MULTISCALE HYPOTHESIS TESTING WITH APPLICATION TO ANOMALY CHARACTERIZATION FROM TOMOGRAPHIC PROJECTIONS

Austin B. Frakt, Alan S. Willsky

Massachusetts Institute of Technology
77 Massachusetts Avenue
Room 35-425
Cambridge, MA 02139
E-mail: frakt@mit.edu

W. Clem Karl

Department of Electrical,
Computer and
Systems Engineering
Boston University
Boston, MA

ABSTRACT

Anomaly characterization from tomographic measurements is of interest in a wide range of fields. In this paper we address the problems of single anomaly detection and localization which we formulate as hypothesis testing problems. While the optimal hypothesis test is easy to formulate, it is computationally infeasible due to the overwhelming number of hypotheses which must be considered. We propose the *multiscale hypothesis test (MSHT)* as an efficient suboptimal alternative. We show how to find decision statistics to achieve maximal composite hypothesis distinguishability for the composite hypothesis tests which comprise the MSHT.

1. INTRODUCTION

In many applied problems tomographic data are used to study properties of the interior of objects. Such problems arise in a wide range of fields including medical imaging, geophysical remote sensing, radio astronomy, non-destructive testing, etc. [1]. In many cases, the ultimate goal is *not* to reconstruct a detailed image of the region under investigation but is to characterize (e.g., detect, locate) regions which are, in some sense, anomalous. Often only limited or noisy data are available to accomplish this objective.

Many approaches to the anomaly characterization problem begin with image reconstruction. Anomalous regions are then analyzed by post-processing this image. Reconstruction is not necessarily the best way to attack the problem, however, since it may not make optimal use of the data. In fact, it can make the problem harder. For example, in limited data or high noise cases, reconstruction introduces streaking artifacts which are themselves anomalous. Finally, reconstruction is a computationally non-trivial task and, in light of the above, is computationally wasteful when the ultimate goal is the far more modest one of anomaly characterization.

A major challenge, therefore, is to develop methods to characterize anomalies directly from tomographic data. In this paper we present methods for detecting and locating a single anomaly without image reconstruction. We view these problems as ones of change detection. The possible changes (anomalies) relative to a statistically known background form a finite set of hypotheses: $\mathcal{H} \triangleq \{H_i\}_{i=0}^{M-1}$. The anomaly detection and localization problems, therefore, are hypothesis testing problems on the set $\mathcal{H}$.

While the form of the optimal M-ary hypothesis test is clear, even for simple classes of anomalies it requires far too many hypotheses for reasonable execution. Our alternative is to conduct a sequence of small composite hypothesis tests. This sequence can be viewed as a scale recursion in hypothesis space. At coarser scales, relatively large numbers of hypotheses are discarded so that at finer scales attention is focussed on relatively small subsets of the global set of hypotheses. Each composite hypothesis test in the sequence is associated with subsets of $\mathcal{H}$ (composite hypotheses) and decision statistics. These subsets and statistics are designed so that the resulting sequence of tests zooms in on the correct hypothesis via a multiscale search over the set $\mathcal{H}$. We call this multiscale search a *multiscale hypothesis test (MSHT)*.

This work was supported by the Advanced Research Project Agency (ARPA) through the Air Force Office of Scientific Research (grant F49620-93-1-0604), the Army Research Office (grant DAAL03-92-G-0115), and the Office of Naval Research (grant N00014-91-J-1004). A.B. Frakt is supported by a National Defense Science and Engineering Graduate Fellowship awarded by ARPA. We also thank Dr. Hamid Krim and Dr. Paul Fieguth for their support in the development of this work.

This approach is similar to the one introduced in [2] for geophysical inverse problems based on scattered radiation measurements. The structure of the composite hypothesis tests employed in [2] is motivated by a desire to effect spatial zooming. The sequence of hypotheses corresponds to a nested sequence of regions in the image domain and at each stage of the search a finer scale region is considered. The form of the statistics employed in each composite hypothesis test has a natural matched filter interpretation. However, while intuitively natural, these statistics are in no sense optimal are not the best statistics for all problems.

We extend the work in [2] by taking a broader perspective of multiscale hypothesis testing. We recognize that the efficacy of a MSHT relies upon the ability to distinguish between composite hypotheses. Achieving such distinguishability is most difficult at the coarsest scale of a MSHT. Therefore, the choice of decision statistics is of crucial importance. We shall introduce a method to find statistics which achieve maximal composite hypothesis distinguishability.

