Edinburgh Research Explorer

Maximum likelihood signal parameter estimation via track before detect

Citation for published version:

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Maximum Likelihood Signal Parameter Estimation via Track Before Detect

Murat Úney and Bernard Mulgrew
Institute for Digital Communications, School of Engineering,
The University of Edinburgh,
EH9 3JL, Edinburgh, UK
Emails: {M.Uney, B.Mulgrew}@ed.ac.uk

Abstract—In this work, we consider the front-end processing for an active sensor. We are interested in estimating signal amplitude and noise power based on the outputs from filters that match transmitted waveforms at different ranges and bearing angles. These parameters identify the distributions in, for example, likelihood ratio tests used by detection algorithms and characterise the probability of detection and false alarm rates. Because they are observed through measurements induced by a (hidden) target process, the associated parameter likelihood has a time recursive structure which involves estimation of the target state based on the filter outputs. We use a track-before-detect scheme for maintaining a Bernoulli target model and updating the parameter likelihood. We use a maximum likelihood strategy and demonstrate the efficacy of the proposed approach with an example.

I. INTRODUCTION

Active sensors send energy packets towards a surveillance region in order to locate objects within from the reflections. For example, radars transmit radio frequency (RF) electromagnetic (EM) pulses and locate reflectors by searching for the pulse waveform in the spatio-temporal energy content of the received signal. This search is often performed using the matched filtering technique in which the received signal is projected onto versions of the transmitted waveform shifted so as to encode the desired reflector locations [1]. The more the energy a projection has, the more likely that it is due to the presence of a reflector.

In this work, we are interested in estimating parameters related to the signal at the output of the matched filters (MF) of a radar. This signal is composed of a distorted version of the waveform auto-correlation function in the presence of a reflector and additive thermal noise. Detection algorithms aim to decide on the existence of an object based on these outputs sampled at selected time instances so as to give the energy of the aforementioned projected signal [2]. These decisions are characterised by a probability of detection and a probability of false alarm which can be found given the energy of the reflected pulse at the receiver front-end $E$, and, the noise power $\beta^2$ (or, the standard deviation of the noise process). These parameters also determine the signal-to-noise ratio which can be used to characterise the expected accuracy in further levels of processing [1].

Often only the noise power $\beta^2$ is estimated using spatial windows over one snapshot of the outputs from MFs associated with different range-bearing bins. These estimates are then used as design parameters in constant-false-alarm rate detection algorithms [3]. However, this approach is prone to errors due to incorrect identification of bins that contain only noise or signal-and-noise as a result of that this identification task requires the temporal information in the received signal which is ignored. Estimation of the signal energy $E$ requires collection of temporal samples, as well. One way of doing this is to replace the MF bank in the baseband processing chain with iterative processing algorithms preferably working with high sampling rates (e.g., [4], [5]). This places requirements on the hardware architecture that are hard to satisfy in practice.

In this work, we use multiple snapshots from the MF bank collected across a time interval and perform spatio-temporal processing to jointly estimate $E$ and $\beta^2$. We treat the problem as a parameter estimation problem in state space models. This allows us to integrate all the information in the measurements in a single likelihood function for both $E$ and $\beta^2$. Such likelihood functions require the state distribution of the underlying (target) process given the measurement history, which can be found by the prediction stage of Bayesian filtering (or, tracking) recursions [6]. As the measurements are the MF outputs (as opposed to the outputs from a detection algorithm as in widely studied tracking scenarios), the corresponding recursions describe a track-before-detect algorithm.

In this framework, we derive explicit formulae for the parameter likelihood and its score function, i.e., the log-likelihood gradient. We use a maximum likelihood (ML) approach to design an unbiased and minimum variance estimator. In particular, we maximise the log-likelihood by using a coordinate ascent algorithm [7] in which we select the directions of increase based on the gradient and perform (golden section) line search along these directions.

This article is outlined as follows: We give the ML problem definition in Sec. II and detail the parameter likelihood in Sec. III. In Sec. IV, we derive the gradient of the objective function and detail an iterative optimisation procedure. We demonstrate the proposed approach with an example in Sec. V.

