TOMOGRAPHIC FEATURE DETECTION AND CLASSIFICATION USING PARALLELOTOPE BOUNDED ERROR ESTIMATION

Alfred O. Hero III, Yong Zhang** and W. Leslie Rogers****

*Dept. of EECS, ** Dept. of Radiology and ***Division of Nuclear Medicine
The University of Michigan, Ann Arbor, MI 48109-2122

ABSTRACT

We give a novel method for performing statistically significant detection of specified object features which operates directly on X-ray (Gaussian) or radio-isotope (Poisson) tomographic projection data. The method is based on constructing an exact $(1-\alpha)100\%$ confidence region on the object derived by backprojecting a projection-domain confidence region into object space. The projection-domain confidence region is a minimal volume hyper-rectangle specified by the projection data and the appropriate quantiles of the standard Gaussian or Poisson distribution. We implement the backprojection step using a very accurate bounded error estimation algorithm which sequentially approximates the feasible set (object-domain confidence region) given the data and its specifed error bounds (known Gaussian or Poisson quantiles). By testing whether this object-domain $(1-\alpha)100\%$ confidence region contains objects with hypothesized features we obtain a feature detection algorithm which has constant false alarm rate (CFAR) α and is adaptive in the sense that no image reconstruction is required and no unknown nuisance parameters need be estimated.

1. BACKGROUND

In a 1992 paper [1] we applied bounded error estimation, also known as set theoretic estimation, to image reconstruction from projections and in a followup 1993 paper [2] the resultant set estimates were used to specify a multi-dimensional statistical confidence region on the true object. In particular we showed that when the error bounds were suitably chosen as quantiles of the projection noise distributions the Ellipsoid Parallel Cuts (EPC) algorithm [3] could be used to specify a set estimate of the object which corresponds to a region that is guaranteed to contain the true object with probability at least $1-\alpha$. Consistent with statistical terminology, we called this set estimate a level $(1-\alpha)100\%$ confidence region for the object. In a subsequent paper [4] we used this methodology to study confidence levels and statistical feasibility of popular image reconstructions such as filtered back projection, weighted-least-squares, and iterative maximum likelihood (EM algorithm) for emission computed tomography. Recently, Combettes published a paper which nicely lays out the equivalence between set theoretic estimation and statistical confidence regions in a general context [5].

The EPC algorithm is a bounded error estimation algorithm which finds a minimal volume ellipsoid containing the set of images which are consistent with the error bounds, called the *feasible set*. EPC is usually applied to measurements of a system output to find a (point) estimate of the input of the system which is robust to bounded error perturbations of the system matrix (mismodeling error) or the system output (additive measurement noise). Various point estimation strategies have been proposed which use the centroid or some other point within the final ellipsoid estimate of the feasible set [6]. When a feasible set estimator such as EPC is used to obtain confidence regions it is important that the size of the estimated set have coverage probability as close to $1-\alpha$ as possible to ensure a high precision, i.e. small volume, confidence region. However, we have found that the EPC algorithm may yield feasible sets whose true coverage probability greatly exceeds $1-\alpha$. This finding is consistent with remarks made by other investigators that the EPC algorithm tends to significantly oversestimate the size of the feasible set.

Recently a new feasible set algorithm was introduced in the context of robust adaptive control which finds a minimal volume paralellotope containing the feasible set [7]. This remarkable algorithm, which we call parallelotope parallel cuts (PPC), is faster than EPC, appears to be less sensitive to roundoff errors, and converges to the exact feasible set in a finite number of iterations for square system matrices. In the present context this means that for square matrices PPC can be applied to obtain confidence regions which have coverage probability which is exactly equal to $1-\alpha$

THIS RESEARCH WAS SUPPORTED IN PART BY THE NATIONAL CANCER INSTITUTE UNDER GRANT R01-CA-54362-02

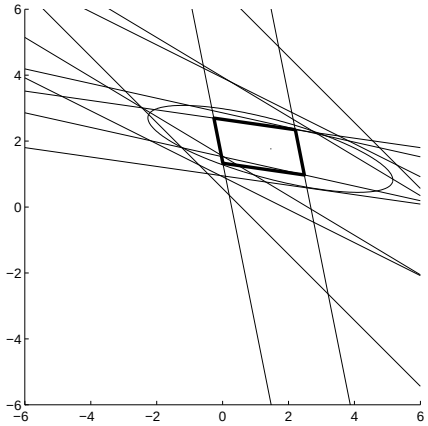

Figure 1: Feasible set, sequence of parallel cuts (PPC), final ellipsoid (EPC) and parallelotope (PPC) (dark outline). Note how PPC only slightly over-estimates the feasible set.

