



Light Scattering data analysis

via set inversion

L. Jaulin[◇], J.-L. Godet^{*},
E. Walter^{*}, A. Elliasmine^{*} and Y. Le Duff^{*}

[◇] *Laboratoire d'Ingénierie des Systèmes Automatisés, Université d'Angers, Faculté
des Sciences, 2 boulevard Lavoisier, 49045 Angers, France*

^{*} *Laboratoire des Propriétés Optiques des Matériaux et Applications, Université
d'Angers, Faculté des Sciences, 2 boulevard Lavoisier, 49045 Angers, France*

^{*} *Laboratoire des Signaux et Systèmes, CNRS, Supélec, Plateau de Moulon,
91192 Gif-sur-Yvette, France*

January 8, 1999

Abstract

Set-inversion is applied to recent collision-induced scattering data concerning gaseous CF_4 . It makes it possible to approximate the set of all vectors of independent components of the CF_4 dipole-quadrupole and dipole-octopole polarizability tensors. Numerical analysis shows that short range effects must be taken into account in the high frequency range of each dipole-multipole contribution to the CIS isotropic spectrum of CF_4 . It also demonstrates the agreement between experiment and recent *ab initio* calculations.

1 Introduction

In low density fluids, interactions between molecules are binary. Thus, collision-induced light scattering (CIS) results from collisional polarizabilities of molecular pairs [1]. For optically isotropic molecules like CF_4 , pure collision-induced depolarized and isotropic spectra are observed in the vicinity of the Rayleigh line, where no monomolecular scattering is allowed [2]. These spectra provide information on molecular interactions and may be used to estimate the origin-independent parameters of the dipole-multipole polarizability tensors (*e.g.* A and E which characterize the dipole-quadrupole and dipole-octopole tensors \mathbf{A} and \mathbf{E} of any tetrahedral molecule). For pairs of molecules that cannot be easily described in terms of quantum mechanics, a semi-classical model may be considered [3]. In this case, the CF_4 spectral intensity I^s can be written as follows [4, 5]:

$$\begin{aligned} I^s(\omega) = & I_{DID}^s(\omega) + c_{\alpha\mathbf{TA}}^s \alpha^2 A^2 I_{\alpha\mathbf{TA}}(\omega) + c_{\alpha\mathbf{TE}}^s \alpha^2 E^2 I_{\alpha\mathbf{TE}}(\omega) \\ & + c_{\mathbf{ATA}}^s A^4 I_{\mathbf{ATA}}(\omega) + c_{\mathbf{ATE}}^s A^2 E^2 I_{\mathbf{ATE}}(\omega) + c_{\mathbf{ETE}}^s E^4 I_{\mathbf{ETE}}(\omega) + \dots \quad (1) \end{aligned}$$

where ω denotes the frequency shift and α is the main polarizability. The subscript DID refers to the dipole-induced-dipole interaction. Subscripts $\alpha\mathbf{TA}$, $\alpha\mathbf{TE}$, \mathbf{ATA} , \mathbf{ATE} and \mathbf{ETE} refer to the successive dipole-multipole light scattering mechanisms [4, 5]. The coefficients c^s depend on the nature s of the spectrum (depolarized

or isotropic) and on the dipole-multipole mechanisms [4, 5]. Unfortunately, point estimation of A and E is highly uncertain due to the following four reasons:

i) Short range effects such as overlap and exchange effects are not taken into account in the aforementioned semi-classical model. As a result, in the CF_4 case, depolarized and isotropic spectra lead to different conclusions. The isotropic spectrum has been shown to be the more adapted [5].

ii) Because of the competition between the dipole-multipole mechanisms (like $\alpha\mathbf{TA}$ and $\alpha\mathbf{TE}$) several different parameter vectors (A, E) may correspond to similar fits, *i.e.* the model is said to be almost unidentifiable.

iii) Errors on measurement of the isotropic spectrum are large [5] and many unsimilar fits can be considered as consistent with all data.

iv) The model is nonlinear (*cf.* Eq. (1)) and local minimization procedures may converge to any local minimum.

