Maximization of the Fisher Information in PDA

Jérémy Payan, Annie-Claude Pérez, Claude Jauffret

Maximization of the Fisher Information in PDA

Jérémy Payan, Claude Jauffret, Annie-Claude Pérez
Aix Marseille Université, Université de Toulon, CNRS, IM2NP, Marseille, France
CS 60584 83041 TOULON Cedex 9, France
jeremy-payan@etud.univ-tln.fr *(Student)*
jauffret, annie-claude.perez @univ-tln.fr

***Abstract*—In a cluttered environment, the probabilistic data association (PDA) model allows constructing efficient estimators. In this case, the Fisher information matrix (FIM) is equal to the FIM in the clean environment multiplied by the so-called information reduction factor. This factor depends implicitly on the detection and false alarms probabilities, hence on the threshold prior to the estimation step. The topic of this paper is to seek the optimal threshold, the one that maximizes the information reduction factor. Whereas the Poisson law is used as an approximation in PDA model, here we consider the binomial law for the large false alarms probabilities. An example, coming from signal processing, illustrates our analysis.**

I. Introduction

Depending on the architecture of a surveillance system, target tracking takes place at three different levels:

- "Track-before-detect" (TBD) based on unthresholded measurements (for example in [1]).
- "Track-after-detect": after the detection step, the available data are composed of false alarms and correct detections. The tracking can be made by the Probabilistic Data Association (PDA) algorithm. We are facing a cluttered environment.
- "Track-after-detect" and extraction of the line of measurements: the input of the tracking function is a line of supposedly 'good' measurements, which is an optimistic and unrealistic situation. Such tracking is said to be in a clean environment.

The topic of this paper takes place in the second scheme. We are concerned with the determination of the 'optimal' threshold in the detection step. Indeed, a poor choice of the threshold can lead to too high a number of false alarms or a very low level of detection probability. In both cases, we intuitively understand that no information may be drawn from these situations. We deduce that the information might be maximal at a particular threshold. Fortunately, we have a tool to quantify the information contained in the available measurements: the Fisher Information Matrix (FIM).

In the PDA model, the FIM in the presence of clutter equals the FIM in a clean environment multiplied by a scalar factor [2]. The so-called q_2 factor lies between 0 and 1, and explains the information loss due to a certain amount of false alarms and non-unity detection probability [2], [3]. This factor is a function of the detection probability and the amount of false alarms in a restricted space. With all these considerations, the existence of an optimal detection threshold that would maximize q_2 is a fundamental question. This paper deals with the search for an optimal threshold that maximizes q_2. Optimal thresholds have been investigated in [4], [5], with a specific criterion at each time (target position error in [4], track lifetime and track loss in [5]).

Our paper is composed of four main sections. Section II presents briefly the PDA with its uses and necessary statistical assumptions, as well as the computation of the FIM in PDA. Section III shows how to optimize q_2, by changing the false alarms repartition model. After detailing a simple example to solve our problem, we present in Section IV some numerical results. Then, the conclusion, appendix and references follow.

II. The Fisher Information Matrix in Probabilistic Data Association

A. Introduction to Probabilistic Data Association

Probabilistic Data Association (PDA) is well described in [3]. PDA is used in realistic tracking contexts. Indeed, it takes into account the presence of false alarms (which are the element of the clutter), as well as a non-unity detection probability. The main aim is to estimate the parameters of interest when true detections are drowned in false alarms, and don't surely appear. Historically, two estimation techniques have been studied:

- A recursive technique : Probabilistic Data Association Filter (PDAF) allowing real-time estimation, based upon a modified Kalman filter.
- A batch technique : PDA Likelihood Maximization (PDA-MLE).

In our study, the latter is used. Some statistical assumptions are necessary to establish our model, following [2], [3] :

- The detections due to the target are corrupted by an additional zero-mean white Gaussian noise. The power of the noise is σ^2. Subsequently, these detections will be called "true detections".
- The clutter is distributed according to an uniform law in a finite scan space u. The clutter is composed of the false alarms.
- The amount of false alarms at a specific sample k follows a discrete probability law μ_{fa}. Usually, $\mu_{fa} = \mathcal{P}$, the

Poisson law of parameter λu (expected number of false alarms).