and have minimal volume. As a result we can use the well known duality between minimal volume confidence regions and rejection regions of optimal binary hypothesis tests [8] to specify an optimal detector. Specifically, by testing whether the minimal volume $(1-\alpha)100\%$ confidence region contains objects with hypothesized features we obtain a feature detection algorithm which has constant false alarm rate (CFAR) α and is adaptive in the sense that no reconstruction is required and no unknown nuisance parameters need be estimated.

Simulation results for a parallel ray projection geometry in emission and transmission tomography applications will be given in the full paper to illustrate the methodology.

2. BOUNDED ERROR ESTIMATION

Assume a nominal linear measurement model: $\overline{Y}_k = \underline{\phi}^T(k)\underline{\theta}$, $k = 1, ..., N$, where $\overline{Y}_k$ is the model output, $\underline{\phi}^T(k)$ is a vector specific to the system, e.g. the k-th row of the tomographic system matrix A, $\underline{\theta} \in \mathbb{R}^p$ is the parameter vector to be estimated, e.g. the object intensity $\underline{\lambda}$, and k is the measurement index. If it is known that the error $e(k) = Y_k - \overline{Y}(k)$ is bounded within $[e_{min}(k), e_{max}(k)]$:

$$e_{min}(k) \leq Y_k - \underline{\phi}^T(k)\underline{\theta} \leq e_{max}(k), \qquad k = 1, ..., N, \quad (1)$$

then the set of all values of $\underline{\theta}$ consistent with (1) is given by the polytope $\Lambda = \cap_{k=1}^N H_k$, called the feasible set, which is the intersection of the hyperslabs: $H_k = \{\underline{\theta} : Y_k - e_{max}(k) \leq \underline{\phi}^T(k)\underline{\theta} \leq Y_k - e_{min}(k)\}$, $k = 1, ..., N$.

The Ellipsoid Parallel Cuts (EPC) and the Parallelotope Parallel Cuts (PPC) algorithms can be put in the common framework of finding a sequence of successively smaller sets $\{E_k\}_{k=1}^N$ all of which contain the feasible set Λ and such that E_N is set of minimal volume:

$$E_k = E(\underline{\theta}_k, C_k) = \{\underline{\theta} : \|C_k(\underline{\theta} - \underline{\theta}_k)\| \leq 1\} \qquad (2)$$

where $\underline{\theta}_k$ is a vector governing the center of E_k, C_k is a non-singular matrix governing the shape of E_k, and $\|\cdot\|$ is a norm. For EPC $\|\cdot\|$ corresponds to the l_2 norm ($\|\underline{z}\| = \sqrt{\underline{z}^T\underline{z}}$), the E_k's are ellipsoids, $\underline{\theta}_k$ is the centroid of E_k, and C_k is a matrix governing the principal and minor axes of E_k. For PPC $\|\cdot\|$ corresponds to the l_∞ norm ($\|\underline{z}\| = \max|z_i|$), the E_k's are parallelopipeds, $\underline{\theta}_k$ is the geometric center of E_k, and C_k determines the length and orientation of the edges of E_k. In [3] and [7], respectively, the EPC and PPC algorithms are introduced and various theoretical properties are established. In particular, they establish that as long as $N \geq p$, the EPC and PPC algorithms are guaranteed to generate sequences E_k, $k = 1, \ldots, N$ which after N steps yields the minimal volume ellipsoid (EPC) or parallelopiped (PPC) containing Λ.

As reported in [7] the PPC algorithm has similar computational complexity as EPC yet consistently converges to a smaller feasible set estimate than EPC. Furthermore, unlike EPC, the PPC algorithm converges to the exact polyhedral feasible set Λ when the system matrix A is square and invertible. Our simulations show that in fact the Parallotope algorithm requires only about 80% of the runtime per iteration as compared to the EPC algorithm when implemented on a HP735.