For all these reasons, a bounded-error estimation approach (see [6], [7] and references therein) is considered in this work. It consists of characterizing the set \mathcal{S} of all values of the vector (A, E) such that the associated model output is consistent with all

experimental data, *i.e.* goes through all error bars. The method to be used is new for most physicists and based on set inversion [8]. Set inversion, presented in Section 2, is particularly suited in our case because of the nonlinearities involved in the semi-classical model; it uses *interval analysis* which is a numerical tool for computing with sets (also presented in Section 2). Section 3 gives an approximation of the set \mathcal{S} of all feasible vectors (A, E) and compares it with former estimations. We show that recent *ab initio* computations by Maroulis [9, 10] partly confirm our results. Finally, we conclude on the advantages that set inversion analysis offers to spectroscopists.

2 Set inversion

Let \vec{f} be a nonlinear continuous vector function mapping R^n into R^m and \mathcal{Y} be a subset of R^m . The set inversion problem is to characterize the set \mathcal{X} defined by

$$\mathcal{X} = \{ \vec{x} \mid \vec{f}(\vec{x}) \subset \mathcal{Y} \} = \vec{f}^{-1}(\mathcal{Y}).$$

The set function \vec{f}^{-1} is the *reciprocal function* of \vec{f} . The set-inversion algorithm to be presented is based on interval arithmetic which is a numerical tool originally developed in order to quantify the effect of finite-precision arithmetic on results obtained by a computer [11]. Interval arithmetic extends classical operators on real numbers to

intervals in a natural way. Thus, if $[x] = [x^-, x^+]$ and $[y] = [y^-, y^+]$,

$$[x] + [y] = [x^- + y^-, x^+ + y^+],$$

$$[x] - [y] = [x^- - y^+, x^+ - y^-],$$

$$[x] \cdot [y] = [\min(x^-y^-, x^-y^+, x^+y^-, x^+y^+), \max(x^-y^-, x^-y^+, x^+y^-, x^+y^+)].$$

For example, we have $([1, 2] + [-3, 4]) \cdot [-1, 5] = [-2, 6] \cdot [-1, 5] = [-10, 30]$,

As another example, let us consider the real function $f(x) = x^2 + 2x + 4$. An interval evaluation for f is $[f]([x]) = [x] \cdot [x] + 2[x] + 4$. For $[x] = [-3, 4]$, we have:

$$[f]([-3, 4]) = [-3, 4] \cdot [-3, 4] + 2[-3, 4] + 4 = [-12, 16] + [-6, 8] + 4 = [-14, 28].$$

Note that the actual image by f of the interval $[x]$, $f([-3, 4]) = [3, 28]$ is a subset of the interval evaluation $[f]([-3, 4]) = [-14, 28]$. This illustrates that interval evaluation is usually pessimistic [11].

A *box* or *vector interval* $[\vec{x}]$ of R^n is defined as the Cartesian product of n intervals.

$$[\vec{x}] = [x_1^-, x_1^+] \times \cdots \times [x_n^-, x_n^+] \quad (2)$$

The i^{th} component of the box $[\vec{x}]$ is an interval denoted by $[x]_i$. It can be proven (see [11]) that the interval evaluation $[f]([\vec{x}]) = [f]([x]_1, \dots, [x]_n)$ always contains the image interval $f([\vec{x}])$, i.e.

$$\forall [\vec{x}], f([\vec{x}]) \subset [f]([\vec{x}]). \quad (3)$$

The *width* $w([\vec{x}])$ of a box $[\vec{x}]$ is the size of its largest side. For instance, the width of the box $[\vec{x}] = [1, 2] \times [-1, 3]$ is equal to 4. A *principal plane* of $[\vec{x}]$ is a symmetry plane of $[\vec{x}]$ normal to a side of maximum length. To *bisect* a box $[\vec{x}]$ means to cut it along one of its principal planes. Bisecting $[\vec{x}] = [1, 2] \times [-1, 3]$ produces two boxes $[\vec{x}](1) = [1, 2] \times [-1, 1]$ and $[\vec{x}](2) = [1, 2] \times [1, 3]$.