II. PROBLEM DEFINITION

We consider a pulse transmitted towards a surveillance region which gets reflected if it interacts with an object at state
\[ x = [x_1^T, x_2^T]^T \] where \( x_1 \) is the location, \( x_2 = \dot{x}_2 \) is the velocity of the object and \((\cdot)^T\) denotes the transpose of a vector. These reflections are sought in the the spatio-temporal energy content of the received signal by matched filtering [1]. Typically, the filter output is sampled with a period of the pulse length so as to compute the correlation of the transmitted waveforms with the received signal corresponding to the \( i \)th range bin of width \( \Delta r \) and \( j \)th bearing bin of width \( \Delta \phi \). Therefore, at time step \( k \), the filter output for the bin \((i, j)\) is given by

\[
r_k(i, j) = \langle \mathbf{m}(i, j), \mathbf{r} \rangle
\] (1)

where \( \mathbf{r} \) is the (complex) received signal and \( \mathbf{m}(i, j) \) represents the transmitted waveform shifted to the \((i, j)\)th bin.

Let \( E_i \) represent the energy of the transmitted pulse, i.e., \( E_i = \mathbf{m}^H \mathbf{m} \) where \((\cdot)^H\) denotes the Hermitian transpose. In the presence of a reflecting object at state \( x_k \), the inner product above leads to

\[
r_k(i, j) = E_i e_j \theta_k h_{i,j}(x_k) + n_k
\] (2)

where \( n_k \sim \mathcal{N}(0; 0, \beta^2) \) is circularly symmetric complex Gaussian noise with complex power \( \beta^2 \). \( h_{i,j}(x_k) \) specifies the ratio of \( E \) that has been reflected from the \((i, j)\)th bin with (an unknown) phase \( \theta_k \sim \mathcal{U}(0, 2\pi] \).

In this work, we assume a sensor resolution such that an object in the surveillance region affects only a single range-bearing bin, i.e.,

\[
h_{i,j}(x_k) = H \delta \mathcal{C}(x_k, i, j)
\] (3)

where \( \delta \) is the Kronecker’s delta function, \( C : \mathcal{X} \rightarrow \mathcal{M} \times \mathcal{N} \) maps object states to range-bearing bins and \( H \) is the reflection coefficient.

We consider (2) and (3), and, are interested in estimating the received signal energy \( E = E_i H \) and, the noise power \( \beta^2 \). These signal parameters determine the signal-to-noise ratio at the matched filter by

\[
SNR = 10 \log_{10} \frac{E^2}{\beta^2},
\] (4)

and are also required to compute false alarm rates and object detection probabilities of threshold rules [1].

We treat these parameters as (non-random) unknown constants and consider an ML solution. We now specify the arguments of the likelihood function: The reflection phase \( \theta_k \) in (2) models the ambiguity related to the exact position of the reflector within the \((i, j)\)th bin which cannot be mitigated. The modulus of (2), i.e.,

\[
z_k(i, j) \triangleq |r_k(i, j)|
\]

neglects the phase and is a sufficient statistic when testing whether \( r_k(i, j) \) is induced by a reflector at state \( x_k \) or noise alone. Therefore, we treat the intensity map given by \( z_k(i, j) \) as measurements based on which the ML estimation will be relying upon.

Let us denote with \( z_k \) the concatenation of \( z_k(i, j) \)s for all range-bearing bins, i.e., the intensity map. We would like to base the parameter likelihood on all measurements covering time step 1 through \( k \). Therefore, the ML estimation problem can be formulated as

\[
\hat{(E, \beta^2)} = \arg \max_{E, \beta^2} \log l(z_1, \ldots, z_k \mid E, \beta^2),
\] (5)

the computation of which will be described next.

