A statistical detection of an anomaly/target from noisy tomographic projections

Lionel Fillatre and Igor Nikiforov
LM2S/UTT, 12, rue Marie-Curie, BP 2060, 10010 Troyes Cedex, France

Abstract
The problem of detecting an anomaly/target from a limited number of noisy tomographic projections is addressed from the statistical point of view. We study an unknown two-dimensional scene composed of an environment, considered as a nuisance parameter set, with a possibly hidden anomaly/target. An optimal statistical invariant test is proposed to solve such a problem.

1 Introduction
Computerized tomography 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 non-invasive nature of tomography has made it very useful for a variety of applications, including medical imaging and quantitative non-destructive testing, among others [2, 5, 10].

The detection of an anomaly/target from projections have been already studied in [11] where the authors estimate the unknown location of a parameterized object by maximum likelihood estimation. Recently, [4] proposes 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. Otherwise, such a problem can be viewed as an object recognition problem [13].

In certain practical applications, like baggage X-ray scanning or non-destructive testing, only few projections are available and a perfect reconstruction of the scene is impossible [5]. It is necessary to detect an anomaly/target possibly hidden by an unknown environment, considered as a nuisance parameter, directly from a very limited number of noisy tomographic projections. The paper is organized as follows. An elementary introduction to the tomographic problem is given in Section 2. Next, the anomaly/target detection problem statement is proposed in Section 3. An optimal invariant decision rule is designed for detecting the anomaly/target in Section 4. Finally, a simple numerical example is presented to illustrate the relevance of the developed tools in Section 5.

2 Physical background
Background material on the tomographic problem can be found in many sources [8, 11]. To simplify the presentation, we only study the two-dimensional (2D) case (the original scene is plane) but the three-dimensional (3D) case can be easily derived from the following results. Let us define the X-ray attenuation coefficient $s$ as a function of two variables $x$ and $y : (x, y) \rightarrow s(x, y)$. It is assumed that $s$ is an element of the space $L^2(D)$ of square-integrable real-valued functions supported on a compact set $D$ in the plane.

This situation is depicted in Figure 1. We consider a set of parallel-beam line integral projections, taken at $P$ view angles $\eta_i \in [-\pi, \pi]$, $i \in \{1, \ldots, P\}$. A projection $g$ at a particular angle $\eta$ is a real-valued function defined on the real line by:

$$g(t, \eta) = \int_0^L s(t \cos \eta - l \sin \eta, t \sin \eta + l \cos \eta) dl.$$
In many practical applications, projections are obtained from a linear numerical detector composed of \( n \) elementary cells counting the total amount of X-photons passing through the object (see Figure 1). Calling \( t_k \in \mathbb{R} \) the abscissa of the centre of the \( k \)-th cell, we finally dispose of a family \( \tau = \{t_1, \ldots, t_n\} \) of \( n \) points of interest, regularly spaced on the linear detector. Hence, the projection at a particular angle \( \eta \) is related to the original object \( s \) by the discrete Radon transform \( R_{\eta, \tau} : \mathcal{L}_2(\mathcal{D}) \to \mathbb{R}^n \) defined by:

\[
R_{\eta, \tau} s = \left( g(t_1, \eta), g(t_2, \eta), \ldots, g(t_n, \eta) \right)^T.
\]

(1)

From the practical point of view, the vector \( R_{\eta, \tau} s \) can be interpreted as a collection of line integrals calculated along the “source-to-sensor” direction at distinct points \( t_k \), where each line integral measures the X-ray attenuation coefficient distribution of the original scene.

3 Problem statement

In this section, we describe physical models related to the original scene and we finally introduce the decision problem to resolve. The original scene is composed of two main elements: the environment and the anomaly or target. Let us start with a short description of these main elements.

3.1 Environment model

In many practical applications, the original scene without target, also called environment, is space limited and it is natural to limit our study to a compact set \( \mathcal{D} \subset \mathbb{R}^2 \). Obviously, a global system of coordinates, noted \( \mathcal{R} = (O, x, y) \), is attached to the acquisition system, and, consequently, to the compact \( \mathcal{D} \). Now, the environment can be easily assimilated to a real-valued function \( h \in \mathcal{L}_2(\mathcal{D}) \). Physically, the function \( \{x, y\} \mapsto h(x, y) \) corresponds to the X-ray attenuation coefficient of the environment defined at the point \( (x, y) \in \mathcal{D} \). This situation is depicted in Figure 2. Moreover, for theoretical convenience, the environment...
Figure 2: The “source-to-sensor” direction attenuation coefficient (hypotheses $\mathcal{H}_0$ : there is no anomaly/target and $\mathcal{H}_1$ : there is an anomaly/target).