2. ELEMENTS OF MULTISCALE HYPOTHESIS TESTING

A multiscale hypothesis test consists of a sequence of composite hypothesis tests. Each test in the sequence consists of composite hypotheses which form a cover for some subset of $\mathcal{H}$. A decision statistic is associated with each composite hypothesis. Each test in the sequence is given an index which we call *scale*. At a given scale, one (or several) composite hypotheses are selected for finer scale investigation on the basis of these statistics (e.g., the one associated with the maximum statistic value is selected, others are discarded). The range of the covers (i.e., the subset of $\mathcal{H}$ which the composite hypotheses cover) becomes smaller in cardinality as the scale index increases. At the coarsest scale a cover is defined for all of $\mathcal{H}$. The finest scale consists of a cover for just a proper subset of $\mathcal{H}$. Some of the $H_i \in \mathcal{H}$ are not included in the composite hypotheses at an intermediate scale. Those which are not included are said to have been *discarded*. Any hypothesis, H_i, which has been discarded cannot ultimately be selected as the one which we think is true. The efficiency of a MSHT is achieved by discarding many H_i at each scale.

An example is illustrated in Figure 1. At each scale in the tree shown a choice is made between two composite hypotheses (written in italics) based on two statistics. The chosen composite hypothesis is indicated with an arc with an arrow. The superscripts on the composite hypotheses and statistics indicate the scale. Notice that the subset of $\mathcal{H}$ for which a cover is defined at

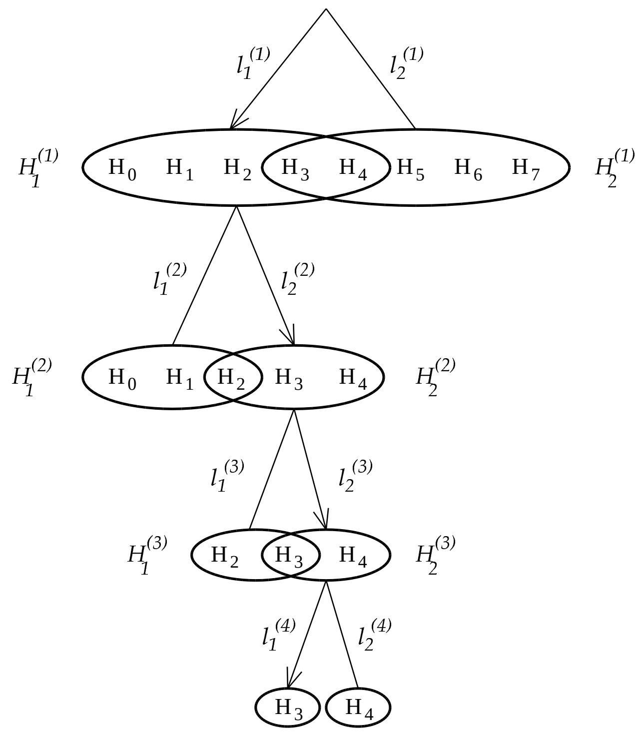

Figure 1: A multiscale hypothesis test.

scale k is precisely the subset which is contained in the composite hypothesis which has been selected at scale $k-1$. The elements of a MSHT discussed above and illustrated in Figure 1 apply equally to the case where an arbitrary number of composite hypotheses $N^{(k)}$ are defined at scale k.

The challenge in designing a MSHT is to select good covers and statistics for every scale and for every possible sequence of decisions. These covers and statistics should be selected so that the resulting sequence of composite hypothesis tests effectively zooms in on the true hypothesis via the multiscale search on the domain $\mathcal{H}$. We shall have more to say about choosing good statistics in Section 4.

3. MODELS AND ASSUMPTIONS

We assume that the $N \times N$ field, $\mathbf{f}$, for which we have tomographic data is the sum of a background field, $\mathbf{f}_b$, and an anomaly field, $\mathbf{f}_a$, where all fields are represented by vectors of rectangular pixel values lexicographically ordered. The background is a zero-mean Gaussian random field with known covariance $\mathbf{\Lambda}_b$. The anomaly field is zero everywhere except over an unknown square patch where the intensity is constant and non-negative:

$$\mathbf{f}_a = c\mathbf{b}_{s,N}(i,j)\,,$$

where c is the anomaly intensity and $\mathbf{b}_{s,N}(i,j)$ is the vector associated with the $N \times N$ field which is zero everywhere except over the $s \times s$ patch with upper left corner at pixel (i,j) where it is one. Each element of $\mathcal{H}$ is associated with one possible size and shift of a square, constant intensity anomaly. That is, each $H_k \in \mathcal{H}$ corresponds to exactly one vector $\mathbf{b}_k \stackrel{\triangle}{=} \mathbf{b}_{s_k,N}(i_k,j_k)$ where s_k is a member of some subset of sizes $\{1,2,3,\ldots,s_{max}\}$ and (i_k,j_k) takes on one of its possible allowable values for the given s_k. We assume knowledge of the maximum possible size, s_{max}, of the anomaly where $s_{max} \ll N$.