- True detections appear at most one time at each sampling time, with a detection probability P_d.
- The available measurement vector at scan k is $z_k = (z_{1,k}, \cdots, z_{m_k,k})^T$
- The vectors z_k are independent conditionally to X for $k = 1 : K$.

Applying the total probability theorem, we end up with the likelihood[1] of X given the measurement vector z_k:

$$L(X|z_k) = \frac{1-P_d}{u^{m_k}}\mu_{fa}(m_k) + \frac{P_d}{u^{m_k-1}}\frac{\mu_{fa}(m_k-1)}{m_k} \times \sum_{i=1}^{m_k}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{z_{i,k}-s_k(X)}{\sigma}\right)^2\right] \tag{1}$$

with $s_k(X)$ the state model at time k. Thanks to the independence assumption, we get $L(X|z) = \prod_{k=1}^{K} L(X|z_k)$.

B. The Fisher Information Matrix

We recall the well-known expression for the FIM in a clean environment, called J:

$$J = \sum_{k=1}^{K}\frac{1}{\sigma^2}\nabla_X s_k(X)\nabla_X^T s_k(X) \tag{2}$$

In the presence of clutter, the computation of the FIM is more complicated, but can be done with the above assumptions. The technique consists in restricting the measurement space to a gate around the true detection [2]–[4]. Its size, denoted by v_g, depends on the standard deviation of the measurements:

$$v_g = 2g\sigma \tag{3}$$

Since the true detections have a Gaussian distribution, the most informative measurements are located in this gate when $g = 5$. Indeed, more than 99 % of a Gaussian population are within five standard deviations of the mean. Conversely, measurements outside the gate contain no additional information. As a consequence, the likelihood function can be approximated by the one computed only with measurements in the gate. Doing this, the FIM in a cluttered environment is proven to be equal to the FIM in clean environment multiplied by a factor q_2:

$$F = q_2 J \tag{4}$$

When μ_{fa} is a Poisson law (the usual assumption), q_2 is expressed with respect to parameters λv_g and detection probability P_d: $q_2 = q_2(P_d, \lambda v_g)$ [1]–[3], [7].

III. Optimization of the Information Reduction Factor

A. Poisson law versus binomial law

The use of the Poisson law in (1) is justified by the fact that the threshold of the prior test is chosen for a low probability of false alarms P_{fa} [8]. Indeed, if M is the number of measurement cells on which the test is carried out, we have $\lambda u \simeq MP_{fa}$, provided P_{fa} is small enough. Consequently, $\lambda v_g \simeq N_g P_{fa}$, where N_g is the number of measurement cells in the gate. Hence, q_2 is a function of the couple $(P_d; P_{fa})$. Finally, using the ROC curves of the detector, we end up with a new expression for q_2, depending only on the threshold t: $q_2(P_d, \lambda v_g) = q_2(t)$

The search for the maximum q_2 is nothing else than the search for the optimal t. But, the optimal t could define a probability of false alarm which could be large. In this case, the use of the Poisson law is no longer justified. This is why we have to introduce the 'natural' false alarm probability into (1): the binomial law with N_g and P_{fa} as parameters. Generally speaking, N_g depends on an analysis of the methods of measurement, and the way the data are processed and quantified.

B. Information Reduction Factor Derived from a Binomial Law

The reduction information factor q_2 may then be derived from the binomial distribution:

$$q_2(t) = \sum_{m=1}^{N_g}\frac{2P_d(t)\mathcal{B}(m-1)}{\sqrt{2\pi}g^{m-1}} \times \int_0^g\cdots\int_0^g \frac{\exp\left(-\xi_1^2\right)\xi_1^2}{\frac{a(t)}{\beta_m(t)}+\sum_{i=1}^{m}\exp\left(-\frac{1}{2}\xi_i^2\right)}d\xi_1\cdots d\xi_m \tag{5}$$

q_2 is then a sum of m-fold integrals, with $\mathcal{B}(m-1)$ being the binomial distribution with parameters N_g and P_{fa}, t the detection threshold, and

$$a(t) = 1 - P_d(t) \tag{6}$$

$$\beta_m(t) = \frac{2gP_d(t)\left[1-P_{fa}(t)\right]}{\sqrt{2\pi}P_{fa}(t)(N_g-m+1)} \tag{7}$$

$$g = 5 \tag{8}$$

The derivation of the above is detailed in the Appendix, and based on [2], [7].