### III. THE SIGNAL PARAMETER LIKELIHOOD

Let us represent with a random set \( X_k \), the events that there exists a reflector with state \( x_k \) and, none, in which cases \( X_k = \{x_k\} \) and \( X_k = \emptyset \), respectively. \( X_k \) is referred to as a Bernoulli random finite set (RFS) [8]. The likelihoods for \( z_k(i, j) \) given the signal parameters to be estimated \( E, \beta^2 \) for the cases that \( X_k = \{x_k\} \) and \( X_k = \emptyset \) are well known results in the literature: After the uniformly distributed phase is marginalised-out, the modulus \( z_k(i, j) \) is distributed a Rician distribution, i.e.,

\[
l_1(z_k(i, j) \mid x_k \mid E, \beta^2) \triangleq p(z_k(i, j) \mid x_k = \{x_k\} \mid E, \beta^2),
\]

\[
= \frac{2z_k(i, j)}{\beta^2} \exp \left( -\frac{z_k(i, j)^2 + E^2}{\beta^2} \right) \mathcal{I}_0 \left( \frac{2z_k(i, j)E}{\beta^2} \right),
\] (6)

if \( X_k = \{x_k\} \) and \( x_k \in \mathcal{C}^{-1}(i, j) \), where \( \mathcal{I}_0 \) is the zero order modified Bessel function of the first kind. Otherwise, \( z_k(i, j) \) follows a Rayleigh law given by [1, Chp.6]

\[
l_0(z_k(i, j) \mid \emptyset \mid E, \beta^2) \triangleq p(z_k(i, j) \mid x_k = \emptyset \mid E, \beta^2),
\]

\[
= \frac{2z_k(i, j)}{\beta^2} \exp \left( -\frac{z_k(i, j)^2}{\beta^2} \right),
\] (7)

Let us define the intensity map \( z_k \) related quantities:

\[
\Lambda(z_k \mid x_k \mid E, \beta^2) \triangleq l_1(z_k(C(x_k))) \mid x_k \mid E, \beta^2
\]

\[
= \prod_{(i, j) \in \mathcal{C}(x_k)} l_0(z_k(i, j) \mid \emptyset \mid E, \beta^2),
\] (8)

where \( C(x_k) \) denotes the set of range-bearing bins complementing \( C(x_k) \), and, \( l_0 \) and \( l_1 \) are given in (6) and (7), respectively. Similarly

\[
\Lambda(z_k \mid \emptyset \mid E, \beta^2) \triangleq \prod_{i,j} l_0(z_k(i, j) \mid \emptyset \mid E, \beta^2)
\] (9)

where the product is over all the range-bearing bins.

We assume that the noise processes for different bins and time steps are independent given the state of the object process \( X_k \). Hence,

\[
p(z_k \mid X_k \mid E, \beta^2) = \begin{cases} \Lambda(z_k \mid x_k \mid E, \beta^2), & \text{if } X_k = \{x_k\} \\ \Lambda(z_k \mid \emptyset \mid E, \beta^2), & \text{if } X_k = \emptyset \end{cases}
\] (10)

We would like to compute the parameter likelihood based on all measurements covering time step 1 through \( k \) which can be found as

\[
l(z_1, \ldots, z_k \mid E, \beta^2) = \prod_{t=1}^{k} p(z_t \mid z_{1:t-1} \mid E, \beta^2)
\]

\[
= l(z_1, \ldots, z_{k-1} \mid E, \beta^2)
\]

\[
\times p(z_k \mid z_{1:k-1} \mid E, \beta^2)
\]
after using the chain rule of probabilities in the first line. The recursive structure revealed in the second line is typical to parameter estimation problems in state space models [9], [10], in which the update term is found by marginalising-out the underlying process $X_k$, i.e.,

$$p(z_{k} | z_{1:k-1}, E, \beta^2) = \int p(z_{k} | X_k, E, \beta^2)p(X_k | z_{1:k-1}, E, \beta^2) \, dX_k$$ (11)

where the right hand side is a set integral [8, Chp.11] as $X_k$ is a set random variable. The first term inside the integral is the likelihood at $k$ given by (10), and, the second term is the prediction of $X_k$ based on the previous measurements.