3. CONFIDENCE REGIONS AND DETECTION

Let $\underline{Y}$ be random data which has probability distribution $P_{\underline{\theta}}(\underline{Y})$ which depends on an unknown parameter vector (an image) $\underline{\theta} \in \Theta$. A $(1-\alpha)100\%$ confidence region for $\underline{\theta}$ is defined as a subset $\Lambda_{1-\alpha}(\underline{Y})$ of Θ such that $P(\underline{\theta} \in \Lambda_{1-\alpha}) \geq 1-\alpha$, which is to be read "the probability that the random subset $\Lambda_{1-\alpha}(\underline{Y})$ contains the non-random parameter $\underline{\theta}$. Now assume that we want to test at some level of false alarm α whether the true image $\underline{\theta}$ has some simple feature, e.g. $\underline{\theta} = \underline{\theta}^o$ for some test image $\underline{\theta}^o$. Specifically, we need to specify a test of

$$\begin{aligned} H_0 &: \underline{\theta} = \underline{\theta}^o \\ H_1 &: \underline{\theta} \neq \underline{\theta}^o \end{aligned}$$

subject to the false alarm constraint

$$P_{\underline{\theta}^o}(\text{decide } H_1) \leq \alpha.$$

It is obvious that the $1-\alpha$ confidence region $\Lambda_{1-\alpha}$ specifies such a test: *if* $\underline{\theta}^o \in \Lambda_{1-\alpha}$ *then decide* H_0, *otherwise decide* H_1. More generally, if we have a composite feature that we want to test, e.g. specified by some set

of images $\mathcal{S}$, a level α test is specified by the following decision rule: *decide the true image has the composite feature when* $\Lambda_{1-\alpha} \supset \mathcal{S}$. For more discussion of the duality between tests and confidence regions the reader is referred to [8].

When $\underline{e}_{max}$ and $\underline{e}_{min}$ are selected to correspond to specific quantiles of the projection noise distribution the feasible set Λ can be identified as a $(1-\alpha)100\%$ confidence region for θ. For X-ray computed tomography (X-ray CT) the projection data $Y_1, \ldots, Y_N$, are distributed as independent Gaussian random variables with variances $\sigma_1^2, \ldots \sigma_N^2$ which do not depend on the object. When the variance is known the measurement variables $(Y_i - A\underline{\lambda})/\sigma_i$ can be used to give the standard $(1-\beta)100\%$ confidence rectangle on $A\underline{\lambda}$:

$$\begin{aligned} &\mathrm{X}_{k=1}^{N}[L_{min}(k), L_{max}(k)] = \\ &\quad \mathrm{X}_{k=1}^{N}\left[Y_k - \sigma_k Z_{1-\beta/2}, Y_k + \sigma_k Z_{1-\beta/2}\right] \end{aligned} \quad (3)$$

where Z_ξ is the ξth percentile of the standard Gaussian distribution. When the variance is unknown the studentized $(1-\beta)100\%$ confidence rectangle can be applied. This rectangle is identical to (3) except that σ_k is replaced by the square root ofthe sample variance $\hat{\sigma}_k$ and $Z_{1-\beta/2}$ is replaced by the $T_{1-\beta/2}$ where T_ξ is the ξth percentile of the student-t distribution. One can also use a studentized confidence interval by replacing σ_k with some asymptotically independent estimate of standard deviation.

On the other hand, for emission computed tomography (ECT), the projection data are distributed as independent Poisson random variables with rates $E[\underline{Y}] = A\underline{\lambda}$. It can be shown [9] that for a Poisson variable Y_k, a $(1-\beta)\times 100\%$ confidence rectangle for the rate $E[Y_k]$ is:

$$\begin{aligned} &\mathrm{X}_{k=1}^{N}[L_{min}(k), L_{max}(k)] = \\ &\quad \mathrm{X}_{k=1}^{N}\left[\frac{1}{2}\chi^2_{\beta/2}(2Y_k), \frac{1}{2}\chi^2_{1-\beta/2}(2[Y_k+1])\right] \end{aligned} \quad (4)$$

where $\chi^2_\xi(\nu)$ is the ξth percentile of the chi-square distribution with ν degrees of freedom.