We note that under the conditions of the convergence of the binomial towards the Poisson distribution, q_2 expressed with both distributions has the same values. Fig. 1 gives us an example of the two q_2, coming from signal processing and detailed hereafter. So, we have generalized the information reduction factor concept, by extending it to cases with high P_{fa}.

IV. Application to Signal Processing

A. Problem Formulation

The context we describe serves as an example for optimal threshold research. The framework is the following: we have a sine wave signal, with unknown amplitude, frequency and phase [9]. This is our signal of interest. This sine wave is

[1] We implicitly use the notation proposed in [6].

corrupted by an additive zero-mean white Gaussian noise. There are two detection hypotheses:

- $\mathrm{H}_0 \triangleq$ "there is only noise"
- $\mathrm{H}_1 \triangleq$ "there are signal of interest and noise"

Ref. [10] shows that the Neyman-Pearson optimal test is equivalent to comparing the periodogram of the signal to a threshold. The probability distribution of the periodogram under each previous hypothesis is:

- H_0: central χ_2^2 distribution
- H_1: noncentral χ_2^2 distribution

The noncentrality parameter γ depends on the SNR and on the length of the samples as follows [10]:

$$\gamma = \frac{NA^2}{2\sigma^2} \tag{9}$$

with N the number of samples, A the amplitude of the sine wave, and σ^2 the power of the noise. So, if we call the Neyman-Pearson test Λ and the threshold t, we have by definition

$$P_d(t) \triangleq P\left(\Lambda > t | \, \mathrm{H}_1\right) \tag{10}$$

$$P_{fa}(t) \triangleq P\left(\Lambda > t | \, \mathrm{H}_0\right) \tag{11}$$

In this application, the temporal signal is segmented into K non-overlapped scans. The periodogram of each scan is compared to the threshold. The frequency cells of scan k greater than the threshold compose the measurement vector z_k.

By the knowledge of probability density function (pdf), we deduce easily that

$$P_{fa}(t) = \exp\left(-\frac{t}{2}\right) \tag{12}$$

The expression of P_d is a numerical series detailed in [10], in which an algorithm numerically computing P_d is presented. So, we are able to express both P_d and P_{fa} with respect to the threshold. In this context, N_g is the number of frequency cells in the gate.

B. Numerical Results

With the previous knowledge of the pdf under both hypotheses, and the new expression for q_2, we may henceforth compute q_2 with respect to the detection threshold. We compute q_2 numerically using Monte-Carlo runs [11] with 500 000 simulations, and N_g set to 5. The search for the maximum is based on a grid-search method. The optimal threshold in the sense of q_2 may then be placed on a ROC curve, to find the optimal operating point, for different values of γ.

Fig. 1 shows the importance of choosing the binomial law instead of the Poisson law to search for the maximum q_2. The Poisson law may indeed prevent the finding of the maximum in certain cases, as shown by the considerations from previous section. Most of the time, the search for the maximum would be possible under the Poisson distribution if the optimal threshold is not too low. Fig. 2 shows q_2 with respect to the threshold for different values of the noncentrality parameter γ, as well as the optimal threshold in each case. The higher the noncentrality parameter, the higher the optimal threshold. This underscores the differences between the pdf under both hypotheses. Fig. 3 represents the ROC curves for each previous case and the operating point from the optimal threshold, highlighting the powerful performance, even in the case of relatively low γ.

