4\varsigma^3 \sqrt{\pi}}.$$ 

Similarly, for the second order derivatives, a short calculus gives:

$$\hat{f}_{i, t} = \frac{\partial^2 f(x; \eta)}{\partial t^2} = -\frac{x-t}{\varsigma^3 \sqrt{2\pi}} \exp\left(-\frac{(x-t)^2}{2\varsigma^2}\right) = \frac{\hat{f}_i}{\varsigma^3},$$

$$\hat{f}_{i, c} = \frac{\partial^2 f(x; \eta)}{\partial c^2} = \frac{(x-t)^3}{\varsigma^5 \sqrt{2\pi}} \exp\left(-\frac{(x-t)^2}{2\varsigma^2}\right) = \frac{-2\hat{f}_i}{\varsigma^3},$$

$$\hat{f}_{i, \varsigma} = \frac{\partial^2 f(x; \eta)}{\partial \varsigma \partial c} = \frac{\hat{f}_i}{\varsigma^2} - \frac{(x-t)^2}{\varsigma^4} \phi\left(\frac{x-t}{\varsigma}\right).$$

Again, the norms of these functions are given by:

$$\|\hat{f}_{i, t}\|^2 = \frac{1}{4\varsigma^3 \sqrt{\pi}}, \quad \|\hat{f}_{i, c}\|^2 = \frac{7}{16\varsigma^3 \sqrt{\pi}}, \quad \|\hat{f}_{i, \varsigma}\|^2 = \frac{3}{8\varsigma^3 \sqrt{\pi}}.$$ 

The first derivatives, $\hat{f}_i$ and $\hat{f}_c$, are respectively even and odd functions, therefore, those functions are orthogonal:

$$\langle \hat{f}_i, \hat{f}_c \rangle = \int_R \hat{f}_i(x; \eta)\hat{f}_c(x; \eta) \, dx = 0.$$ 

Let us now calculate the projection of the second order derivatives onto the first order derivatives in order to calculate the normal and tangential curvatures. Since $\hat{f}_c = \varsigma^{-1}\hat{f}_i$ that this whole vector belongs to the tangential curvature, and therefore:

$$\hat{f}_{i, t} = \hat{f}_i \quad \text{and} \quad \hat{f}_{i, \varsigma} = 0.$$ 

For the two other second order derivatives the calculus are slightly more complicated and, using a method similar to the Gram-Schmidt process, lead to:

$$\hat{f}_{i, c} = -\frac{1}{2\varsigma}\hat{f}_i \quad \text{and} \quad \hat{f}_{i, \varsigma} = -\frac{1}{2\varsigma}\hat{f}_i,$$

which, from the orthogonality between tangentials and normals components, permits to write:

$$\hat{f}_{i, c} = \hat{f}_{i, \varsigma} = \frac{\hat{f}_{i, t}}{\hat{f}_i} \Rightarrow \|\hat{f}_{i, c}\|^2 = \frac{3}{8\varsigma^3 \sqrt{\pi}},$$

and

$$\hat{f}_{i, \varsigma} = -\frac{1}{2\varsigma}\hat{f}_i \Rightarrow \|\hat{f}_{i, \varsigma}\|^2 = \frac{1}{4\varsigma^3 \sqrt{\pi}}.$$
It finally follows from previous calculations that, after sampling:

$$\left\| \mathbf{W} \mathbf{g} / (\mathbf{u}, \mathbf{m}) \right\|_2^2 \leq \frac{2}{\mathcal{N}} \int_\mathcal{R} \frac{\eta - \hat{\eta}}{\nu} \mathbf{y}(x, \mathbf{u}, \mathbf{m}) \mathbf{y}^\top(x, \mathbf{u}, \mathbf{m}) \, dx,$$

$$= \frac{\nu^2}{\nu^2} \frac{1}{\mathcal{N}} \left( \frac{\left| \mathbf{t} - \mathbf{t}^\star \right|}{\mathcal{N}} + \frac{\left| \mathbf{e} - \hat{\mathbf{e}} \right|}{\mathcal{N}} \right)^2 \, o \left( \frac{\left| \eta - \hat{\eta} \right|}{\mathcal{N}} \right)^2.$ $

Putting the term $|\eta - \hat{\eta}|$ immediately yields:

$$\left\| \mathbf{W} \mathbf{g} / (\mathbf{u}, \mathbf{m}) \right\|_2^2 \leq \frac{t^2}{\nu^2} |\eta - \hat{\eta}| \leq \frac{\nu^2}{\nu^2} \frac{1}{\mathcal{N}} \left( \frac{\left| \mathbf{t} - \mathbf{t}^\star \right|}{\mathcal{N}} + \frac{\left| \mathbf{e} - \hat{\mathbf{e}} \right|}{\mathcal{N}} \right)^2 \, o \left( \frac{\left| \eta - \hat{\eta} \right|}{\mathcal{N}} \right)^2,$$

which finally proves Equation (37) and ends this proof.

Theorem 3 immediately follows from the previous results and straightforward probability calculus very similar to those details in the Appendix A; they are hence not again detailed here.

References