The projection of the field $\mathbf{f}$ is modeled by a linear transformation using a matrix, $\mathbf{T}$, which captures a discrete representation of the line-integral projections. Finally, zero-mean white Gaussian measurement noise, $\mathbf{n}$, with known intensity λ is added to the projection data. Our observation equation and associated models are:

$$\begin{aligned} \mathbf{g} &= \mathbf{T}(\mathbf{f}_a + \mathbf{f}_b) + \mathbf{n}\,, \\ \mathbf{f}_b &\sim \mathcal{N}(0, \mathbf{\Lambda})\,, \\ \mathbf{n} &\sim \mathcal{N}(0, \lambda \mathbf{I})\,. \end{aligned}$$

We also define the data covariance matrix

$$\mathrm{cov}(\mathbf{g}) \stackrel{\triangle}{=} \mathbf{\Lambda} = \mathbf{T}\mathbf{\Lambda}_b\mathbf{T}^T + \lambda \mathbf{I}\,.$$

4. MSHT METHODS FOR ANOMALY DETECTION AND LOCALIZATION

A multiscale hypothesis test possesses two main high level characteristics: the form of the covers (the composite hypotheses) and the form of the statistics. (Actually there is a third characteristic, the number of composite hypotheses selected at each scale. However, for brevity and clarity we ignore this degree of freedom in this paper and focus on the case in which only one is selected at each scale.) In this section we formulate a MSHT for the anomaly detection and localization problems. The form of the covers we use is motivated by the work of Miller and Willsky in [2] and are chosen to effect spatial zooming. In particular, the composite hypotheses are associated with spatially contiguous regions of the image domain. Having fixed the form of the composite hypotheses, we propose and solve an optimization problem for the statistics. This problem aims to provide statistics with maximal composite hypothesis distinguishability as discussed in Section 1. For clarity of presentation we shall restrict our discussion to the coarsest scale composite hypothesis test of a MSHT. The processing at subsequent scales is similar.

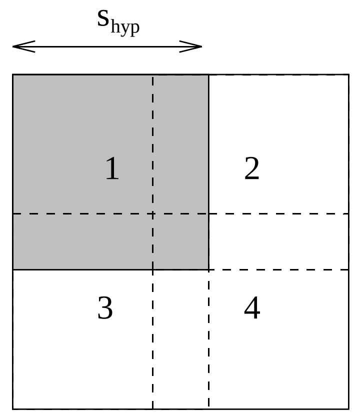

Figure 2: The composite hypotheses at the coarsest scale. The four composite hypotheses overlap so that the chosen one contains the entire anomaly.

4.1. Composite Hypotheses

Figure 2 provides an image domain interpretation of the composite hypotheses at the coarsest scale. There are four composite hypotheses: $\mathcal{H}_i$, for $i \in \{1,2,3,4\}$. Each one corresponds to a square $s_{hyp} \times s_{hyp}$ region of the image domain as shown. We associate each composite hypothesis with an indicator function $\mathbf{B}_i$ which is one over the $s_{hyp} \times s_{hyp}$ region corresponding to $\mathcal{H}_i$ and zero elsewhere. The hypothesis H_k belongs to composite hypothesis $\mathcal{H}_i$ if and only if $\mathbf{b}_k^T \mathbf{B}_i = s_k^2$. For example, composite hypothesis $\mathcal{H}_1$ corresponds to the shaded region in Figure 2. All and only hypotheses associated with anomalies with support *entirely* within this shaded region belong to $\mathcal{H}_1$. The composite hypothesis regions overlap by at least $s_{max} - 1$ pixels so that each possible anomaly lies entirely within at least one region.

The composite hypothesis test conducted at the coarsest scale selects one of $\mathcal{H}_i$, where $i \in \{0,1,2,3,4\}$. The composite hypotheses have the interpretation

$$\begin{aligned} \mathcal{H}_0 &: \text{ no anomaly}\,, \\ \mathcal{H}_i &: \text{ anomaly has support in region } i\,. \end{aligned}$$

One of these composite hypotheses is selected on the basis of a comparison of four statistics, ℓ_i, for $i \in \{1,2,3,4\}$. The composite hypothesis $\mathcal{H}_i$ is selected where

$$i = \begin{cases} 0 & , \text{ for } \max_j [\ell_j] < \eta \\ \arg\max_j [\ell_j] & , \text{ otherwise} \end{cases}\,.$$

We discuss the form of the statistics next.

