

**SET THEORETIC ESTIMATION BY RANDOM SEARCH**

**Patrick L. Combettes**

**July 1989**

**CCSP Report: CCSP-TR-89/11**

## I. INTRODUCTION

In set theoretic estimation, consistency with all the *a priori* knowledge serves as an estimation criterion. Each piece of *a priori* knowledge is represented by a property set  $S_i$  in the solution space. Thus, given a collection  $\{S_1, \dots, S_m\}$  of property sets, the estimation problem is

$$\text{Find } y \in S = \bigcap_{i=1}^m S_i \quad (1)$$

$S$  is called the solution set or the feasible set. It is the class of objects which do not violate those constraints about the problem which are known *a priori*. Although it has met with significant success in several digital signal processing applications (e.g. filter design [1], signal restoration [7],[8], tomographic reconstruction [6]), set theoretic estimation still faces important questions, the most critical being to actually solve (1). Until recently, the only theoretically sound technique available was Brègman's method of successive projections (MOSP) for closed and convex sets in Hilbert spaces [2]. It was used in the aforementioned studies. The MOSP clearly imposes severe restrictions on the property sets and the underlying solution space (e.g. a piece of *a priori* knowledge may not be representable by a convex set in a hilbertian space). It was generalized to a much wider class of sets in arbitrary metric spaces in [3]. Even in that general framework, the MOSP still bears inherent theoretical and computational limitations, leaving the synthesis of set theoretic estimates an open question in many cases. Primarily, carrying out the projections onto certain property sets may amount to solving a very delicate global minimization problem. As a result, in a practical problem, the user may elect to give up a valuable piece of *a priori* knowledge because of the complexity associated with the analytical derivation and the numerical implementation of the projection onto the corresponding set.

In light of these elements, it becomes apparent that set theoretic estimation cannot rely exclusively on the MOSP for the synthesis of solutions. Therefore, to broaden its scope, it is necessary to introduce an alternative method. In sciences, it is common to resort to random strategies when deterministic methods are unsuccessful. This approach will be adopted here and a random search method will be employed to generate set theoretic estimates.

## II. THE METHOD OF RANDOM SEARCH (MORS)

In this report,  $k$  real parameters are to be estimated and, thus, the underlying solution space is  $\mathbb{R}^k$ . All of the sets to be considered are Borel subsets of  $\mathbb{R}^k$  and  $\lambda$  denotes the Lebesgue measure in  $\mathbb{R}^k$ .  $\Gamma = \{S_1, \dots, S_m\}$  is a collection of property sets with intersection  $S$  such that  $\lambda(S) > 0$ ,  $S_1$  is the set of smallest  $\lambda$ -measure, and  $\lambda(S_1) < +\infty$ .

A search region for  $\Gamma$  a set  $G$  such that  $\lambda(G) < +\infty$  and  $\lambda(S \cap G) > 0$ . A set theoretic estimate is generated by sequentially drawing points independently at random from a uniform distribution over  $G$  until a point in  $S$  is found. Henceforth, this method will be called the method of random search (MORS). The algorithmic description of the MORS is as follows.

1. Generate  $x$  from a uniform distribution over  $G$  and set  $i=1$ .
2. If  $x \notin S_i$ , return to 1.
3. If  $i \neq m$ , set  $i=i+1$  and return to 2.
4. Stop.

The efficiency of the MORS can be described by its acceptance rate, i.e. the probability  $\gamma$  that a point drawn from a uniform distribution over  $G$  is in  $S$ . Trivially

$$\gamma = \frac{\lambda(S \cap G)}{\lambda(G)} \quad (2)$$

Under our assumptions,  $\gamma > 0$  and, therefore, the MORS produces a solution with probability one. In practice, we are interested in the number of trials  $\eta$  required to obtain the first success. Clearly,  $\eta$  has a geometric distribution and its expected value is  $\bar{\eta} = 1/\gamma$ .