The confidence rectangle (4) can be calculated recursively using known relations between Gamma(k) and Gamma(k+1) distributions. However this is very time-consuming and for our simulations we implement approximate confidence regions based on the normalizing square root transformations of Poisson random variables used in [1]. There and here we use the fact that $2(\sqrt{\underline{y}} - \sqrt{A\underline{\lambda}})$ is approximately distributed as a vector of i.i.d. $\mathcal{N}(0,1)$ (standard normal) random variables. Then, after applying a square root transformation to the Gaussian distribution, an approximate $1-\beta$ confidence rectangle for $A\underline{\lambda}$ can be obtained by using (3). It has been our experience that for the typical count rates encountered in ECT there is little perceptible difference in size and coverage probability between the simple approximate Gaussian confidence interval and the Chi-square based interval (4).

By setting $\beta = 1-(1-\alpha)^{\frac{1}{N}}$ in either (3) or (4) a $(1-\alpha)100\%$ confidence region for $\underline{\lambda}$ is obtained:

$$\begin{aligned} \Lambda_{1-\alpha} &= \{\underline{\lambda} : \underline{L}_{min} \le A\underline{\lambda} \le \underline{L}_{max}\} \\ &= \{\underline{\lambda} : \underline{e}_{min} \le \underline{Y} - A\underline{\lambda} \le \underline{e}_{max}\} \end{aligned} \quad (5)$$

where $\underline{e}_{min} \stackrel{\text{def}}{=} \underline{Y} - \underline{L}_{max}$, $\underline{e}_{max} \stackrel{\text{def}}{=} \underline{Y} - \underline{L}_{min}$. The $(1-\alpha)\%$ confidence region (5) is in the form of a bounded error (1) to which EPC or PPC can be directly applied. The resulting final set E_N will be a minimum volume ellipsoid (EPC) or polytope (PPC) E_N which contains $\Lambda_{1-\alpha}$ and is therefore a valid $(1-\alpha)100\%$ region.

In order that the PPC or EPC algorithms give a sequence of sets which converge to a bounding set on the feasible region $\Lambda_{1-\alpha}$ it is required that $\Lambda_{1-\alpha}$ be a subset of the initial PPC/EPC region E_1. In the simulations reported below we use an analytical bound on maximum parallelotope facet length to select such a region E_1. With $[L_{min}(i), L_{max}(i)]$ defined as the $[1-(1-\alpha)^{1/N})]\times 100\%$ confidence interval on the i-th element of the vector $A\underline{\lambda}$ define

$$l_o = |||[A^TA]^{-1}A^T|||_\infty \left(\max_i\{L_{max}(i) - L_{min}(i)\}\right).$$

Using well known relations between matrix and vector norms it can be shown that the hypercube of radius r_o centered at

$$[A^TA]^{-1}A^T\bar{\underline{Y}},$$

where

$$\bar{\underline{Y}} \stackrel{\text{def}}{=} \left[\frac{L_{max}(1)+L_{min}(1)}{2}, \ldots, \frac{L_{max}(N)+L_{min}(N)}{2}\right]$$

covers the feasible polytope $\Lambda_{1-\alpha}$. This hypercube was used to initialize the PPC algorithm in the simulations of this paper.

4. NUMERICAL COMPARISONS

First we compare the performance of the EPC and PPC algorithms for a very simple two dimensional problem with additive Gaussian noise. In this problem $\underline{\lambda} = [\lambda_1, \lambda_2]^T$ and the system matrix A is 6×2. Figure 1 shows the actual feasible set Λ, the sequence of PPC parallel cuts and the final ellipsoid and final parallelotope, respectively, for the same Gaussian noise realization. Note that the final ellipsoid is many times larger than the final parallelotope which closely approximates the actual feasible set. In tables 1 and 2 we compare the actual coverage probabilities of the final ellipsoid and final parallelotope, estimated empirically from 1000 trials, vs the prespecified confidence level $1-\alpha$ for 2×2

Confidence level	P_{EPC}	P_{PPC}
0.5	0.792	0.496
0.6	0.860	0.617
0.7	0.928	0.678
0.8	0.971	0.797
0.9	0.997	0.907
0.95	0.998	0.947

Table 1: H_0 coverage probabilities of the final ellipsoid and parallelotope containing for a 2×2 system. Observe that final parallelotope essentially meets the prescribed confidence level and is therefore of much smaller volume than the final ellipsoid satisfying the same confidence level constraint.