4.2. Optimized Statistics

In this section we are concerned with finding a *linear* statistic $\ell_i = \mathbf{a}_i^T \mathbf{g}$ to associate with composite hypothesis $\mathcal{H}_i$ for each i. We wish this statistic to provide maximal composite hypothesis distinguishability. Toward this end, we formulate an optimization problem to choose the linear weight vector $\mathbf{a}_i$ so that ℓ_i is, on average, as large as possible when $\mathcal{H}_i$ is true (i.e., it contains the true hypothesis) and as small as possible otherwise. Before introducing the problem, we make the following definitions and observations. Define the conditional mean and variance of the statistic ℓ_i as

$$\begin{aligned} m_{ij} &\stackrel{\triangle}{=} E[\ell_i | H_j] = \mathbf{a}_i^T \mathbf{T} \mathbf{b}_j \,, \\ \sigma_i^2 &\stackrel{\triangle}{=} \mathrm{var}[\ell_i | H_j] = \mathbf{a}_i^T \mathbf{\Lambda} \mathbf{a}_i \,. \end{aligned}$$

Notice that the conditional mean is linear in $\mathbf{a}_i$ while the conditional variance is quadratic in $\mathbf{a}_i$. Also note that the conditional variance is independent of j.

The optimization problem we consider is

$$\hat{\mathbf{a}}_i = \arg\max_{\mathbf{a}} \min_{(j,k) \in \mathcal{A}_i} \frac{m_{ij} - m_{ik}}{\sigma_i} \,,$$

where $\mathcal{A}_i \stackrel{\triangle}{=} \{(j,k) | H_j \in \mathcal{H}_i \text{ and } H_k \notin \mathcal{H}_i\}$. Reading from right to left, we see that in this optimization problem we consider the difference between two standard-deviation-normalized conditional means. One element in the difference is m_{ij} which we want to be large since we constrain $H_j \in \mathcal{H}_i$. The other element in the difference is m_{ik} which we want to be small since we constrain $H_k \notin \mathcal{H}_i$. Therefore, we want the difference to be large. Taking the worst case difference (with the min), we maximize this with respect to $\mathbf{a}$. Plugging in definitions, the optimization problem is

$$\hat{\mathbf{a}}_i = \arg\max_{\mathbf{a}} \min_{(j,k) \in \mathcal{A}_i} \frac{\mathbf{a}^T \mathbf{T} \mathbf{b}_j - \mathbf{a}^T \mathbf{T} \mathbf{b}_k}{\sqrt{\mathbf{a}^T \mathbf{\Lambda} \mathbf{a}}} \,.$$

Using Lagrange duality theory it can be shown that this problem is equivalent to a quadratic program which is easily solved using off-the-shelf software (e.g., MATLAB's `qp()` function). One such quadratic program must be solved for each composite hypothesis at each scale. However, this processing can be done once off-line.

4.3. Detection and Localization Algorithm

The coarsest scale of the MSHT has already been specified. It is at this scale that the detection decision is made. If $\mathcal{H}_0$ is selected then it is determined that no anomaly exists and no further processing is done. If, on the other hand, $\mathcal{H}_0$ is not selected then the region corresponding to the composite hypothesis which has been selected is subdivided with four overlapping regions (analogous to the subdivisions at the coarsest scale). Since, in this case, it has already been determined that an anomaly exists, there is no null hypothesis at the second and higher scales. This decision-directed, scale-recursive search continues until a pre-specified scale has been reached (in practice we stop at the scale corresponding to subdivisions of size s_{max}). Alternatively a stopping criterion which depends on the statistic values may be employed.

5. CONCLUSION

In this paper we have introduced the MSHT, an efficient suboptimal alternative to a computationally infeasible M-ary hypothesis test. Consideration of the general structure of the MSHT has led us an optimization problem whose solution is a statistic which provides maximum composite hypothesis distinguishability. We have also shown how to apply the MSHT framework with optimized statistics to the single anomaly detection and localization problems from tomographic projections. Our other work in this area (which could not be provided here due to space limitations) includes the computation of performance bounds, the investigation of anomaly ambiguity, examples of the application of our methods to the tomography problem, and investigation of the computational complexity of our algorithm as compared to the optimal M-ary hypothesis test.

6. REFERENCES

[1] D.J. Rossi and A.S. Willsky. Reconstruction from projections based on detection and estimation of objects—parts I and II: Performance analysis and robustness analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-32(4):886–906, 1984.

[2] E. L. Miller and A. S. Willsky. Multiscale, statistical anomaly detection analysis and algorithms for linearized inverse scattering problems, July 1995. Submitted to *Multidimensional Signals and Systems.*