$h$ is supposed to be linearly parameterized:

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

with a known family of basic functions $\{h_1, \ldots, h_m\}$ in $L_2(\mathcal{D})$ and an unknown real-valued vector $\mu = (\mu_1, \ldots, \mu_m)^T$. In other words, the unknown environment $h$ is assumed to be composed of a finite number of a priori known basic elements $h_k$, representing a local X-ray attenuation density, with unknown scaling coefficients $\mu_k$.

### 3.2 Anomaly/target model

The anomaly, that is also called the target when a precise description of its physical/geometrical properties is available, is assimilated to a function $g_\theta \in L_2(\mathcal{D})$, parameterized by the vector $\theta \in \Theta \subset \mathbb{R}^p$ and having a compact support $d$ strictly included in $\mathcal{D}$. This situation is depicted in Figure 2. Hence, the difference between the anomaly/target attenuation coefficient and the environment one can be defined as follows:

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

Physically, the function $f_{\theta, \mu} \in L_2(\mathcal{D})$ represents the local variation of the attenuation coefficient due to the presence of the anomaly/target. In the rest of the paper, it is assumed that the anomaly/target is contrast with respect to the environment (see Figure 2). This requirement can be expressed in the following manner: $\int \int_d (f_{\theta, \mu}(x, y))^2 \, dx \, dy > 0$ for all $\mu \in \mathbb{R}^m$. The parameter vector $\theta = (\vartheta, \gamma, x_0, y_0)^T$ is composed of the sub-vector $\vartheta$ that contains some information about the X-ray attenuation coefficient distribution of the anomaly/target and the shape of the support $d = d_{x_0, y_0, \gamma}$ in the local system of coordinates noted $\mathcal{R}_{\text{loc}} = (G, u, v)$ (see Figure 1), the angle $\gamma$ that defines the orientation of $\mathcal{R}_{\text{loc}}$ in the global system of coordinates $\mathcal{R}$ (defined along the horizontal axis) and the coordinates $(x_0, y_0) \in \mathbb{R}^2$ of the point $G$ in the global system of coordinates (see Figure 1). To simplify the notation of the anomaly/target support $d_{x_0, y_0, \gamma}$, the subscripts $x_0, y_0, \gamma$ will be omitted in the rest of the paper.
3.3 Measurement model

Let us define now the following model of X-ray attenuation coefficient: \( (x, y) \mapsto s(x, y) \) that describes the both possible situations (\( \mathcal{H}_0 \) : there is no anomaly/target and \( \mathcal{H}_1 \) : there is an anomaly/target):

\[
\mathcal{H}_0 \quad \text{and} \quad \mathcal{H}_1
\]

(2)

Putting together equations (1) and (2), we obtain the following measurement model for a particular view angle \( \eta \):

\[
Y_\eta = R_{\eta, \tau} s + \xi,
\]

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

Due to the linearity of the operator \( R_{\eta, \tau} \), the above measurement model can be re-written as:

\[
Y_\eta = M_\eta(\theta, \mu) + H_\eta(\mu) + \xi,
\]

(3)

where \( M_\eta(\theta, \mu) = R_{\eta, \tau} f_{\theta, \mu} \) and \( H_\eta = (H_\eta^1, \ldots, H_\eta^m) \) with \( H_\eta^k = R_{\eta, \tau} h_k \in \mathbb{R}^n \) for \( k = 1, \ldots, m \).

This situation is depicted in Figures 2 and 2. If \( P \) projections are available, the vector \( Y_{\eta_1, \ldots, \eta_P} \), the function \( M_{\eta_1, \ldots, \eta_P}(\theta, \mu) \) and the matrix \( H_{\eta_1, \ldots, \eta_P} \) can be easily designed from “elementary” components \( Y_{\eta_i}, M_{\eta_i}(\theta, \mu) \) and \( H_{\eta_i} \). It is assumed that only one projection is available and the subscript \( \eta \) will be omitted in the rest of the paper to simplify the notations.