The objective function for the ML solution in (5), hence, is given by

$$J(E, \beta^2; z_{1:k}) = J(E, \beta^2; z_{1:k-1}) + \log p(z_{k} | z_{1:k-1}, E, \beta^2).$$ (12)

Let us now consider the likelihood update term in (11). The computation of the predictive term inside the integral is detailed later in Section III-A. For our discussion on the likelihood update term, suppose that it is given by

$$p(X_k | z_{1:k-1}, E, \beta^2) = \begin{cases} r_{k|k-1} s_{k|k-1}(x_k), & \text{if } X_k = \{ x_k \} \\ 1 - r_{k|k-1}, & \text{if } X_k = \emptyset. \end{cases}$$ (13)

Then, (11) expands using the set integration rule in [8, Chp.11] with a Bernoulli RFS characterised by (13) and the likelihood in (10) as

$$p(z_{k} | z_{1:k-1}, E, \beta^2) = p(z_{k} | \emptyset, E, \beta^2)p(\emptyset | z_{1:k-1}, E, \beta^2) + \int p(z_{k} | X_k = \{ x_k \}, E, \beta^2)p(X_k = \{ x_k \} | z_{1:k-1}, E, \beta^2) \, dX_k$$

$$= (1 - r_{k|k-1}) \Lambda(z_k | \beta^2) + r_{k|k-1} \int \Lambda(z_k | x_k, E, \beta^2) s_{k|k-1}(x_k) \, dX_k.$$ (14)

After dividing both parts of the equation above by $\Lambda(z_k | \beta^2)$, it can easily be shown that

$$\log p(z_{k} | z_{1:k-1}, E, \beta^2) = \log \Lambda(z_k | \beta^2) + \log \left( 1 - r_{k|k-1} + r_{k|k-1} \int \frac{l_1(z_k | C(x_k)) |x_k; E, \beta^2)}{L_0(z_k | C(x_k)) |x_k; E, \beta^2)} s_{k|k-1}(x_k) \, dX_k \right)$$ (15)

Therefore, the ML objective can be recursively evaluated using (15) in (12). Next, we discuss the computation of the predictive state distribution in (13) using Bernoulli track before detect.

A. Bernoulli track before detect

Recursive updating of a Bernoulli object model using measurements with likelihood models in the form of (10) can be carried out using the Bayesian recursive filtering principles [11], [12]. An important component of filtering with RFS models is the object appearance, or, birth model. The probability of object appearance at time $k$ is given by $p_b$, and the state of this object is distributed as $b(x)$. The probability of an object that existed at $k-1$ continuing to exist at time $k$ is given by $p_s$. A Bernoulli model at time $k-1$ characterised by $(r_{k-1}, s_{k-1}(x_{k-1}))$, then, leads to the following prediction:

$$r_{k|k-1} = p_b (1 - r_{k-1}) + r_{k-1} p_s$$

$$s_{k|k-1}(x_k) = \frac{p_b (1 - r_{k-1}) b(x)}{r_{k|k-1}} + \frac{r_{k-1} p_s}{r_{k|k-1}} \int s_{k-1}(x_{k-1}) s_{k|k-1}(x_{k-1}) \, dx_{k-1}$$ (16)

where $p_{k|k-1}$ is the state transition density [11]. Upon receiving $z_k$, the posterior model is given by [11]

$$r_k = \frac{r_{k|k-1} g_k(x_k | E, \beta^2) \, dx_{k}}{1 - r_{k|k-1} + r_{k|k-1} g_k(x_k | E, \beta^2) \, dx_{k}}$$

$$s_k(x_k) = \frac{g_k(x_k | E, \beta^2)}{l_1(z_k | C(x_k)) |x_k; E, \beta^2)} s_{k|k-1}(x_k).$$ (17)

As a result, the predictive term in (13) required for the parameter likelihood is found by iterating prediction and update cycles using (16) and (17).

The object birth model we use is selected so as to have a uniform distribution in the location component which is nonzero in the sensor field of view, and, a uniform distribution in the velocity component which is nonzero if the speed is between selected minimum and maximum values:

$$b([x_v^T, x_v^T]^{T}) = U_{FOV}(x) \, |x_v|_{min} \leq x_v \leq |x_v|_{max}.$$ (18)

IV. MAXIMUM LIKELIHOOD SIGNAL AMPLITUDE AND NOISE POWER ESTIMATION

The objective function of the ML solution in (5) is not straightforward to maximise partly because, in practice, only a noisy approximation of it can be obtained via particle methods, and, the term due to $E$ in the right hand side (RHS) of (15) is dominated by the influence of $\beta^2$ through the first term (see, e.g., the example in Section V).