Confidence level	P_{EPC}	P_{PPC}
0.5	0.911	0.617
0.6	0.941	0.729
0.7	0.983	0.792
0.8	0.990	0.879
0.9	0.998	0.941
0.95	0.999	0.978

Table 2: H_0 coverage probabilities of the final ellipsoid and parallelotope confidence regions for a 6×2 system.

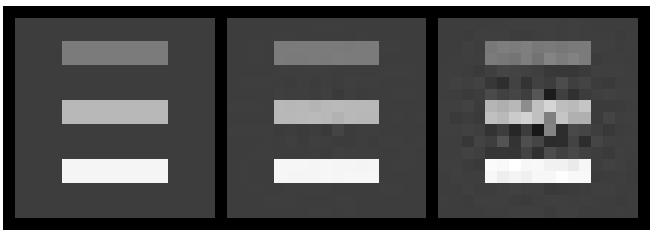

Figure 2: 17×17 phantom image, PPC centroid for high count regime (total mean number of counts $= 3.8 \times 10^6$), and PPC centroid for low count regime (total mean number of counts $= 3.8 \times 10^5$). Initial hypercubecube facet length $= 4.1635 \times 10^{10}$; confidence level $= 80\%$. Object was tomographically sampled using Anger camera geometry with 17 parallel collimated detector bins over 45 detector angles in $[0^o, 180^o]$

and 6×2 A matrices, respectively. Note that only the PPC algorithm gives a final set estimate E_6 having coverage probability close to the specified value $1 - \alpha$ in each of these cases.

We next turn to the tomographic imaging problem for object detection and classification using a small parallel ray Anger tomograph. The phantom is the mean intensity $A\underline{\lambda}$ of the spatial Poisson emission process shown in the first panel of Fig. 2. The two other panels show the true phantom and the centroids of the final parallelopipeds of PPC at two different phantom intensity levels for a single realization of the Poisson noise. Note that for high counts the 80% PPC centroid is very close to the phantom, indicating that the confidence region is centered near the object. For this case the centroid gives a good reconstruction. On the other hand, for lower counts the PPC centroid becomes a much more noisy facsimile of the object; the geometric center of the confidence region is further away from the object. We also found that the centroid is a poor estimator when the system matrix A becomes close to ill conditioned.

5. REFERENCES

[1] Y. Zhang, A. O. Hero, and W. L. Rogers, "A bounded error estimation approach to image reconstruction," in *Proc. of IEEE Nuclear Science Symposium*, pp. 966–968, Orlando, FA, Oct. 1992.

[2] A. O. Hero, Y. Zhang, and W. L. Rogers, "Consistency set estimation for PET reconstruction," in *Proc. of Conference on Information Science and Systems*, Johns Hopkins Univ., MD, March. 1993.

[3] E. Fogel and Y. Huang, "On the value of information in system identification - bounded noise case," *Automatica*, vol. 18, pp. 229–238, 1982.

[4] Y. Zhang, A. O. Hero, and W. L. Rogers, "Simultaneous confidence intervals for image reconstruction problems," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc.*, pp. V.317–320, Adelaide, April 1994.

[5] P. L. Combettes and T. J. Chaussalet, "Combining statistical information in set theoretic estimation," *Signal Processing Letters*, vol. 3, pp. 61–63, March, 1996.

[6] E. Walter and P. Lahanier, "Estimation of parameter bounds from bounded-error data - a survey," *Math. and Comp. Simulat.*, vol. 32, pp. 449–468, 1990.

[7] Vicino and Zappa, "Sequential approximation of feasible parameters for identifiaction with set membership uncertainty," *IEEE Trans. Automatic Control*, vol. 41, pp. 774–786, June 1996.

[8] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, San Francisco, 1977.

[9] E. B. Manoukian, *Modern Concepts and Theorems of Mathematical Statistics*, Springer-Verlag, New York N.Y., 1986.