3.4 Anomaly/target detection problem

The detection problem consists in deciding which hypothesis is the true one, \( \mathcal{H}_0 \) there is no anomaly/target or \( \mathcal{H}_1 \) there is an anomaly/target, while considering the environment as nuisance parameter, with a limited number of projections. In the following, we must distinguish two possible problems according to our final goal. On the one hand, we are interested in detecting the presence of an anomaly without characterizing it precisely. For this reason it is assumed that the vector \( \mu = M(\theta, \mu) \) is unknown. This anomaly detection problem consists in testing between the two following hypotheses:

\[
\mathcal{H}_0 = \{ M = 0, \mu \in \mathbb{R}^m \} \quad \text{and} \quad \mathcal{H}_1 = \{ M \neq 0, \mu \in \mathbb{R}^m \},
\]

(4)

while considering \( \mu \) as an unknown vector, given the measurement vector \( Y \) defined by equation (3). On the other hand, if the physical/geometrical nature of the target \( M = M(\theta, \mu) \) is available in its parameterized form (defined by the vector \( \vartheta \)), the target detection problem consists in testing between the hypotheses (4) while considering the unknown orientation and position \( \gamma, x_0, y_0 \) together with the unknown environment \( H_\mu \) as nuisance. Only the anomaly detection problem will be discussed in the rest of the paper.

4 Statistical hypotheses testing

In this section, we first introduce the hypotheses testing. Particular emphasis is put on dealing with nuisance parameters and deciding between linear composite hypotheses.

4.1 Testing between two hypotheses

Key features of mathematical statistics theories and tools for solving hypotheses testing problems are their ability to handle noises and uncertainties, to select one among several hypotheses and to reject nuisance parameters.
Composite hypotheses testing problems. Let Y be a random variable with distribution $P_{\theta}$ belonging to the parametric family $\mathcal{P} = \{P_{\theta}\}$, where $\theta \in \Theta \subset \mathbb{R}^p$. Two types of hypotheses have to be distinguished. A simple hypothesis $H_i$ is defined by a unique value of the parameter vector: $H_i: \theta = \theta_i$, $i = 0, 1$. In contrast, with a simple hypothesis a composite one refers to a set of parameters $H_i: \theta \in \Theta_i$, with $\Theta_i \subset \Theta \subset \mathbb{R}^m$, $i = 0, 1$. We assume that $\Theta_0 \cap \Theta_1 = \emptyset$. The usage of composite hypotheses is more relevant in practice, because of limited available amount of information, especially for the alternative hypothesis (presence of anomaly). Assume that a sample of observations $Y \in \mathbb{R}^n$ is available and follows a particular unknown distribution $P_{\theta}$. A statistical test between the hypotheses is a measurable mapping $\delta : \mathbb{R}^n \mapsto \{H_0, H_1\}$ from observations space onto the set of hypotheses. The quality of a statistical test is defined with the probability of false alarm and the power of the test. Let us define the class $\mathcal{K}_\alpha = \{\delta : \text{sup}_{\theta \in \Theta_0} \text{Pr}_\theta(\delta = H_1) \leq \alpha\}$ of tests with upper-bounded maximum false alarm probability, where the probability $\text{Pr}_\theta$ stands for the vector of observations $Y$ being generated by distribution $P_{\theta}$ and $\alpha$ is the prescribed probability of false alarm. The power function $\beta_\delta(\theta)$ is defined with the probability of correct decision: $\beta_\delta(\theta) = \text{Pr}_\theta(\delta = H_1)$. Obviously, $\alpha$ should be as small as possible and $\beta_\delta(\theta)$ should be large for every $\theta \in \Theta_1$. The ideal solution is a uniformly most powerful (UMP) test $\delta^*$ in the class $\mathcal{K}_\alpha$, i.e.
\[
\forall \delta \in \mathcal{K}_\alpha, \forall \theta \in \Theta_1 : \beta_{\delta^*}(\theta) \geq \beta_\delta(\theta).
\] (5)