## III. DETERMINATION OF THE SEARCH REGION

The search region  $G$  for  $\Gamma$  should be determined so as to minimize the search time, an objective which involves two main factors. First, the geometry of  $G$  should be simple so that performing a single trial can be achieved at low computational cost. Indeed, efficient methods for generating points uniformly over a subset  $G$  of the Cartesian space are limited to cases where  $G$  is a simple region (hyperrectangle, hyperparallelogram, simplex, hypersphere, hyperellipsoid) and are described in [5]. More complex regions can be considered if  $G$  can be expressed as the union of disjoint simple regions

$\{G_1, \dots, G_n\}$ . Then, if  $f_i$  denotes the uniform density over  $G_i$ , the uniform density over  $G$  is given by

$$f = \sum_{i=1}^n \frac{\lambda(G_i)}{\lambda(G)} f_i \quad (3)$$

Second,  $G$  should be chosen so that the acceptance rate  $\gamma$  is large, which would in turn guarantee that the number of trials required to find a solution is small. The difficulty of this problem is that  $S$  is unknown. If  $G$  is determined too conservatively,  $\gamma$  will be low; if  $G$  is too small, it may not intersect with  $S$ . Since  $S$  is the intersection of all the  $S_i$ 's, it is contained in each of them. Hence, the best safe choice for the search region is the  $S_i$  of smallest  $\lambda$ -measure, i.e.  $S_1$ .

Thus, in terms of acceptance rate, the best choice for  $G$  is  $S_1$  whereas efficient uniform generation of points requires that  $G$  have a simple geometry. The following guidelines can be used to reach a compromise between these two conditions. If  $S_1$  is a simple region, then, clearly, it is the best candidate for a search region. In general, each  $S_i$  in  $\Gamma$  should be inserted into a set  $R_i$  determined as follows. If  $S_i$  is convex,  $R_i$  should be the smallest simple region which covers it. If not, it will be inefficiently covered by a single simple region. To achieve a better covering,  $R_i$  should be a disjoint union of simple regions. Then,  $G$  is chosen as the  $R_i$  of smallest  $\lambda$ -measure. To keep the cost of the uniform generation as low as possible, in choosing the  $R_i$ 's, hyperrectangles should be considered first. If they do not provide a good boxing of the sets, hyperparallelograms should be considered next, then simplexes, and finally hyperspheres and ellipsoids. In selecting  $G$ , the main objective is the minimization of the overall search time. Thus, for example, a simplex would be preferable to a hyperrectangle only if the improved acceptance rate offset the increase in the cost of the uniform generation.

#### IV. DISCUSSION

The MORS is a simple and easily implementable method for generating set theoretic estimates in  $\mathbb{R}^k$ . Most of all, it truly provides total flexibility with regard to the incorporation of the *a priori* knowledge since no topological or geometrical restrictions are placed on the property sets. Moreover, in the MORS, using a property set  $S_i$  reduces to testing points for membership in  $S_i$ , which is a much easier task than projecting onto  $S_i$ , as is done in the MOSP. Computationally, another plus for the

MORS is that it lends itself to parallel implementation. Indeed, all the trials of the MORS can be made simultaneously since the outcome of a trial is independent from the preceding ones.

From the above exposition, it may seem that the MORS is an ideal method for synthesizing set theoretic estimates. Unfortunately, it is limited to problems in which it is feasible to approximate the solution set  $S$  by a region  $G$  over which uniform generation of points can be performed efficiently and such that a good acceptance rate is achievable. Without the latter restriction, the MORS would eventually generate a solution, but it might require a prohibitive number of trials. As a general rule, it becomes increasingly difficult to achieve a high acceptance rate as the dimension  $k$  of the problem increases. This "curse of dimensionality" is a well known problem of any search procedure and certainly precludes the use of the MORS in problems of large dimension such as image processing. Rather, the MORS should be considered for applications in which the number of unknown parameters is typically low (e.g. system identification, parametric spectral estimation, filter design).

## V. APPLICATION TO HARMONIC RETRIEVAL

In this section, we shall demonstrate an application of the MORS to the problem of estimating the frequencies of two real sinusoids in additive noise. The data are given by

$$x_i = h_3 \sin(2\pi h_1 i) + h_4 \sin(2\pi h_2 i) + w_i \quad 1 \leq i \leq n \quad (4)$$

The true frequencies are  $h_1=0.10$  and  $h_2=0.14$ , and the true amplitudes are  $h_3=h_4=1.0$ . Moreover,  $n=16$  data samples are available and the noise process is zero mean white and Gaussian with power  $\sigma^2=0.05$  yielding a SNR of 10 dB on each sinusoid. It is noted that the frequency spacing is well beyond the resolution  $1/n$  of DFT methods. The quadruple  $h=(h_1, h_2, h_3, h_4)$  constitutes the object to be estimated. Three easily accepted pieces of *a priori* knowledge will be assumed.

- (i) The data were obtained by sampling at least at twice the Nyquist rate.
- (ii) The power of one sinusoid is not less than ten times that of the other.
- (iii) The noise is zero mean white and Gaussian with power  $\sigma^2=0.05$ .