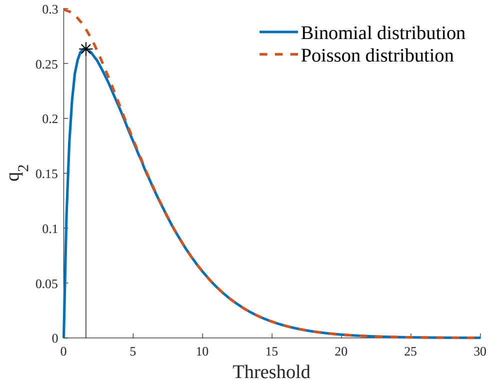

Fig. 1. q_2 with both Poisson and binomial laws, $\gamma = 2$

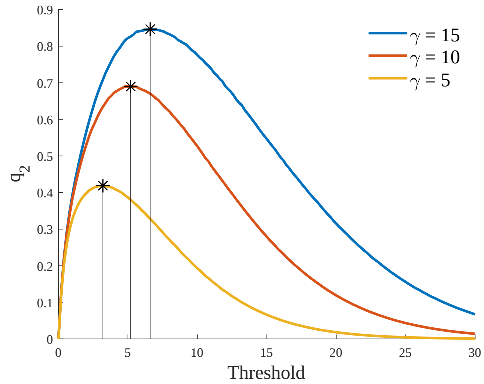

Fig. 2. q_2 with respect to threshold for different values of γ

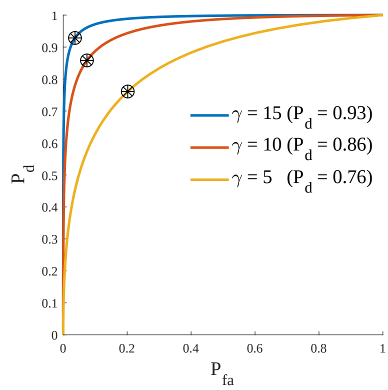

Fig. 3. ROC curves for different values of γ

V. Conclusion

In this paper, we have posed and solved the problem of choosing the optimal detection threshold in a surveillance system. The criterion is the amount of information relative to the parameter of interest and contained in the retained binary measurements. Based upon the assumptions of the PDA, the optimal threshold is the one for which the information reduction factor is the largest. An example of signal processing illustrates our approach. In the future, a similar problem will be studied when the measurements are no longer binary, but coupled with their own energy as in [1].

Appendix
Derivation of the Information Reduction Factor with Binomial Law

We derive q_2 with a binomial law, using the same principles as explained in [2], [7]. First, we express the likelihood at time k with (1) restricted to the validation region gate of volume v_g. Moreover, μ_{fa} is replaced with the binomial distribution with parameters N_g and P_{fa}. For the sake of clarity, the subscript k is dropped:

$$\begin{aligned} L(X|z) = & \frac{1-P_d(t)}{v_g^m}\mathcal{B}(m) + \frac{P_d(t)}{v_g^{m-1}}\frac{\mathcal{B}(m-1)}{m} \\ & \times \sum_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{z_i - s(X)}{\sigma}\right)^2\right] \end{aligned} \tag{13}$$

where $v_g = 2g\sigma$ is the volume of the gate. After factorization and highlighting that

$$\frac{\mathcal{B}(m-1)}{\mathcal{B}(m)} = \frac{m(1-P_{fa}(t))}{P_{fa}(t)(N_g - m + 1)} \tag{14}$$

we get

$$\begin{aligned} L(X|z) = & \frac{\mathcal{B}(m)}{v_g^m}\left\{1 - P_d(t) + \frac{P_d(t)v_g(1-P_{fa}(t))}{P_{fa}(t)(N_g - m + 1)}\right. \\ & \left. \times \sum_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{z_i - s(X)}{\sigma}\right)^2\right]\right\}. \end{aligned} \tag{15}$$

Then, it is possible to compute FIM in clutter from (15). The latter is then the FIM in a clean environment multiplied by the scalar factor q_2, which is now

$$\begin{aligned} q_2(t) = & \sum_{m=1}^{N_g} \sigma^2 \frac{1}{2\pi\sigma^2}\left[\frac{P_d(t)v_g(1-P_{fa}(t))}{P_{fa}(t)(N_g - m + 1)}\right]^2 \frac{1}{\sigma^2}\frac{\mathcal{B}(m)}{v_g^m} \\ & \times \int_{v_g}\cdots\int_{v_g} \frac{1}{\Phi(X|z)} \sum_{i=1}^{m}\sum_{j=1}^{m} \exp\left[-\frac{1}{2}\left(\frac{z_i - s(X)}{\sigma}\right)^2\right. \\ & \left. -\frac{1}{2}\left(\frac{z_j - s(X)}{\sigma}\right)^2\right] \frac{z_i - s(X)}{\sigma}\frac{z_j - s(X)}{\sigma} dz_1\cdots dz_m \end{aligned} \tag{16}$$

with

$$\begin{aligned} \Phi(X|z) = & 1 - P_d(t) + \frac{P_d(t)v_g(1-P_{fa}(t))}{P_{fa}(t)(N_g - m + 1)} \\ & \times \sum_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{z_i - s(X)}{\sigma}\right)^2\right]. \end{aligned} \tag{17}$$