Unfortunately, UMP tests scarcely exist, except when the parameter $\theta$ is scalar, the family of distributions $\mathcal{P} = \{P_{\theta}, \theta \in \Theta\}$ has a monotone likelihood ratio and the test is one-sided, namely: $H_0 : \theta \leq \theta_0$ and $H_1 : \theta > \theta_0$ [1, 7]. In 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, Wald [12] proposes to impose an additional constraint on the class of considered tests, namely, a constant power function over a family of surfaces $S$ defined on the parameter space $\Theta$, in order to avoid the existence of UMP tests over a subspace $\overline{\Theta}$ of $\Theta$ which are very inefficient over $\Theta \setminus \overline{\Theta}$. A test $\delta^* \in \mathcal{K}_\alpha$ is said to have uniformly best constant power (UBCP) on the family of surfaces $S$, if the following conditions are fulfilled [12]:
1. for any pair of points $\theta_1$ and $\theta_2$ which lies on the same surface $S_\theta \in S$, $\beta_{\delta^*}(\theta_1) = \beta_{\delta^*}(\theta_2)$, where $\beta_\delta(\theta) = \text{Pr}_\theta(\delta = H_1)$ is the power function of the test $\delta$.
2. for another test $\delta \in \mathcal{K}_\alpha$, which satisfies the previous condition, we have $\beta_{\delta^*}(\theta) \geq \beta_\delta(\theta)$.

Nuisance parameters rejection. As it follows from equation (3), the vector of observations $Y$ is an additive sum of the discrete Radon transformations $H\mu$ and $M$ of the environment and anomaly, respectively. Hence, it is necessary to detect the anomaly 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 decision function $\Lambda(Y)$ of the test $\delta$ should be independent of its value. To obtain a statistics which is independent of the nuisance parameter, the theory of invariance can be used [1, 7].

4.2 Anomaly detection

We consider now the hypotheses testing problem (4), given the following measurement model:
\[
Y = H\mu + \xi (+ M).
\] (6)

First of all, let us note that the family of distributions $Y \sim N(M + H\mu, \sigma^2 I_n)$ remains invariant (see [7] for details and definitions) under the group of translations $G = \{g : g(Y) = Y + HC, \ C \in \mathbb{R}^m\}$, which in the parameter space $\mathbb{R}^n$ induces the group $\overline{G} = \{\overline{g} : \overline{g}(N) = N + HB, \ B \in \mathbb{R}^m\}$, $N \in \mathbb{R}^n$, that preserves both $\Omega_0 = \{E(Y) = H\mu, \ \mu \in \mathbb{R}^m\}$ and $\Omega_1 = \{E(Y) = H\mu + M, \ \mu \in \mathbb{R}^m, \ M \neq 0\}$, i.e. $\overline{g}\Omega_0 = \Omega_0$ and $\overline{g}\Omega_1 = \Omega_1$. Hence, the hypotheses testing problem (4) remains invariant under $G$ [7]. The optimal
Let us consider a 2D scene, without (Figure 3.(a)) and with (Figure 3.(d)) anomaly, numerically simulated with Matlab computing software. The environment \( h_u \) is composed of a polynomial function of degree 2 with unknown coefficients [10]. The anomaly consists in a small constant attenuation coefficient disk located at the upper left corner [11].

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 a parameter of non-centrality equal to \( c^2 = \frac{1}{\sigma^2}M^T P_H M \). Hence, the power function of \( \delta^* \) is directly given by:

\[
\beta_{\delta^*}(c^2) = \text{Pr}_{c^2}(\delta^* = \mathcal{H}_1) = \int_{\lambda_\alpha}^{+\infty} \varphi_{c^2}(x)dx,
\]

where \( \varphi_{c^2} \) is the probability density function (pdf) of the non-central \( \chi^2 \) with \( n - q \) degrees of freedom and the non-centrality parameter \( c^2 \).

5 Numerical example

Let us consider a 2D scene, without (Figure 3.(a)) and with (Figure 3.(d)) anomaly, numerically simulated with Matlab computing software. The environment \( h_u \) is composed of a polynomial function of degree 2 with unknown coefficients [10]. The anomaly consists in a small constant attenuation coefficient disk located at the upper left corner [11].