From (i), the true frequencies  $h_1$  and  $h_2$  lie in  $[0,0.25]$ . The power of the data process is

$p_x = \sigma^2 + (h_3^2 + h_4^2)/2$ . Hence, we obtain from (ii) that  $B \leq h_3, h_4 \leq \sqrt{10B}$  where  $B = \sqrt{2(p_x - \sigma^2)/11}$ . An estimate for  $p_x$  is  $\hat{p}_x = 1/n \sum_{i=1}^n x_i^2$ . Extensive simulation using 16-point data frames formed in accordance with (4) provides numerical evidence that the probability that  $|p_x - \hat{p}_x|/\hat{p}_x$  be greater than 0.30 is negligible. Hence, since our data record gives  $\hat{p}_x = 0.803$ ,  $p_x$  ranges from 0.56 to 1.05. Substituting these extreme values yields  $0.30 \leq h_3, h_4 \leq 1.35$ . Whence, the property set of constraints on  $h$  is

$$S_1 = [0, 0.25] \times [0, 0.25] \times [0.30, 1.35] \times [0.30, 1.35] \quad (5)$$

Given an estimate  $a = (a_1, a_2, a_3, a_4)$  of  $h$ , the residual samples are defined as

$$(\forall i \in \{1, \dots, n\}) \quad r_i = r_i(a) = x_i - a_3 \sin(2\pi a_1 i) - a_4 \sin(2\pi a_2 i) \quad (6)$$

If  $h$  was estimated with no error by  $a$ , the sample statistics of the residual should agree, within some confidence coefficient, with those known properties of the noise. Hence, (iii) can be used to construct the set  $S_2$  of estimates which give residual samples within some confidence interval [3]

$$S_2 = \bigcap_{i=1}^n \left\{ a \in \mathbb{R}^4 \mid |r_i| \leq \alpha \sigma \right\} \quad (7)$$

and the set  $S_3$  of estimates which give a residual periodogram consistent with the whiteness and the normality of the noise [3]

$$S_3 = \bigcap_{l=0}^{n/2} \left\{ a \in \mathbb{R}^4 \mid \left| \sum_{i=1}^n r_i \exp(-j \frac{2\pi}{n} li) \right|^2 \leq \xi_l \right\} \quad \text{where} \quad \xi_l = \begin{cases} n\sigma^2\beta & \text{if } l=0 \text{ or } n/2 \\ -n\sigma^2 \ln(\epsilon) & \text{if } 0 < l < n/2 \end{cases} \quad (8)$$

The confidence coefficient was fixed to  $1 - \epsilon = 0.95$ . Thus,  $\alpha = 1.96$  in (7) and  $\beta = 3.84$  in (8) from the tables of the normal and  $\chi_1^2$  distributions respectively.

Although  $S_1$  is not the set of smallest  $\lambda$ -measure, it will be chosen as a search region because, as a hyperrectangle, it allows uniform generation of points at minimum cost. In the presence of more *a priori* knowledge (e.g. the sinusoids have equal power, or are harmonically related), the search region could be significantly improved. To give a precise description of the performance of the MORS,  $10^8$  points were drawn from a uniform distribution over  $G = S_1$ . Among these, 3093 were found to be feasible. Thus, the expected number of trials required to obtain a solution is  $\bar{\eta} \approx 3.2 \times 10^4$ . Since we are