This way, q_2 is a finite sum of m-fold integrals. We define

$$\xi_i = \frac{z_i - s(X)}{\sigma} \tag{18}$$

Reintroducing in (16) the notations of $a(t)$ and $\beta_m(t)$ given in (6) and (7), we get

$$\begin{aligned} q_2(t) = & \sum_{m=1}^{N_g} \beta_m^2(t)\frac{\mathcal{B}(m)}{v_g^m}\sigma^m \\ & \times \int_{-g}^{g}\cdots\int_{-g}^{g} \frac{1}{a(t) + \beta_m(t)\sum_{i=1}^{m}\exp\left(-\frac{1}{2}\xi_i^2\right)} \\ & \times \sum_{i=1}^{m}\sum_{j=1}^{m} \exp\left(-\frac{1}{2}\xi_i^2 - \frac{1}{2}\xi_j^2\right)\xi_i\xi_j d\xi_1\cdots d\xi_m \end{aligned} \tag{19}$$

For ξ_j fixed, $q_2(t)$ is an odd function with respect to ξ_i. Then, the cross-terms vanish from the integral because of the symmetry of the integration domain:

$$\begin{aligned} q_2(t) = & \sum_{m=1}^{N_g} \beta_m^2(t)\frac{\mathcal{B}(m)}{v_g^m}\sigma^m \\ & \times \int_{-g}^{g}\cdots\int_{-g}^{g} \frac{\sum_{i=1}^{m}\exp\left(-\xi_i^2\right)\xi_i^2}{a(t) + \beta_m(t)\sum_{i=1}^{m}\exp\left(-\frac{1}{2}\xi_i^2\right)} d\xi_1\cdots d\xi_m \end{aligned} \tag{20}$$

We use the parity of the function with respect to ξ_i and note that the integrals are the same for each considered index. After some manipulations, we obtain

$$\begin{aligned} q_2(t) = & \sum_{m=1}^{N_g} \beta_m(t)\frac{\mathcal{B}(m)}{v_g^m}\sigma^m 2^m m \\ & \times \int_{0}^{g}\cdots\int_{0}^{g} \frac{\exp\left(-\xi_1^2\right)\xi_1^2}{\frac{a(t)}{\beta_m(t)} + \sum_{i=1}^{m}\exp\left(-\frac{1}{2}\xi_i^2\right)} d\xi_1\cdots d\xi_m \end{aligned} \tag{21}$$

After noting that

$$\beta_m(t)\frac{\mathcal{B}(m)}{v_g^m}\sigma^m 2^m m = \frac{2}{\sqrt{2\pi}}P_d(t)\mathcal{B}(m-1)\frac{1}{g^{m-1}} \tag{22}$$

we finally get the information reduction factor expressed with binomial distribution given in (5).

References

[1] T. Kirubarajan, Y. Bar-Shalom, and Y. Wang, "Passive ranging of a low observable ballistic missile in a gravitational field," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 481–494, Apr 2001.

[2] C. Jauffret and Y. Bar-Shalom, "Track formation with bearing and frequency measurements in clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 26, no. 6, pp. 999–1010, Nov 1990.

[3] Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association*. Academic Press, 1988.

[4] T. Fortmann, Y. Bar-Shalom, M. Scheffe, and S. Gelfand, "Detection thresholds for tracking in clutter–A connection between estimation and signal processing," *IEEE Transactions on Automatic Control*, vol. 30, no. 3, pp. 221–229, Mar. 1985.

[5] X. R. Li and Y. Bar-Shalom, "Stability evaluation and track life of the PDAF for tracking in clutter," *IEEE Transactions on Automatic Control*, vol. 36, no. 5, pp. 588–602, May 1991.

[6] J. M. Mendel, *Lessons in estimation theory for signal processing, communications, and control*. Pearson Education, 1995.

[7] C. Jauffret, "Trajectographie passive, observabilité et prise en compte des fausses alarmes," Ph.D. dissertation, Université de Toulon, 1993.

[8] R. W. Sittler, "An optimal data association problem in surveillance theory," *IEEE Transactions on Military Electronics*, vol. 8, no. 2, pp. 125–139, Apr. 1964.

[9] D. Rife and R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *IEEE Transactions on Information Theory*, vol. 20, no. 5, pp. 591–598, Sep. 1974.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice-Hall, 1998.

[11] C. Robert, *L'analyse statistique bayésienne*. Economica, 1992.