The algorithm SIVIA (Set Inverter Via Interval Analysis) partitions some prior box of interest $[\vec{x}](0)$ into a set of non-overlapping boxes. For the sake of simplicity, it is assumed that the set \mathcal{Y} to be inverted is a box $[\vec{y}]$. SIVIA uses the two following tests to decide whether a given box $[\vec{x}]$ is inside or outside the solution set \mathcal{X} :

$$\begin{aligned} \text{(i)} \quad & \forall i, [f]_i([\vec{x}]) \subset [y]_i \quad \Rightarrow \quad [\vec{x}] \subset \mathcal{X} \\ \text{(ii)} \quad & \exists i \mid [f]_i([\vec{x}]) \cap [y]_i = \emptyset \quad \Rightarrow \quad [\vec{x}] \cap \mathcal{X} = \emptyset. \end{aligned} \tag{4}$$

Proofs

(i) If $\forall i, [f]_i([\vec{x}]) \subset [y]_i$, then from (3), $\forall i, f_i([\vec{x}]) \subset [y]_i$, i.e. $\vec{f}([\vec{x}]) \subset [\vec{y}]$.

Therefore $[\vec{x}] \subset \mathcal{X}$.

(ii) If $\exists i \mid [f]_i([\vec{x}]) \cap [y]_i = \emptyset$, then from (3), $\exists i \mid f_i([\vec{x}]) \cap [y]_i = \emptyset$.

Therefore $\vec{f}([\vec{x}]) \cap [\vec{y}] = \emptyset$, i.e. $[\vec{x}] \cap \mathcal{X} = \emptyset$.

SIVIA is a recursive routine that brackets the solution set \mathcal{X} between an inner set of boxes and an outer set of boxes. For the sake of simplicity, it will be presented

here in the case of a two dimensional solution set, but the algorithm readily extends to higher dimensions [8]. Boxes that have been proved to belong to \mathcal{X} via test (i) are drawn in dark grey, those that have been proved to be outside \mathcal{X} via test (ii) are drawn in light grey and those that satisfy neither (i) nor (ii) and are too small to be bisected are drawn in white. The accuracy ε is a small positive real number.

SIVIA(\vec{x})

Step 1 $\forall i, [f]_i([\vec{x}]) \subset [y]_i, \{\text{draw}([\vec{x}], \text{'darkgrey'}); \text{return}\};$

Step 2 $\exists i \mid [f]_i([\vec{x}]) \cap [y]_i = \emptyset, \{\text{draw}([\vec{x}], \text{'lightgrey'}); \text{return}\};$

Step 3 If $w(\vec{x}) < \varepsilon, \{\text{draw}([\vec{x}], \text{'white'}); \text{return}\};$

Step 4 Bisect $[\vec{x}]$ and store the two resulting boxes into $[\vec{x}](1)$ and $[\vec{x}](2)$;

Step 5 SIVIA($[\vec{x}](1)$); SIVIA($[\vec{x}](2)$);

SIVIA is first called for $[\vec{x}] = [\vec{x}](0)$, where $[\vec{x}](0)$ is a box assumed to contain the solution set \mathcal{X} . If we denote by $\Delta\mathcal{X}$ the union of all white boxes and by \mathcal{X}^- the union of all dark grey boxes, then the solution set \mathcal{X} is bracketed by:

$$\mathcal{X}^- \subset \mathcal{X} \subset \mathcal{X}^- \cup \Delta\mathcal{X} \tag{5}$$

Remark 1 *When the box \mathcal{Y} to be inverted is a singleton $\{\overrightarrow{y}\}$ (for example when dealing with error-free data), the solution set \mathcal{X} is often reduced to a singleton $\{\overrightarrow{x}\}$ which is easily found by SIVIA or by other punctual approaches. When two or more solutions exist, SIVIA detects all of them in a guaranteed way, contrary to punctual approaches.*

Remark 2 *Inversion methods generally considered are punctual: they try to find the best fit. In a linear context, numerical instability appears when the matrix to be inverted is almost non invertible. With a set inversion approach, the problem of instability does not exist. Even if the model is non identifiable (which corresponds to a situation where the matrix is non invertible in a punctual and a linear context), a set with a stretched shape is obtained. All informations about uncertainties (numerical, errors on measurements, ...) are given by \mathcal{X} : if \mathcal{X} is big, as in the application treated in the next section, the problem can be considered as "ill-posed" in a punctual point of view.*

3 Results and discussion

The experimental CF_4 isotropic spectrum I^{iso} reported in [5] had been recorded for a set of Raman frequency shifts ω_i relative to the green spectral line ($\lambda_L = 514.5 \text{ nm}$) of an argon laser. Following the bounded-error approach, uncertainties on experimental data are assumed to be bounded, *i.e.* the i th ideal measurement I_i^{iso} (the measurement that should be obtained if no measurement errors occurred) is assumed to belong to the interval $[I_{min}^{iso}(i), I_{max}^{iso}(i)]$ provided in [5] and recalled in Table 1.