Our ML realisation strategy is to apply coordinate ascent along the projections of the log-likelihood gradient. This can be viewed as a subgradient approach with the difference that we do not explicitly select step sizes to move along the selected subgradient direction. Instead, we perform (golden ratio) line search along the subgradient direction. The gradient equals the sum of the gradients of the log update term in (15) over time. Note that, the first term in the RHS of (15) is independent of $E$, hence,

$$\frac{\partial \log p(z_{k} | z_{1:k-1}, E, \beta^2)}{\partial E} = \frac{r_{k|k-1} \frac{\partial \log l_1}{\partial E} s_{k|k-1}}{1 - r_{k|k-1} + r_{k|k-1} \int l_1^{R}(z_k | C(x_k)) |x_k; E, \beta^2)}$$

$$= \frac{r_{k|k-1} \frac{\partial \log l_1}{\partial E}}{1 - r_{k|k-1} + r_{k|k-1} g_k}$$ (19)

where the first line follows after differentiating the term inside the integral and using the identity $\partial l_1/\partial E = l_1 \times \partial \log l_1/\partial E$. The second line is obtained through the definition of $g_k$ in (17).
The partial derivative of \( \log l_1 \) with respect to \( E \) can easily be found [13] as
\[
\frac{\partial \log l_1}{\partial E} = \frac{2E}{\beta^2} + \frac{2z_k(C(x_k))}{\beta^2} \frac{I_1(2z_k(C(x_k))E/\beta^2)}{I_0(2z_k(C(x_k))/\beta^2)}
\]
where \( I_0 \) and \( I_1 \) are the modified Bessel functions of the first kind of order zero and one, respectively. After substituting (20) in (19), we get
\[
\frac{\partial \log p(z_k|z_{1:k-1}, E, \beta^2)}{\partial E} = -\frac{2Er_k}{\beta^2} + \frac{2r_{k|k-1} \left\{ \frac{z_k}{2} \frac{I_1(z_kE/\beta^2)}{I_0(2z_kE/\beta^2)} \right\}}{1 - r_{k|k-1} + r_{k|k-1} \left\{ \frac{z_k}{2} \frac{I_1(z_kE/\beta^2)}{I_0(2z_kE/\beta^2)} \right\}}
\]
where \( r_k \) and \( g_k \) are given in (17).

Next, we consider the partial derivative of the log update term with respect to \( \beta^2 \):
\[
\frac{\partial \log p(z_k|z_{1:k-1}, E, \beta^2)}{\partial \beta^2} = \sum_{i,j} \frac{\partial \log l_0}{\partial \beta^2} + \frac{r_{k|k-1} \left\{ \frac{z_k}{2} \frac{I_1(z_kE/\beta^2)}{I_0(2z_kE/\beta^2)} \right\}}{1 - r_{k|k-1} + r_{k|k-1} \left\{ \frac{z_k}{2} \frac{I_1(z_kE/\beta^2)}{I_0(2z_kE/\beta^2)} \right\}} g_k
\]
The first partial derivative inside the integral above can be found [13] as
\[
\frac{\partial \log l_1}{\partial \beta^2} = -\frac{1}{\beta^2} \left( 1 - \frac{z_k^2 + E^2}{\beta^2} + \frac{2z_k E I_1(2z_kE/\beta^2)}{\beta^2} \right) \frac{I_0(2z_kE/\beta^2)}{I_0(2z_kE/\beta^2)}
\]
and the derivative of log-noise term with respect to the noise power is given by
\[
\frac{\partial \log l_0}{\partial \beta^2} = -\frac{1}{\beta^2} + \frac{z_k^2}{\beta^4}
\]
Therefore, the first term inside the integral in (22) is given by
\[
\frac{\partial \log l_1}{\partial \beta^2} \frac{\partial \log l_0}{\partial \beta^2} = \frac{E^2}{\beta^4} - \frac{2Ez_k I_1}{I_0}
\]
(23)