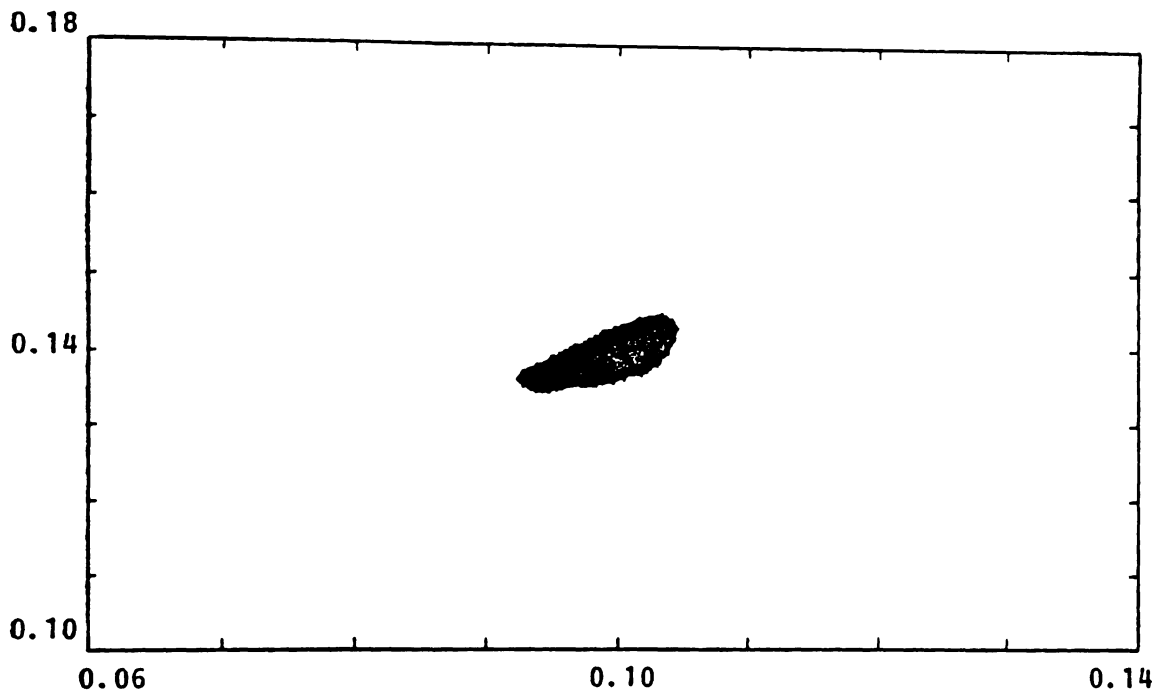
interested in the frequencies here, consider the intersection  $S$  of the solution set given by the MORS in  $\mathbb{R}^4$  with the  $(a_1, a_2)$  plane.  $S$  is displayed in Figure 1, the smallest frequency being assigned to  $a_1$  and the largest to  $a_2$ . The statistics of the points in  $S$  are as follows. The mean and the standard deviation are  $m_1=9.90 \times 10^{-2}$  and  $\sigma_1=2.30 \times 10^{-3}$  for  $a_1$ ,  $m_2=13.92 \times 10^{-2}$  and  $\sigma_2=1.61 \times 10^{-3}$  for  $a_2$ . Since the MORS is equally likely to produce any of these solutions, the worst case should be considered: in terms of the Euclidean distance from the true object  $(h_1, h_2)$ , the worst feasible frequencies are found to be  $(a_1, a_2)=(0.094, 0.136)$ . This compares favorably with the standard Pisarenko method [4] which gave  $(a_1, a_2)=(0.095, 0.189)$ .

## VI. CONCLUSION

We have presented a simple and easily implementable method of random search (MORS) to synthesize set theoretic estimates in the Cartesian space. Unlike projection methods, it places no limitation on the analytical complexity of the pieces of *a priori* knowledge that can be involved in the description of the solution. Because of its potential low efficiency in high dimensional problems, the MORS is best fitted for applications in which the number of unknown parameters is low.

## REFERENCES

- [1] A. Abo-Taleb and M. M. Fahmy, "Design of FIR Two-Dimensional Digital Filters by Successive Projections," *IEEE Trans. on Circuits and Systems*, vol. CAS-31, pp 801-805, Sept. 1984.
- [2] L. M. Brégman, "The Method of Successive Projection for Finding a Common Point of Convex Sets," *Doklady Akademii Nauk SSSR*, vol. 162, pp 487-490, 1965. (English translation in *Soviet Mathematics - Doklady*, vol. 6, pp 688-692, May 1965).
- [3] P. L. Combettes, *Set Theoretic Estimation in Digital Signal Processing*, Ph.D. Dissertation, Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, June 1989.
- [4] V. F. Pisarenko, "The Retrieval of Harmonics from a Covariance Function," *Geophysical Journal of the Royal Astronomical Society*, vol. 33, pp 347-366, Sept. 1973.
- [5] R. Y. Rubinstein, "Generating Random Vectors Uniformly Distributed inside and on the Surface of Different Regions," *European Journal of Operational Research*, vol. 10, pp 205-209, June 1982.
- [6] M. I. Sezan and H. Stark, "Tomographic Image Reconstruction from Incomplete View Data by Convex Projections and Direct Fourier Inversion," *IEEE Trans. on Medical Imaging*, vol. MI-3, pp 91-98, June 1984.
- [7] H. J. Trussell and M. R. Civanlar, "The Feasible Solution in Signal Restoration," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, pp 201-212, Apr. 1984.
- [8] D. C. Youla and H. Webb, "Image Restoration by the Method of Convex Projections: Part 1 - Theory," *IEEE Trans. on Medical Imaging*, vol. MI-1, pp 81-94, Oct. 1982.



**Figure 1. The Set Theoretic Solutions.**