Possible location for Table 1

As regards the theoretical spectrum, and according to Eq. (1), the i th model output is given by

$$\begin{aligned}
 I_i^{iso}(A, E) = & DID_i + \alpha^2 A^2 \alpha \mathbf{TA}_i + \alpha^2 E^2 \alpha \mathbf{TE}_i \\
 & + A^4 \mathbf{ATA}_i + A^2 E^2 \mathbf{ATE}_i + E^4 \mathbf{ETE}_i.
 \end{aligned} \tag{6}$$

The dipole-dipole and dipole-multipole components ($DID_i = I_{DID}^{iso}(\omega_i)$, $\alpha \mathbf{TA}_i = c_{\alpha \mathbf{TA}}^{iso} I_{\alpha \mathbf{TA}}(\omega_i)$, *etc.*) are provided in Table 1 (see [5] for the details of their computation). Since the value of the CF_4 polarizability is known ($\alpha = 2.93 \text{ \AA}^3$ for

$\lambda_L = 514.5 \text{ nm}$ [12]), the unknown parameters are A and E . Prior feasible intervals for A and E are given by

$$A \in [0, 2] \text{ and } E \in [0, 4]. \quad (7)$$

where A and E are expressed in \AA^4 and \AA^5 units, respectively. Table 2 shows how the problem of estimating A and E can be cast into the framework of set inversion

Possible location for Table 2

In less than 5 seconds on a Pentium 100 Personal Computer, SIVIA brackets \mathcal{S} as represented in Fig. 1.

Possible location for Figure 1

Because of some mechanism not taken into account in our semi-classical model, it may happen that SIVIA eliminates a part of the parameter space that could contain the true values of A and E . To protect against this, one would like to be especially careful about data points that turn out to have a critical influence on the size of \mathcal{S} .

For this purpose, we define the *safety* of the i th interval data by

$$\gamma(i) = \frac{Vol(\mathcal{S})}{Vol(\mathcal{S}_i)}, \quad (8)$$

where Vol corresponds to the volume (an area in our two-dimensional case) and \mathcal{S}_i is the set of all parameters that are consistent with all data but the i th. The smaller the safety is, the more careful one must be with the corresponding data point. SIVIA can easily be adapted to compute volumes of sets [8] and therefore the coefficients $\gamma(i)$. We obtain $\gamma(1) = \gamma(3) = \gamma(5) = 1.$, which means that any of the associated data points can be removed without changing the feasible domain for (A, E) . Moreover, $\gamma(2) = 0.98, \gamma(4) = 0.98, \gamma(6) = 0.93$ and $\gamma(8) = 0.74$. The safety of the 8th measurement is the smallest and the reliability of the semi-classical model at the corresponding frequency $\omega_8 = 120 \text{ cm}^{-1}$ can be questioned.. After removing the 8th measurement, SIVIA brackets \mathcal{S}_8 as represented in Fig. 2. Note that the upper boundary of \mathcal{S}_8 is not significantly different from that of \mathcal{S} , but the lower boundary of \mathcal{S}_8 lie much lower.

Possible location for Figure 2

In [5], an empirical estimation leads to the intervals $[0.5, 1.2] \text{ \AA}^4$ and $[1.0, 3.5] \text{ \AA}^5$ for A and E , respectively. The corresponding rectangle, drawn with dashed lines in Figs.1-

2, contains a large part of the solution sets \mathcal{S} and \mathcal{S}_8 . Moreover, the maximum values of A and E given in [5] are very close to those provided here. The aforementioned rectangle, however, also contains parameter vectors that are not consistent with the data. Finally, note that the vector $(A, E) = (0.97\text{\AA}^4, 1.15\text{\AA}^5)$, recently calculated *ab initio* by Maroulis [9, 10] and represented by a white cross on Figs. 1-2, is located outside \mathcal{S} and deep inside \mathcal{S}_8 . This corroborates the hypothesis that the 8th data is not consistent with the semi-classical model. The discrepancy may be due to short range effects that take place for high frequencies of each dipole-multipole contribution. When scattering data for frequency shifts higher than 120 cm^{-1} are considered, similar discrepancies can be observed [5].