As a result, the gradient of the log-likelihood function in (12) can be computed using (19)–(23) in the recursive form given by
\[
\nabla J(E, \beta^2; z_{1:k}) = \nabla J(E, \beta^2; z_{1:k-1}) + \left[ \frac{\partial \log p(z_k|z_{1:k-1}, E, \beta^2)}{\partial \beta^2} \frac{\partial \log p(z_k|z_{1:k-1}, E, \beta^2)}{\partial \beta^2} \right]_{E, \beta^2} \nabla J(E, \beta^2; z_{1:k}) + \left[ \frac{\partial \log p(z_k|z_{1:k-1}, E, \beta^2)}{\partial \beta^2} \frac{\partial \log p(z_k|z_{1:k-1}, E, \beta^2)}{\partial \beta^2} \right]_{E, \beta^2} \nabla J(E, \beta^2; z_{1:k})
\]
(24)

In order to maximise \( J \), we adopt a coordinate ascent approach: Starting from an initial point \((E_0, \beta^2_0)\), at each iteration \(m\), we find the gradient vector (24) and select a line search direction as follows:
\[
d_m = \nabla J(E, \beta^2; z_{1:k})^T e_m
\]
(25)
\[
e_m = \begin{cases} \{0, 1\}^T, & \text{if } m = 0, 2, 4, \ldots \\ \{1, 0\}^T, & \text{if } m = 1, 3, 4, \ldots \end{cases}
\]
Then, we solve the following one dimensional problem
\[
(E_{m+1}, \beta^2_{m+1}) = \arg \min_{\lambda \in [0, \lambda_{\max}]} J([E_m, \beta^2_m]^T + \lambda d_m; z_{1:k})
\]
(26)
using golden section search [14].
Correspondingly, the Cramer-Rao lower bound (CRLB) for $E$ is higher and estimates of $E$ will be less accurate [16, App. 8A]. For a closer look to the log-likelihood objective function, we provide the normalised profiles of the surface along $E$ and $\beta^2$ axis in Fig. 1(b) and (c), respectively. Note that, for almost all values of $E$ in the grid, it is possible to estimate $\beta^2$ through maximising the objective along $\beta^2$ direction.

We repeat the proposed ML scheme for 200 Monte Carlo realisations of this scenario. The iterations are depicted in Fig. 2, in which the resulting estimates are shown by red crosses. The average number of iterations is 3.6231. The average of $(\hat{E}, \beta^2)$ is $(1.972, 0.104)$, which is very close to the true value of $(2, 0.1)$. This indicates that the optimisation scheme is approximately unbiased. The variance of the estimation scheme as measured by the empirical average of the squared errors for $\hat{E}$ and $\beta^2$ are found as $1.78 \times 10^{-2}$ and $1.7 \times 10^{-5}$, respectively. The variance of $\hat{E}$ is much higher than that for $\beta^2$, as expected.

VI. CONCLUSIONS AND FUTURE WORK

In this work, we proposed a ML scheme for jointly estimating the signal amplitude and noise power using the matched filter outputs of an active sensor. This likelihood involves track-before-detect in order predict the underlying target process which is modelled by a Bernoulli RFS. We derived explicit formulae for the score function and proposed an iterative maximisation procedure using a coordinate ascent approach.

It is possible to improve the optimisation step of the proposed algorithm. For example, Newtonian methods can be used after finding the Hessian of the log-likelihood. Such an approach would remove the need for line search and potentially provide more accurate estimates for fewer number of iterations and objective evaluations.

There is a tradeoff between the observation length and the accuracy in which $E$ can be estimated (see, e.g., [17]). The relations between the Hessian, Fisher information and the associated Cramer-Rao lower bound can be explored in order to investigate this tradeoff. Another possible extension of this work is to accommodate multi-Bernoulli models in order to handle multiple moving objects via track-before-detect [18] and other Swerling target types [1].

ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) grants EP/J015180/1 and EP/K014277/1, and the MOD University Defence Research Collaboration in Signal Processing.

REFERENCES