The attenuation coefficient of the disk does not dominate over the scene but we assume that a local contrast \( f_{0,u} \neq 0 \) exists between the disk and the environment. The considered tomographic projections \( Y \) are taken at the angle of view \( \eta = 75^\circ \) (the linear X-ray detector corresponds to the black straight line plotted on Figure 3.(a) and Figure 3.(d)), of the scene without (Figure 3.(b)) and with the anomaly (Figure 3.(e)). It is worth noting that the mean value of projections has been centered for increasing the visual contrast but, nevertheless, the attenuation peak due to the anomaly is nearly invisible. To reject the nuisance parameters (environment), the transformation of the observations \( Y \)
into the parity vector $Z = WY$ has been applied. Here, the pick corresponding to the anomaly is clearly visible (see Figures 3.(c) and 3.(f)). This visual result confirms the augmentation of the statistics $\Lambda(Y)$ of the test $\delta^\ast$ given by equation (8) under $\mathcal{H}_1$. Let us examine the statistical properties of the proposed test by using this numerical example. The probability of false alarm is fixed at 1% : $\Pr_0(\delta(Y) = \mathcal{H}_1) = 10^{-2}$. The variance $\sigma^2$ is assumed to be known and constant. First, let us compute the probability of non-detection $1 - \beta_\delta^\ast$ as a function of the anomaly location $(x_0, y_0) : (x_0, y_0) \mapsto 1 - \beta_\delta^\ast(x_0, y_0)$. It is assumed that the abscissa $x_0$ varies from $-10$ cm to $10$ cm and the ordinate $y_0$ also varies between $-10$ cm et $10$ cm. The numerical results are shown in Figure 4.(a) by using a logarithmic scale to attenuate their variations. A “valley” of good detection clearly appears on the surface. It corresponds to the locations $(x_0, y_0)$ where the environment is rejected with a great efficiency. In contrast with this “valley”, the edges of the surface $(x_0, y_0) \mapsto 1 - \beta_\delta^\ast(x_0, y_0)$ correspond to locations where the anomaly is well hidden by the environment.

To explain this phenomenon, let us consider the variations of the non-centrality parameter $\lambda = \frac{1}{\sigma^2}M^T\Pi_n M$ of the $X_{n-m,\lambda}$ distribution of the statistics $\Lambda(Y)$ under the hypothesis $\mathcal{H}_1$ (see Figure 4.(b)). The non-centrality parameter can be rewritten as $\lambda = \frac{1}{\sigma^2}M^T(I_n - \tilde{H})M$ where $\tilde{H} = H(H^T H)^{-1}H^T$. Assuming that the anomaly $M$ differs from zero for only $k$ adjacent positions (the anomaly is “compact”), we can write $M = \sum_{i=p}^{p+k-1} m_i e_i$ where $M = (0, \ldots, 0, m_p, \ldots, m_{p+k-1}, 0, \ldots, 0)^T$, where $p$ is the subscript of the first non-zero component of $M$ and $(e_1, \ldots, e_n)$ is the canonic basis of the linear space $\mathbb{R}^n$. It follows that:

$$\lambda = \frac{1}{\sigma^2} \left( \sum_{i=p}^{p+k-1} m_i^2 - \sum_{i=p}^{p+k-1} \sum_{j=p}^{p+k-1} m_i m_j \tilde{h}_{i,j} \right) = A(M, \sigma^2) - E(M, \tilde{H}, \sigma^2, p), \quad (9)$$

where $A(M, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=p}^{p+k-1} m_i^2$, $E(M, \tilde{H}, \sigma^2, p) = \frac{1}{\sigma^2} \sum_{i=p}^{p+k-1} \sum_{j=p}^{p+k-1} m_i m_j \tilde{h}_{i,j}$ and $\tilde{h}_{i,j}$ is the element $ij$ of the matrix $\tilde{H}$. For a given angle of view $\eta$, the anomaly projection has a constant geometrical shape for all possible locations of the anomaly. Hence, the finite set $m_p, m_{p+1}, \ldots, m_{p+k-1}$ is independent of the first subscript $p$. In other words, the positive term $A(M, \sigma^2)$ represents the constant anomaly impact.

Figure 3: (a) Original scene without anomaly. (b) Projection of (a) at $75^\circ$. (c) Parity vector calculated from (b). (d) Original scene with anomaly. (e) Projection of (d) at $75^\circ$. (f) Parity vector calculated from (e).
6 Conclusion

The problem of anomaly/target detection in a linearly parameterized environment has been stated as a composite hypotheses testing problem with nuisance parameters. An optimal invariant decision rule has been proposed in the case of anomaly detection. The detection of an a priori specified target is much more difficult because of the non-linear measurement model. In the future works, a very special attention will be paid to the target detection problem in order to propose an appropriate decision rule with some warranted optimality properties.
Figure 5: The probability of non-detection as a function of the radius $R$ and the contrast $d$ characterizing the anomaly.

References