In conclusion, the set inversion approach, advocated here, makes it possible to estimate unknown parameters, their uncertainties, as well as their correlations, when bounded-error data and nonlinear models are involved. Such problems arise often in spectroscopy where many efforts are focused on estimating physical coefficients from experimental data (*e.g.* collision-induced scattering or absorption processes). The example of recent CIS studies on gaseous CF_4 is but one illustration of the advantages of set-inversion over more conventional methods.

References

- [1] *Collision–and Interaction–Induced Spectroscopy*, vol. 452 of *NATO ASI Series C: Mathematical and Physical Sciences* Edited by G.C. Tabisz and M.N. Neuman, Kluwer Academic Publishers, Dordrecht (1995).
- [2] A. D. Buckingham and G. C. Tabisz, *Mol. Phys.* **36**, 583 (1978).
- [3] H. Posch, *Mol. Phys.* **46**, 1213 (1982).
- [4] A. Elliasmine, J.-L. Godet, Y. Le Duff, and T. Bancewicz, *Mol. Phys.*, **90**, 147 (1997).
- [5] A. Elliasmine, J.-L. Godet, Y. Le Duff, and T. Bancewicz, *Phys. Rev. A*, **55**, (June 1997).
- [6] E. Walter and L. Pronzato, *Identification of Parametric Models from Experimental Data*, Springer, London, (1997).
- [7] *Bounding Approaches to System Identification*, Edited by M. Milanese, John Norton, H. Piet-Lahanier and E. Walter, Plenum, New York (1997).
- [8] L. Jaulin and E. Walter, *Automatica*, **29**, 1053 (1993).
- [9] G. Maroulis, *Chem. Phys. Lett.*, **259**, 654 (1996)

- [10] G. Maroulis, *J. Chem. Phys.*, **105**, 8467 (1996).
- [11] R.E. Moore, *Methods and Applications of Interval Analysis*, SIAM, Philadelphia, 1979.
- [12] H. E. Watson and K. L. Ramaswamy, *Proc. Roy. Soc. London A* **156**, 144 (1936).

Table 1

Experimental intensities of the CF_4 isotropic spectrum versus frequency shifts together with corresponding dipole-dipole and dipole-multipole theoretical contributions.

| i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | units |
|--------------------|-------|-------|-------|-------|-------|------|------|------|---|
| ω | 50 | 60 | 70 | 80 | 90 | 100 | 110 | 120 | cm^{-1} |
| I_{min}^{iso} | 192 | 177 | 116 | 75 | 41 | 22 | 15 | 10 | 10^{-58}cm^6 |
| I_{max}^{iso} | 1732 | 1003 | 656 | 299 | 165 | 89 | 61 | 41 | 10^{-58}cm^6 |
| DID | 27.78 | 14.17 | 7.53 | 4.18 | 2.34 | 1.34 | 0.78 | 0.46 | 10^{-58}cm^6 |
| α TA | 133.1 | 80.57 | 44.55 | 22.55 | 10.47 | 4.53 | 1.87 | 0.77 | $10^{-58} \text{cm}^6 \text{\AA}^{-14}$ |
| α TE | 6.39 | 4.78 | 3.39 | 2.27 | 1.44 | 0.87 | 0.50 | 0.27 | $10^{-58} \text{cm}^6 \text{\AA}^{-16}$ |
| ATA | 196.1 | 153.7 | 114.9 | 81.9 | 55.7 | 36.1 | 22.4 | 13.3 | $10^{-60} \text{cm}^6 \text{\AA}^{-16}$ |
| ATE | 20.20 | 16.94 | 13.73 | 10.74 | 8.12 | 5.93 | 4.19 | 2.86 | $10^{-60} \text{cm}^6 \text{\AA}^{-18}$ |
| ETE | 0.70 | 0.61 | 0.52 | 0.43 | 0.35 | 0.27 | 0.21 | 0.16 | $10^{-60} \text{cm}^6 \text{\AA}^{-20}$ |

Table 2

Translation table between the set inversion formalism and the problem of estimating the A and E values of CF_4 .

| Set inversion | Estimation of A and E |
|--|---|
| $[y]_i$ | $[I_{min}^{iso}(i), I_{max}^{iso}(i)]$ |
| $\mathcal{Y} = [\vec{y}]$ | $[I_{min}^{iso}(1), I_{max}^{iso}(1)] \times \dots \times [I_{min}^{iso}(8), I_{max}^{iso}(8)]$ |
| \vec{x} | $(A, E)^T$ |
| $f_i(\vec{x})$ | $I_i^m(A, E)$ |
| \vec{f} | $(I_1^m(A, E), \dots, I_8^m(A, E))^T$ |
| $\mathcal{X} = \{\vec{x} \mid \vec{f} \subset \mathcal{Y}\}$ | $\mathcal{S} = \{(A, E) \mid \forall i, I_i^m(A, E) \in [I_{min}^{iso}(i), I_{max}^{iso}(i)]\}$ |
| $[\vec{x}](0)$ | $[0, 2] \times [0, 4]$ |

Figure Captions

Figure 1: Set of A and E that are consistent with all data (dark grey area) in the (A, E) plane. The units are in \AA^4 for A and in \AA^5 for E . The dashed rectangle corresponds to the intervals for A and E provided in [5], the white cross represents the vector (A, E) found by Maroulis [9, 10], and the light grey area is the set of forbidden parameter vectors.

Figure 2: Set of A and E that are consistent with the first seven data.

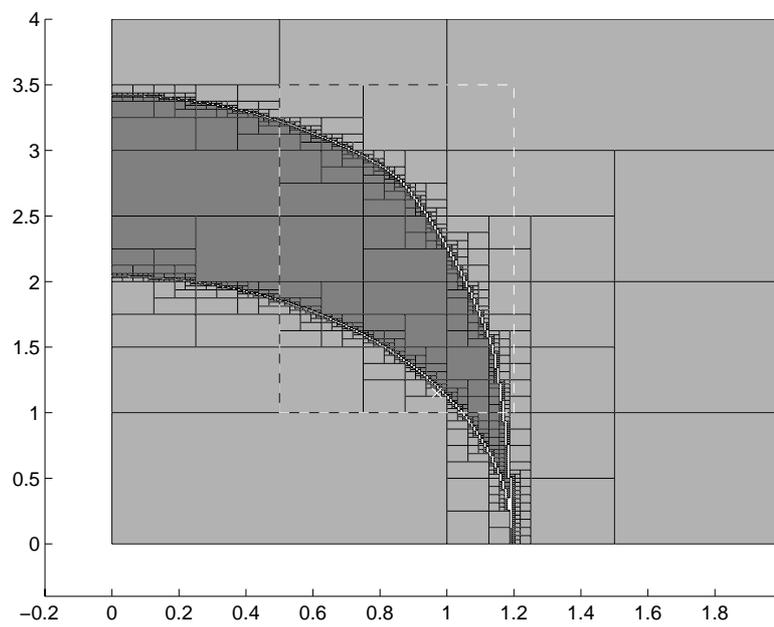


Figure 1:

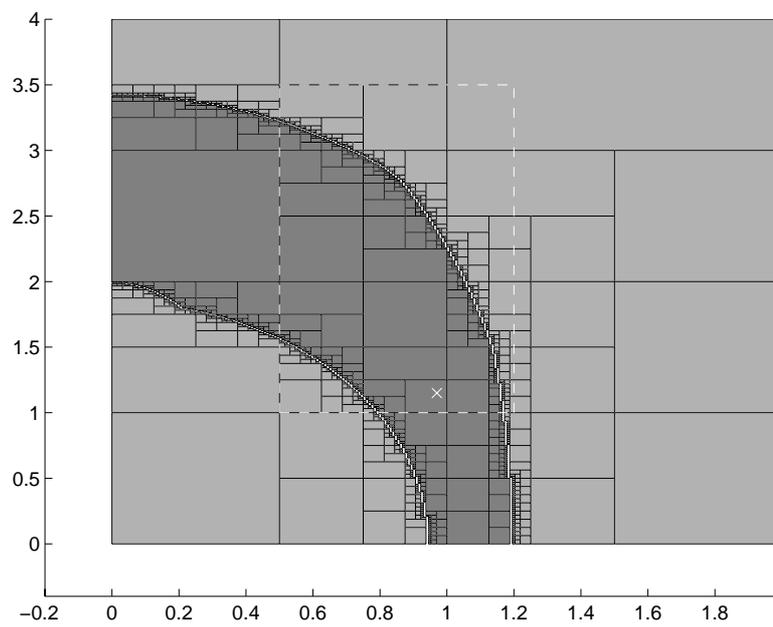


Figure 2: