

# Matrix CRLB Scaling Due to Measurements of Uncertain Origin

Ruixin Niu, Peter Willett, *Senior Member, IEEE*, and Yaakov Bar-Shalom, *Fellow, IEEE*

**Abstract**—In many estimation situations, measurements are of uncertain origin. This is best exemplified by the target-tracking situation in which at each scan (of a radar, sonar, or electro-optical sensor), a number of measurements are obtained, and it is not known which, if any, of these is target originated. The source of extraneous measurements can be false alarms—especially in low-SNR situations that force the detector at the end of the signal processing chain to operate with a reduced threshold—or spurious targets. In several earlier papers, the surprising observation was made that the Cramér–Rao lower bound (CRLB) for the estimation of a fixed parameter vector (e.g., initial position and velocity) that characterizes the target motion, for the special case of multidimensional measurements in the presence of additive white Gaussian noise, is simply a multiple of that for the case with no uncertainty. That is, there is a scalar *information-reduction factor*; this is particularly useful as it allows comparison in terms of a scalar. In this paper, we explore this result to determine how wide the class of such problems is. It turns out to include many non-Gaussian situations. Simulations corroborate the analysis.

## I. INTRODUCTION

### A. Multiparameter CRLB

IN many estimation problems, one is faced with the problem that one's data is of uncertain origin. For example, in the target-tracking situation, the data set may consist of "hits" (threshold exceedances) indicative of a target's presence in a particular location at a particular time; however, hits may be of spurious origin (that is, they are *false alarms*), and indeed, it may be that the true target is unrepresented in the data set at the time in question (a *missed detection*). At issue is how well one can estimate a fixed parameter or parameter vector that characterizes target motion given such uncertainty.

The measure in which we are interested is, not unnaturally, the mean-squared error (MSE) of the estimate. The classical result for this is known as the Cramér–Rao lower bound (CRLB) (e.g., [12], [15]). Let us assume access to an observation  $\mathbf{Z}$  that has probability density function (pdf)  $p(\mathbf{Z}; \mathbf{x})$ , meaning that the pdf depends on a parameter vector  $\mathbf{x}$  to be estimated. Let us assume the existence of an *unbiased* estimator  $\hat{\mathbf{x}}(\mathbf{Z})$ , meaning that we have

$$\mathcal{E}\{\hat{\mathbf{x}}(\mathbf{Z})\} = \mathbf{x} \quad (1)$$

Manuscript received May 21, 1999; revised April 3, 2001. This work was supported by the ONR/BMDO-IST under Contract N00014-00-1-0740, by the AFOSR under Contract F49620-00-1-0052, and by the ONR under Contract N00014-97-1-0502. The associate editor coordinating the review of this paper and approving it for publication was Prof. S. M. Jesus.

The authors are with the Department of Electrical and Systems Engineering, University of Connecticut, Storrs CT 06269-2157 USA (e-mail: willett@engr.uconn.edu).

Publisher Item Identifier S 1053-587X(01)05176-5.

in which  $\mathcal{E}$  denotes expectation parametrized by  $\mathbf{x}$ . Then, under fairly broad regularity conditions, the CRLB has it that

$$\mathcal{E}\{[\hat{\mathbf{x}}(\mathbf{Z}) - \mathbf{x}][\hat{\mathbf{x}}(\mathbf{Z}) - \mathbf{x}]^T\} \geq \mathbf{J}^{-1} \quad (2)$$

in which

$$\mathbf{J} \equiv \mathcal{E}\{[\nabla_{\mathbf{x}} \log(p(\mathbf{Z}; \mathbf{x}))][\nabla_{\mathbf{x}} \log(p(\mathbf{Z}; \mathbf{x}))]^T\} \quad (3)$$

is Fisher's information matrix. Again, under broad regularity conditions, if a maximum-likelihood estimator (MLE) for  $\mathbf{x}$  exists, then it achieves the CRLB asymptotically.

### B. Measurements of Uncertain Origin

The general multiparameter CRLB is fairly standard; let us now turn to the case of particular interest: that of measurement origin uncertainty. Without loss of generality, we define the aggregate observation from time 1 to time  $T$

$$\mathbf{Z} = \{\mathbf{Z}(1), \mathbf{Z}(2), \dots, \mathbf{Z}(T)\} \quad (4)$$

in which the  $t$ th observation is

$$\mathbf{Z}(t) = \{\mathbf{z}_i(t)\}_{i=1}^{m_t} \quad (5)$$

meaning that there are  $m_t$  individual observations that comprise it.

In the target tracking scenario,  $\mathbf{Z}(t)$  is comprised of all observations collected at time  $t$ , and these observations can be all false alarms (the detection from the target has been missed) or can contain exactly one true detection and  $(m_t - 1)$  false alarms. In this situation, it is common to assume that  $p_0(\cdot)$  is uniform over the observation volume (or gated volume), i.e.,  $p_0(\cdot)$  is independent of  $\mathbf{x}$ , and that the number of false-alarms is Poisson, meaning that (valid for  $m_t \geq 1$ )

$$\begin{aligned} p(m_t) &= (1 - P_d) \frac{(\lambda V)^{m_t} e^{-\lambda V}}{m_t!} + P_d \frac{(\lambda V)^{(m_t-1)} e^{-\lambda V}}{(m_t - 1)!} \\ \epsilon(m_t) &= \frac{P_d}{p(m_t)} \frac{(\lambda V)^{(m_t-1)} e^{-\lambda V}}{(m_t - 1)!} \end{aligned} \quad (6)$$

are, respectively, the *a priori* probability that there are  $m_t \in \{0, \infty\}$  observations at time  $t$  and the probability that one measurement is target generated, given that there are  $m_t$  measurements at time  $t$ . In the above

- $P_d$  probability of detection;
- $\lambda$  average number of false alarms per unit observation volume;
- $V$  observation (surveillance) volume.

Although  $m_t$  is, of course, known to the estimator, to compute the MSE, it is necessary to average over the possible values of  $m_t$ . We further assume that

$$\begin{aligned}
p(\mathbf{Z}; \mathbf{x}) &= \prod_{t=1}^T p(\mathbf{Z}(t); \mathbf{x}) \\
&= \prod_{t=1}^T \left( \sum_{m_t=0}^{\infty} p(m_t) [p(\mathbf{Z}(t)|m_t, \text{all false}; \mathbf{x}) \right. \\
&\quad \cdot p(\text{all false}|m_t) \\
&\quad \left. + p(\mathbf{Z}(t)|m_t, \text{one true detection}; \mathbf{x}) \right. \\
&\quad \left. \cdot p(\text{one true detection}|m_t) \right) \\
&= \prod_{t=1}^T \left( \sum_{m_t=0}^{\infty} p(m_t) [(1 - \epsilon(m_t)) \right. \\
&\quad \cdot p(\{\mathbf{z}_j(t)\}_{j=1}^{m_t} | \text{all false}; \mathbf{x}) \\
&\quad \left. + \sum_{i=1}^{m_t} \frac{\epsilon(m_t)}{m_t} p(\{\mathbf{z}_j(t)\}_{j=1}^{m_t} | \mathbf{z}_i(t) \text{ from target} \right. \\
&\quad \left. \text{all others clutter}; \mathbf{x}) \right] \\
&= \prod_{t=1}^T \left( \sum_{m_t=0}^{\infty} p(m_t) \left[ \prod_{j=1}^{m_t} p_0(\mathbf{z}_j(t)) \right] \right. \\
&\quad \left. \cdot \left[ (1 - \epsilon(m_t)) + \frac{\epsilon(m_t)}{m_t} \sum_{i=1}^{m_t} \frac{p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))}{p_0(\mathbf{z}_i(t))} \right] \right) \quad (7)
\end{aligned}$$

which requires some explanation. First, we assume that conditioned on  $\mathbf{x}$ ,  $\{\mathbf{Z}(1), \dots, \mathbf{Z}(T)\}$  are independent. Second, we have assumed that given  $m_t$  observations that comprise  $\mathbf{Z}(t)$ , any of  $m_t + 1$  events are possible: Either all  $m_t$  observations  $\{\mathbf{z}_i(t)\}_{i=1}^{m_t}$  are distributed according to  $p_0(\cdot)$  (which does not depend on  $\mathbf{x}$ ), or *exactly one* of these is distributed according to  $p_1(\cdot)$  (which *does* depend on  $\mathbf{x}$ ), whereas the rest remain distributed according to  $p_0(\cdot)$ , with each of these  $m_t$  events equally likely.<sup>1</sup> Regardless of which event is true, all  $m_t$  observations are independent given that event. Third, we have written the dependence of the target-generated observations on the unknown parameter  $\mathbf{x}$  as  $\mu_t(\mathbf{x})$ ; although variations are possible [7], this model is most appropriate for a deterministic track (such as straight-line or ballistic motion) in which the target's location is a function of a few "initial" parameters such as position and velocity. Finally, it is important that the dependence of the observation  $\mathbf{z}_i(t)$  on  $\mu_t(\mathbf{x})$  is as a direct translation (mean-shift).

### C. Information-Reduction Factor

In order to apply (7) to (2), we first note that due to independence among  $\{\mathbf{Z}(t)\}$  and to the logarithm, we have

$$\mathbf{J} = \sum_{t=1}^T \mathbf{J}_t \quad (8)$$

<sup>1</sup>The first "all-false" event has prior probability  $(1 - \epsilon(m_t))$ , whereas each of the others ("one true and  $m_t - 1$  false") has probability  $\epsilon(m_t)/m_t$ .

in which

$$\mathbf{J}_t \equiv \mathcal{E}\{(\nabla_{\mathbf{x}} \log[p(\mathbf{Z}(t); \mathbf{x})])(\nabla_{\mathbf{x}} \log[p(\mathbf{Z}(t); \mathbf{x})])^T\}. \quad (9)$$

Further, since the number of observations  $m_t$  is known to the estimator, we also have

$$\begin{aligned}
\mathbf{J}_t &= \sum_{m_t=1}^{\infty} p(m_t) \mathcal{E}\{(\nabla_{\mathbf{x}} \log[p(\mathbf{Z}(t); \mathbf{x})]) \\
&\quad \cdot (\nabla_{\mathbf{x}} \log[p(\mathbf{Z}(t); \mathbf{x})])^T | m_t\} \\
&= \sum_{m_t=1}^{\infty} p(m_t) \mathcal{E}\{(\nabla_{\mathbf{x}} \log[p(\{\mathbf{z}_i(t)\}_{i=1}^{m_t}; \mathbf{x})]) \\
&\quad \cdot (\nabla_{\mathbf{x}} \log[p(\{\mathbf{z}_i(t)\}_{i=1}^{m_t}; \mathbf{x})])^T\} \\
&= \sum_{m_t=1}^{\infty} p(m_t) \mathbf{J}_t(m_t) \quad (10)
\end{aligned}$$

in which one should note that the  $m_t = 0$  term, corresponding to the lack of a target-generated measurement, naturally has no contribution. In the above equation

$$\begin{aligned}
\mathbf{J}_t(m_t) &= \mathcal{E}\{(\nabla_{\mathbf{x}} \log[p(\{\mathbf{z}_i(t)\}_{i=1}^{m_t}; \mathbf{x})]) \\
&\quad \cdot (\nabla_{\mathbf{x}} \log[p(\{\mathbf{z}_i(t)\}_{i=1}^{m_t}; \mathbf{x})])^T\} \quad (11)
\end{aligned}$$

reflects estimation efficiency at a particular time  $t$  for a particular number of observations  $m_t$ . We have the probability density function of the  $m_t$  observations at time  $t$ , parametrized by  $\mathbf{x}$

$$\begin{aligned}
p(\{\mathbf{z}_i(t)\}_{i=1}^{m_t}; \mathbf{x}) &= \left[ \prod_{i=1}^{m_t} p_0(\mathbf{z}_i(t)) \right] \left[ (1 - \epsilon(m_t)) \right. \\
&\quad \left. + \frac{\epsilon(m_t)}{m_t} \sum_{i=1}^{m_t} \frac{p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))}{p_0(\mathbf{z}_i(t))} \right] \quad (12)
\end{aligned}$$

In [9] and [11], a surprising result was obtained. Under the target-tracking assumptions above and further assuming a Gaussian model for the true-observation pdf  $p_1(\cdot)$ , it was shown that

$$\mathbf{J}_t(m_t) = \kappa(P_d, \lambda V) \mathbf{J}_t^0 \quad (13)$$

where  $\kappa$  is a scalar less than unity (in the literature usually denoted " $q_2$ " for historical reasons), and

$$\begin{aligned}
\mathbf{J}_t^0 &\equiv \mathcal{E}\{(\nabla_{\mathbf{x}} \log[p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))]) \\
&\quad \cdot (\nabla_{\mathbf{x}} \log[p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))])^T\} \quad (14)
\end{aligned}$$

is Fisher's information matrix in scan  $t$  for the *measurement-certain* case. Assuming, therefore, that  $P_d$  does not vary with  $t$  (this is a reasonable approximation in the far-field tracking situation), it therefore follows that the total Fisher information matrix in the case of uncertain-origin measurements is

$$\mathbf{J} = \kappa \mathbf{J}^0 \quad (15)$$

where

$$\mathbf{J}^0 = \sum_{t=1}^T \mathbf{J}_t^0. \quad (16)$$

In other words, there is a scalar *information-reduction factor*  $\kappa$  (less than unity) in the proportionality to account for the estimation algorithm's need to judge which of its observations (if

any) are relevant and which are spurious.<sup>2</sup> This, in our opinion, is a remarkable result: The presence of spurious data in one's observation set affects observation efficiency, quantified by the total information in the observations—which is given by a matrix—in a *scalar* way. Thus, the presence of false measurements and, indeed, the possible absence of a target-originated measurement have an isotropic effect in the parameter space as far as the “existing” information about the parameter is concerned.

The task in the present paper is to expound on this: How general is (15)? In the following section, we develop some conditions that are sufficient. Specifically, we isolate attention to the satisfaction of a condition to be denoted A2) [the other condition A1) is trivial], which is a somewhat technical requirement on the observation distribution  $p_1(\cdot)$ . We further require that the “false-alarm” distribution  $p_0(\cdot)$  be either uniform B1), which is a sensible condition, or target-centered B2), which is an unlikely situation. We finally interpret A2) as C1), C2), or C3). The last is the most general of these and means that (15) is true for any  $p_1(\cdot)$  that has elements that are *iid* and symmetric or can be made so via a linear transformation.

A key benefit of (15) is in its distillation of the effect of measurement origin uncertainty<sup>3</sup> to a *scalar* information-reduction factor as compared with the origin-uncertainty-free case. Thus, in Section III, we explore the result for several non-Gaussian measurement distributions  $p_1(\cdot)$  with the goal of seeing their relative effects. Section IV corroborates our results via simulation, and in Section V, we summarize.

## II. SOME SUFFICIENT CONDITIONS FOR A MULTIPLICATIVE CRLB

To look for *necessary* conditions for (15) to be true is likely to be unrewarding since matrices can always add up in weird ways; we will instead develop as wide a class of *sufficient* situations as possible. From (8) and (11), the task is clear: We must

<sup>2</sup>Calculation of this proportionality constant is remarkably involved. Unfortunately, it arises from the need to evaluate a high-dimensional integral. However, the proportionality constant has been tabulated for a number of cases, and since it is sparsely parametrized, this is sufficient for many needs.

<sup>3</sup>It seems reasonable at this point to remind the ourselves that the “measurement origin uncertainty,” which is referred to in this paper, is not that caused by corrupting noise directly added to a quantity of interest, but rather by the appearance of spurious irrelevant measurements, by a lack of assurance that *any* measurement at a particular time is relevant, and overall by there being no labeling among the measurements.

characterize problems for which  $\mathbf{J}_t(m_t) \propto \mathbf{J}_t^0$  [see (13)], regardless of the number of observations  $m_t$  of the time index  $t$ .

We thus examine (12) and rewrite it as

$$p(\{\mathbf{z}_i(t)\}_{i=1}^{m_t}; \mathbf{x}) = \left[ \prod_{i=1}^{m_t} p_0(\mathbf{z}_i(t)) \right] \cdot \left[ (1 - \epsilon(m_t)) + \frac{\epsilon(m_t)}{m_t} \sum_{i=1}^{m_t} \frac{p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))}{p_0(\mathbf{z}_i(t))} \right].$$

Taking the gradient with respect to  $\mathbf{x}$  of the logarithm, we get

$$\mathcal{E}\{(\nabla_{\mathbf{x}} \log[p(\{\mathbf{z}_i(t)\}_{i=1}^{m_t}; \mathbf{x})]) (\nabla_{\mathbf{x}} \log[p(\{\mathbf{z}_i(t)\}_{i=1}^{m_t}; \mathbf{x})])^T\} = \mathbf{M}_t^T \mathbf{F}_t(m_t) \mathbf{M}_t \quad (17)$$

in which

$$\mathbf{M}_t = \begin{pmatrix} \frac{\partial \mu_t(\mathbf{x})_1}{\partial \mathbf{x}_1} & \frac{\partial \mu_t(\mathbf{x})_1}{\partial \mathbf{x}_2} & \dots & \frac{\partial \mu_t(\mathbf{x})_1}{\partial \mathbf{x}_{n_x}} \\ \frac{\partial \mu_t(\mathbf{x})_2}{\partial \mathbf{x}_1} & \frac{\partial \mu_t(\mathbf{x})_2}{\partial \mathbf{x}_2} & \dots & \frac{\partial \mu_t(\mathbf{x})_2}{\partial \mathbf{x}_{n_x}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mu_t(\mathbf{x})_{n_z}}{\partial \mathbf{x}_1} & \frac{\partial \mu_t(\mathbf{x})_{n_z}}{\partial \mathbf{x}_2} & \dots & \frac{\partial \mu_t(\mathbf{x})_{n_z}}{\partial \mathbf{x}_{n_x}} \end{pmatrix} \quad (18)$$

is the Jacobian matrix of  $\mu_t(\mathbf{x})$  (we assume that  $\mathbf{z}_i(t)$  and  $\mathbf{x}$  have respective dimensions  $n_z$  and  $n_x$ ) and in which we define (19), shown at the bottom of the page. We can thus write

$$\begin{aligned} \mathbf{J} &= \sum_{t=1}^T \mathbf{M}_t^T \left[ \sum_{m_t=1}^{\infty} p(m_t) \mathbf{F}_t(m_t) \right] \mathbf{M}_t \\ &= \sum_{t=1}^T \mathbf{M}_t^T \mathbf{F}_t \mathbf{M}_t \end{aligned} \quad (20)$$

where  $\mathbf{F}_t$  has been defined as the term in brackets. In addition to the obvious scalar case that  $n_x = 1$ , we immediately observe a pair of sufficient conditions under which (15) can be true.

A1) If  $n_z = 1$  and  $\mathbf{M}_t$  does not depend on  $t$ , then we have

$$\begin{aligned} \mathbf{J} &= \sum_{t=1}^T \mathbf{M}_t^T \left[ \sum_{m_t=1}^{\infty} p(m_t) \mathbf{F}_t(m_t) \right] \mathbf{M}_t \\ &= \left( \sum_{t=1}^T \left[ \sum_{m_t=1}^{\infty} p(m_t) \mathbf{F}_t(m_t) \right] \right) \mathbf{M}^T \mathbf{M} \\ &\propto \mathbf{J}^0 \end{aligned} \quad (21)$$

$$\begin{aligned} \mathbf{F}_t(m_t) &\equiv \mathcal{E} \left\{ \left( \frac{\left[ \prod_{i=1}^{m_t} p_0(\mathbf{z}_i(t)) \right] \left[ \frac{\epsilon(m_t)}{m_t} \sum_{i=1}^{m_t} \frac{\nabla_{\mathbf{z}}(p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x})))}{p_0(\mathbf{z}_i(t))} \right]}{\left[ \prod_{i=1}^{m_t} p_0(\mathbf{z}_i(t)) \right] \left[ (1 - \epsilon(m_t)) + \frac{\epsilon(m_t)}{m_t} \sum_{i=1}^{m_t} \frac{p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))}{p_0(\mathbf{z}_i(t))} \right]} \right)^T \right. \\ &\quad \times \left. \left( \frac{\left[ \prod_{i=1}^{m_t} p_0(\mathbf{z}_i(t)) \right] \left[ \frac{\epsilon(m_t)}{m_t} \sum_{i=1}^{m_t} \frac{\nabla_{\mathbf{z}}(p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x})))}{p_0(\mathbf{z}_i(t))} \right]}{\left[ \prod_{i=1}^{m_t} p_0(\mathbf{z}_i(t)) \right] \left[ (1 - \epsilon(m_t)) + \frac{\epsilon(m_t)}{m_t} \sum_{i=1}^{m_t} \frac{p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))}{p_0(\mathbf{z}_i(t))} \right]} \right) \right\} \quad (19) \end{aligned}$$

and (15) is true. This appears to be more interesting than it is since from these conditions, it must follow that the rank of  $\mathbf{J}$  is one. An example is the estimation of the position of a stationary target from range-only measurements—a rather trivial problem resulting in a singular  $\mathbf{J}$ , meaning that there is ambiguity.

A2) If

$$\mathbf{F}_t = \kappa \mathbf{F}^0 \quad (22)$$

in which

$$\mathbf{F}^0 \equiv \mathcal{E}\{(\nabla_{\mathbf{z}}[\log(p_1(\mathbf{z}))])(\nabla_{\mathbf{z}}[\log(p_1(\mathbf{z}))])^T\} \quad (23)$$

corresponds to the measurement-origin-uncertainty-free case [meaning that (14) can be written as  $\mathbf{J}_t^0 = \mathbf{M}_t^T \mathbf{F}^0 \mathbf{M}_t$ ] and where  $\kappa$  is a proportionality constant, then (15) is true.

The second case is the interesting one.

To explore (22), we rewrite (19) as

$$\mathbf{F}_t(m_t) = \mathcal{E} \left\{ \sum_{i=1}^{m_t} P(\chi_i | \mathbf{Z}(t); \mu_t(\mathbf{x}))^2 \cdot \frac{(\nabla_{\mathbf{z}}[p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))])(\nabla_{\mathbf{z}}[p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))])^T}{p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))^2} \right\} \quad (24)$$

in which

$$P(\chi_i | \mathbf{Z}(t); \mu_t(\mathbf{x})) \equiv \frac{\left[ \frac{\epsilon(m_t) p_1(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))}{m_t p_0(\mathbf{z}_i(t))} \right]}{\left[ (1 - \epsilon(m_t)) + \frac{\epsilon(m_t)}{m_t} \sum_{j=1}^{m_t} \frac{p_1(\mathbf{z}_j(t) - \mu_t(\mathbf{x}))}{p_0(\mathbf{z}_j(t))} \right]} \quad (25)$$

is recognizable as the posterior probability of the event  $\chi_i$  that the target-generated observation at time  $t$  is  $\mathbf{z}_i(t)$ , conditioned on the available data  $\{\mathbf{z}_j(t)\}_{j=1}^{m_t}$  and parametrized by  $\mu_t(\mathbf{x})$  [4]. If (22) is to be true, then  $\mathbf{F}_t(m_t)$  should not depend on  $\mu_t(\mathbf{x})$ , and it is seen that two possibilities on the *spurious-observation* distribution  $p_0(\cdot)$  are sufficient for this.

- B1)  $p_0(\mathbf{z}_i(t))$  is uniform, in which case, the substitution  $\tilde{\mathbf{z}}_i(t) = \mathbf{z}_i(t) - \mu_t(\mathbf{x})$  removes the effect of  $\mu_t(\mathbf{x})$  in the expectation (integration).
- B2)  $p_0(\mathbf{z}_i(t)) = \tilde{p}_0(\mathbf{z}_i(t) - \mu_t(\mathbf{x}))$ , in which case, the above substitution has the same effect.

Either of the above allows the integration implied by the expectation over  $\mathbf{Z}_t$  to be shifted by  $\mu_t(\mathbf{x})$ , and  $\mu_t(\mathbf{x})$  disappears from  $\mathbf{F}_t(m_t)$ . The first case B1) is that most commonly applied, and we will continue with it being assumed.

Under the assumption of a uniform  $p_0(\cdot) = 1/V$ , we write (24) (for  $i = 1$ ) as

$$\mathbf{F}_t(m) = \int P(\chi_1 | \mathbf{Z})^2 \left\{ \frac{(\nabla_{\mathbf{z}}[p_1(\mathbf{z}_1)])(\nabla_{\mathbf{z}}[p_1(\mathbf{z}_1)])^T}{p_1(\mathbf{z}_1)^2} \right\} \cdot \frac{m}{V^m} \left[ (1 - \epsilon(m)) + \frac{\epsilon(m)}{m} \sum_{j=1}^m V p_1(\mathbf{z}_j) \right] d\mathbf{Z} \quad (26)$$

in which

$$P(\chi_1 | \mathbf{Z}) \equiv \frac{\left[ \frac{\epsilon(m)}{m} V p_1(\mathbf{z}_1) \right]}{\left[ (1 - \epsilon(m)) + \frac{\epsilon(m)}{m} \sum_{j=1}^m V p_1(\mathbf{z}_j) \right]} \quad (27)$$

The disappearance of  $\mu_t(\mathbf{x})$  is obvious, and the time dependence has been (notationally) ignored. By symmetry, we need only concentrate on the event that the true measurement is labeled  $i = 1$ . Further simplification is possible by rewriting (26) as

$$\mathbf{F}_t(m) = \epsilon(m) \int \left\{ \frac{(\nabla_{\mathbf{z}}[p_1(\mathbf{z}_1)])(\nabla_{\mathbf{z}}[p_1(\mathbf{z}_1)])^T}{p_1(\mathbf{z}_1)^2} \right\} \cdot P(\chi_1 | \mathbf{z}_1) p_1(\mathbf{z}_1) d\mathbf{z}_1 \quad (28)$$

in which

$$P(\chi_1 | \mathbf{z}_1) \equiv \int_{\mathbf{z}_2 \in V} \cdots \int_{\mathbf{z}_m \in V} \frac{1}{V^{m-1}} \left[ \frac{\epsilon(m)}{m} V p_1(\mathbf{z}_1) \right] \cdot \frac{d\mathbf{z}_2 \cdots d\mathbf{z}_m}{\left[ (1 - \epsilon(m)) + \frac{\epsilon(m)}{m} \sum_{j=1}^m V p_1(\mathbf{z}_j) \right]} \quad (29)$$

Comparison of (28) with

$$\mathbf{F}^0 = \int \left\{ \frac{(\nabla_{\mathbf{z}}[p_1(\mathbf{z})])(\nabla_{\mathbf{z}}[p_1(\mathbf{z})])^T}{p_1(\mathbf{z})^2} \right\} p_1(\mathbf{z}) d\mathbf{z} \quad (30)$$

which is from (23), makes the information reduction from the measurements of uncertain origin clear.

To have  $\mathbf{F}_t(m)$  of (26) proportional to  $\mathbf{F}^0$ —that is, to satisfy condition A2)—is not straightforward. We present three cases that have so far been found sufficient.

- C1) If  $n_z = 1$ , then  $\mathbf{F}_t(m)$  is a multiple of  $\mathbf{F}^0$  whose value depends only on  $m$ . Thus, although this condition at first appears as the trivial A1), there is in fact no need for a “stationary target,” given that  $\mathbf{M}_t$  can depend on the time index. This is true regardless of the pdf  $p_1(\cdot)$ .
- C2) If all elements of  $\mathbf{z}_i$  are, given  $\chi_i$ , independent and identically-distributed (iid), and provided this distribution is symmetric, then (A2) is true.
- C3) If all elements of  $\mathbf{z}_i$  can be made to be iid and symmetric via a linear transformation, then (A2) is true.

It has been assumed that  $\epsilon(m_t)$  (the probability that any given measurement out  $m_t$  is target-generated), from (6), does not depend on  $t$ . Intuitively, a time-varying  $\epsilon(m_t)$  may render observations from certain  $t$  relatively more informative than they would be without uncertainty in measurement origin, and this can affect  $\mathbf{J}$ .

To show C2), note that even symmetry with respect to all arguments of  $p_1(\cdot)$  means that  $P(\chi_1 | \mathbf{z}_1)$  is even symmetric, and hence, all off-diagonal terms in  $\mathbf{F}_t(m)$  are zero. Further, since

all elements of  $\mathbf{z}_1$  are distributed identically, we have that  $\mathbf{F}_t(m)$  is a multiple of the identity matrix, as is  $\mathbf{F}^0$ . We thus have

$$\begin{aligned} \mathbf{J} &= \sum_{t=1}^T \mathbf{M}_t^T \left[ \sum_{m_t=1}^{\infty} p(m_t) \mathbf{F}_t(m_t) \right] \mathbf{M}_t \\ &= \sum_{t=1}^T \mathbf{M}_t^T \mathbf{F}_t \mathbf{M}_t \\ &= \sum_{t=1}^T \mathbf{M}_t^T \kappa \mathbf{F}^0 \mathbf{M}_t \\ &= \kappa \mathbf{J}^0. \end{aligned} \quad (31)$$

To show (C3), suppose

$$\mathbf{z}_i = \mathbf{B}\tilde{\mathbf{z}}_i \quad (32)$$

with the elements of  $\tilde{\mathbf{z}}_i$  iid. We then have

$$\begin{aligned} \mathbf{F}_t(m) &= \epsilon(m) \int \left\{ \frac{(\mathbf{B}^{-1})^T (\nabla_{\tilde{\mathbf{z}}} [p_1(\mathbf{B}\tilde{\mathbf{z}})]) (\nabla_{\tilde{\mathbf{z}}} [p_1(\mathbf{B}\tilde{\mathbf{z}})])^T \mathbf{B}^{-1}}{p_1(\mathbf{B}\tilde{\mathbf{z}})^2} \right\} \\ &\quad \cdot P(\chi_1 | \tilde{\mathbf{z}}_1) p_1(\mathbf{B}\tilde{\mathbf{z}}_1) |\mathbf{B}| d\tilde{\mathbf{z}}_1 \\ &= \epsilon(m) |\mathbf{B}| (\mathbf{B}^{-1})^T \int \frac{(\nabla_{\tilde{\mathbf{z}}} [p_1(\mathbf{B}\tilde{\mathbf{z}})]) (\nabla_{\tilde{\mathbf{z}}} [p_1(\mathbf{B}\tilde{\mathbf{z}})])^T}{p_1(\mathbf{B}\tilde{\mathbf{z}})^2} \\ &\quad \cdot P(\chi_1 | \tilde{\mathbf{z}}_1) p_1(\mathbf{B}\tilde{\mathbf{z}}_1) d\tilde{\mathbf{z}}_1 \mathbf{B}^{-1} \end{aligned} \quad (33)$$

in which

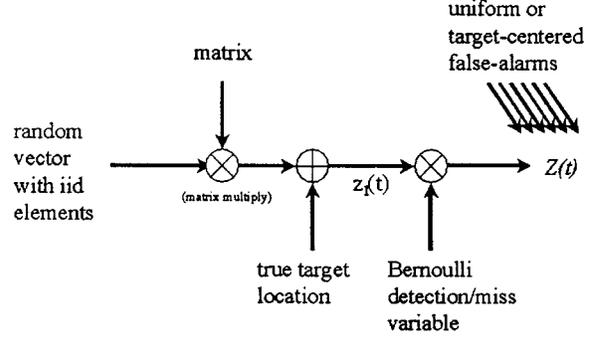
$$\begin{aligned} P(\chi_1 | \tilde{\mathbf{z}}_1) &\equiv \int_{\tilde{\mathbf{z}}_2 \in \tilde{V}} \cdots \int_{\tilde{\mathbf{z}}_m \in \tilde{V}} \frac{1}{V^{m-1}} \left[ \frac{\epsilon(m)}{m} V p_1(\mathbf{B}\tilde{\mathbf{z}}_1) \right] \\ &\quad \cdot \frac{\left[ (1 - \epsilon(m)) + \frac{\epsilon(m)}{m} \sum_{j=1}^m V p_1(\tilde{\mathbf{z}}_j) \right]}{|\mathbf{B}|^{m-1} d\tilde{\mathbf{z}}_2 \cdots d\tilde{\mathbf{z}}_m}. \end{aligned} \quad (34)$$

We also have

$$\begin{aligned} \mathbf{F}^0 &= \int \left\{ \frac{(\mathbf{B}^{-1})^T (\nabla_{\tilde{\mathbf{z}}} [p_1(\mathbf{B}\tilde{\mathbf{z}})]) (\nabla_{\tilde{\mathbf{z}}} [p_1(\mathbf{B}\tilde{\mathbf{z}})])^T \mathbf{B}^{-1}}{p_1(\mathbf{B}\tilde{\mathbf{z}})^2} \right\} \\ &\quad \cdot p_1(\mathbf{B}\tilde{\mathbf{z}}) |\mathbf{B}| d\tilde{\mathbf{z}} \\ &= |\mathbf{B}| (\mathbf{B}^{-1})^T \left\{ \int \frac{(\nabla_{\tilde{\mathbf{z}}} [p_1(\mathbf{B}\tilde{\mathbf{z}})]) (\nabla_{\tilde{\mathbf{z}}} [p_1(\mathbf{B}\tilde{\mathbf{z}})])^T}{p_1(\mathbf{B}\tilde{\mathbf{z}})^2} \right. \\ &\quad \left. \cdot p_1(\mathbf{B}\tilde{\mathbf{z}}) d\tilde{\mathbf{z}} \right\} \mathbf{B}^{-1}. \end{aligned} \quad (35)$$

Comparing (33) and (35) and proceeding as with C2), C3) follows if the elements of  $\tilde{\mathbf{z}}_i$  are iid.

Case C3), with an identity linear transformation, becomes C2), and C2), with  $n_z = 1$ , becomes C1). It can be seen that case C3) includes any case in which  $p_1(\cdot)$  is Gaussian. However,  $p_1(\cdot)$  need not be Gaussian for C3) to be satisfied: We show the class schematically in Fig. 1.



(31) Fig. 1. Schematic diagram of ways to generate observations such that multiplicative CRLB is obtained.

### III. COMPUTATION OF THE INFORMATION REDUCTION FACTOR

Equations (28) and (29) appear remarkably complicated for numerical calculation since they require an  $m$ -fold integral for each number of measurements  $m$ , and indeed, exact evaluation would be a nasty chore. However, under case C2) [or case C3) after linear transformation of the integration], it is only necessary to evaluate one diagonal element of the matrix, and let us assume this is the first element  $z_{11}$ . In principle, this still involves a difficult integration, but rewriting of (29) as

$$\begin{aligned} P(\chi_1 | z_{11}, m) &\equiv \int_{z_2 \in V} \cdots \int_{z_m \in V} \int_{\{z_{12} \cdots z_{1n_z}\}; z_1 \in V} \\ &\quad \frac{1}{V^{m-1}} p_1(z_{12}, \cdots, z_{1n_z}) \left[ \frac{\epsilon(m)}{m} V p_1(z_1) \right] \\ &\quad \times \frac{\left[ (1 - \epsilon(m)) + \frac{\epsilon(m)}{m} \sum_{j=1}^m V p_1(z_j) \right]}{\left[ (1 - \epsilon(m)) + \frac{\epsilon(m)}{m} \sum_{j=1}^m V p_1(z_j) \right]} \\ &\quad \cdot dz_2 \cdots dz_m dz_{12} \cdots dz_{1n_z} \end{aligned} \quad (36)$$

for insertion into

$$\mathbf{F}_t(m) = \left[ \epsilon(m) \int \left( \frac{\partial p_1(z_{11})}{\partial z_{11}} \right)^2 P(\chi_1 | z_{11}, m) p_1(z_{11}) dz_{11} \right] \cdot \mathbf{I}(n_z) \quad (37)$$

reveals that the averaged posterior probability need only be calculated at a *one-dimensional* (1-D) grid of values. This is easily and efficiently accomplished via a Monte Carlo technique [13]. In the above equation,  $\mathbf{I}(n_z)$  is an identity matrix of size  $n_z$ .

Similarly, we have

$$\mathbf{F}^0 = \left[ \int \left( \frac{\partial p_1(z_1)}{\partial z_1} \right)^2 p_1(z_1) dz_1 \right] \cdot \mathbf{I}(n_z) \quad (38)$$

and it is easy to see (39), shown at the bottom of the next page.

Since “far out” measurements are accorded small weight when a reasonable data association algorithm is used, we restrict our measurements to a validation gate [4], i.e.,

$$\begin{aligned} |z_i(t)_1 - \mu_t(\mathbf{x})_1| &< g\sigma_1 \\ &\vdots \\ |z_i(t)_{n_z} - \mu_t(\mathbf{x})_{n_z}| &< g\sigma_{n_z} \end{aligned} \quad (40)$$

where  $\sigma_1, \dots, \sigma_{n_z}$  are the standard deviations of corresponding dimensions of measurements, and we take  $g = 5$ . When condition C2) is met, i.e., all elements of  $\mathbf{z}_i$  are iid, then

$$\sigma_1 = \sigma_2 = \dots = \sigma_{n_z} = \sigma. \quad (41)$$

Thus, in (36),  $V$  can be replaced by the volume of validation gate  $V_g$ , which is defined as

$$V_g = \prod_{i=1}^{n_z} 2g\sigma_i = (2g\sigma)^{n_z}. \quad (42)$$

We discuss the idea of a gate only insofar as it impacts the Monte Carlo integration necessary to evaluate  $\kappa$ —strictly speaking, gating is never *necessary* for estimation, and for theoretical results to be best understood, the gate should be ignored. We will be interested in the situations where condition C2) or C3) is met. The case where the distribution is Gaussian has been discussed in [9] and [11], and hence, we will explore non-Gaussian measurement distributions  $p_1(\cdot)$ .

#### A. Generalized Gaussian Distribution

This family of symmetric distribution includes the Laplace and the Gaussian. The density of a generalized Gaussian random variable (zero mean) has the following form:

$$p_1(z) = \frac{k\alpha^{1/k}}{2\Gamma(\frac{1}{k})} \exp(-\alpha|z|^k) \quad (43)$$

where  $\alpha$  is a scale parameter, and  $k$  controls the shape. In order to measure the distributional tail weight, we introduce the “kurtosis,” which is defined as

$$K = \frac{\mathcal{E}\{z^4\}}{(\mathcal{E}\{z^2\})^2}.$$

For  $k = 1$ , we have

$$p_1(z) = \frac{\alpha}{2} \exp(-\alpha|z|) \quad (44)$$

which is Laplace (double-exponential), and the kurtosis is  $K = 6$ . For  $k = 2$ , we have

$$p_1(z) = \sqrt{\frac{\alpha}{\pi}} \exp(-\alpha z^2) \quad (45)$$

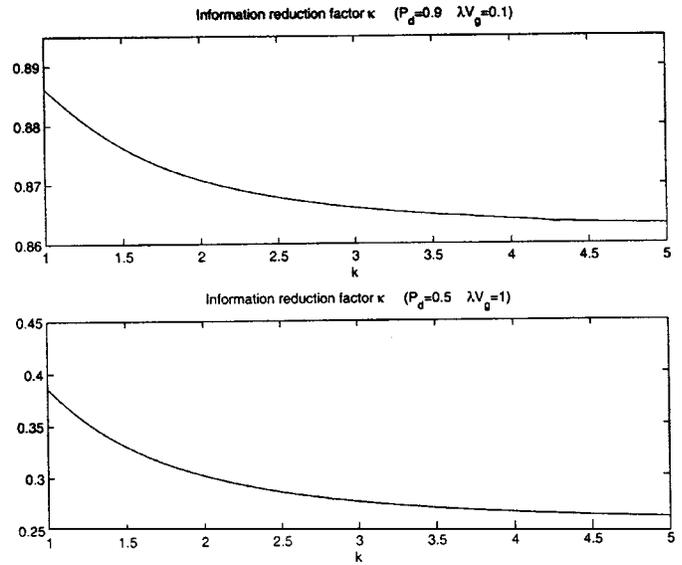


Fig. 2. Information reduction factor  $\kappa$  as a function of shape parameter  $k$  for generalized Gaussian noise. Upper plot:  $P_d = 0.9$  and  $\lambda V_g = 0.1$ . Lower plot:  $P_d = 0.5$  and  $\lambda V_g = 1$ .

which is Gaussian distribution, and the kurtosis is  $K = 3$ . For  $k = 4$ , we get

$$p_1(z) = \frac{2\alpha^{1/4}}{\Gamma(\frac{1}{4})} \exp(-\alpha z^4) \quad (46)$$

and the kurtosis is  $K = 2.19$ . The shape parameter  $k$  may be any positive number; in general, the smaller  $k$  is, the larger the kurtosis  $K$  and, hence, the heavier the tail of the distribution.

The results are shown for  $n_z = 2$  in Fig. 2. It is readily apparent that the information reduction factor  $\kappa$  is more significant (lower estimation fidelity) in the lower plot, corresponding to greater uncertainty (lower probability of detection and more false alarms). It is also interesting that  $\kappa$  increases with the relative tail weight of the measurement pdf. Precisely why this is so is open to interpretation, and we offer the following. First, recall that  $\kappa$  is the information reduction factor *relative* to the case of no measurement uncertainty but with the same measurement pdf  $p_1(\cdot)$ . We therefore speculate that a heavy-tailed distribution (low values of the shape parameter  $k$ ) is more forgiving of and robust to outlying observations that may be spurious than is a distribution that “expects” all to be reasonably centralized about the mean.

$$\kappa = \frac{\sum_{m=1}^{\infty} \left[ p(m)\epsilon(m) \int \left( \frac{\partial p_1(z_{11})}{\partial z_{11}} \right)^2 P(\chi_1|z_{11}, m) p_1(z_{11}) dz_{11} \right]}{\int \left( \frac{\partial p_1(z_1)}{\partial z_1} \right)^2 p_1(z_1) dz_1}. \quad (39)$$

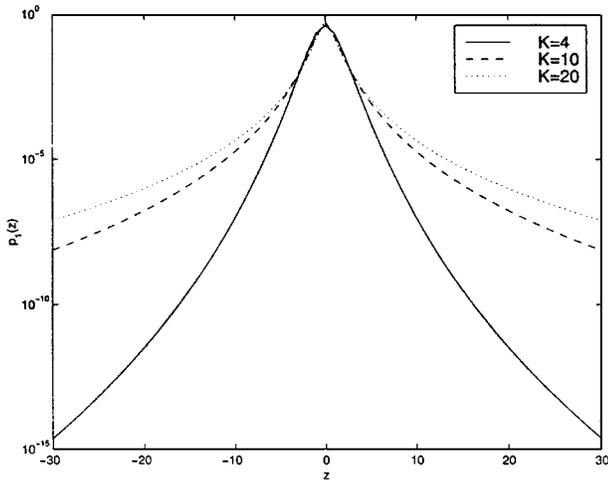


Fig. 3. The pdf of Johnson random variables for different kurtosis  $K$  ( $\sigma = 1$ ).

### B. Johnson Distribution

A Johnson random variable is the result of a nonlinear transformation upon a Gaussian random variable [10]. The density of a Johnson r.v. is

$$p_1(z) = \frac{\delta}{\tau \sqrt{2\pi \left( \left( \frac{z}{\tau} \right)^2 + 1 \right)}} \cdot \exp \left( -\frac{\delta^2}{2} \left[ \ln \left( \frac{z}{\tau} + \sqrt{\left( \frac{z}{\tau} \right)^2 + 1} \right) \right]^2 \right) \quad (47)$$

where

$$\delta = \sqrt{\frac{2}{\ln(\sqrt{2K-2}-1)}} \quad (48)$$

$$\tau = \sqrt{\frac{2\sigma^2}{\sqrt{2K-2}-2}} \quad (49)$$

in which  $K$  denotes the “kurtosis.” Examples are shown in Fig. 3.

Results for the Johnson case are shown in Fig. 4. As for the generalized Gaussian distribution, the information reduction factor increases (less information loss caused by measurement uncertainty) as the tail weight increases; even for quite heavy-tailed distributions (for a kurtosis value  $K = 20$ , the tail is very heavy, and by comparison, the kurtosis of a Laplace random variable is 6). It is interesting that the increase, as opposed to the kurtosis, appears to flatten out. In this paper, we have not considered the question, but it is possible that an asymptote is reached; it seems unlikely that the value unity is ever reached with data uncertainty.

## IV. SIMULATION RESULTS

To corroborate our results, we have performed several simulations. We explore a situation in which  $n_z = 2$ , which is the narrowband sonar case with bearing and frequency measurements (similar to that in [11]). Specifically, we have a constant-velocity

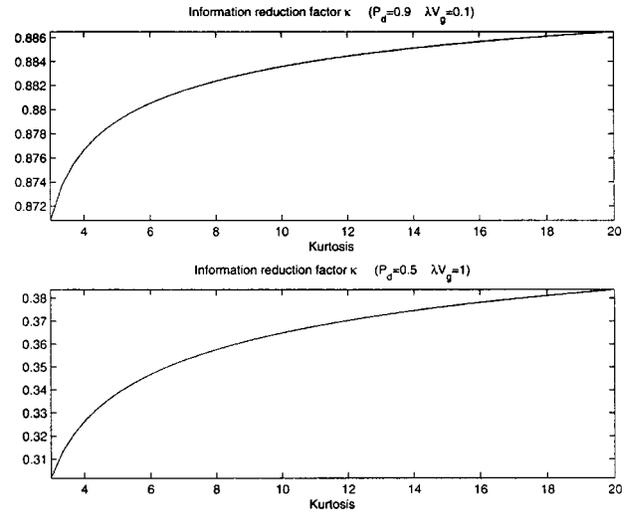


Fig. 4. Information reduction factor  $\kappa$  as a function of kurtosis for the Johnson noise. Upper plot:  $P_d = 0.9$  and  $\lambda V_g = 0.1$ . Lower plot:  $P_d = 0.5$  and  $\lambda V_g = 1$ .

target from which an unknown narrowband signal emanates; the observations are of angle (as observed from a moving platform) and frequency (with an implied Doppler shift from the relative target motion). Our intentions are

- i) to simulate this system with missed detections, false-alarms, and with non-Gaussian noise added to true observations when present;
- ii) to estimate the target parameters (initial position, initial velocities, and emitted frequency) via maximum likelihood;
- iii) to compare the resulting estimation covariance to the (modified) CRLB theory just developed.

One subtlety encountered is that estimation proceeds by maximization over a likelihood surface having a reasonably complex topography with the result that there is an occasional convergence to a “local” likelihood maximum. To avoid having our results skewed from this, we accept only estimates that pass a sensible test on residual errors—the theory is nicely confirmed.

### A. Target Model and ML Estimation

We assume that we have  $T$  sets of measurements that are indexed by  $t = 1, 2, \dots, T$  with succeeding snapshots separated by  $\Delta t$  seconds. The target motion is parametrized by a 5-D vector

$$\mathbf{x}(1) = [\xi(1) \ \eta(1) \ \dot{\xi} \ \dot{\eta} \ \gamma]^T. \quad (50)$$

In (50), we use the  $(t = 1)$ -element of the quantities  $\xi(t)$  and  $\eta(t)$ , which are the positions of the target in the east and north directions, respectively, at the time  $t$ ;  $\dot{\xi}$  and  $\dot{\eta}$  are the corresponding velocities, and  $\gamma$  is the unknown emitted frequency, which is assumed constant. This parametrization is possible because the target is assumed to be moving at a constant velocity, meaning that

$$\begin{aligned} \xi(t) &= \xi(1) + (t-1)(\Delta t)\dot{\xi} \\ \eta(t) &= \eta(1) + (t-1)(\Delta t)\dot{\eta}. \end{aligned} \quad (51)$$

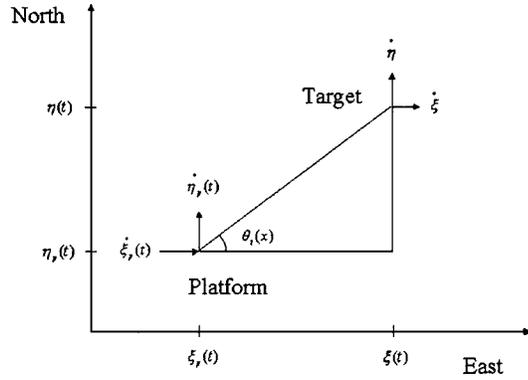


Fig. 5. Underlying coordinate frame for simulation.

The target is observed from a “platform” whose state at time  $t$  ( $t = 1, \dots, T$ ) is

$$\mathbf{x}_p(t) = [\xi_p(t) \eta_p(t) \dot{\xi}_p(t) \dot{\eta}_p(t)]'. \quad (52)$$

The true noiseless measurements of the target are of bearing

$$\theta_t(\mathbf{x}) = \tan^{-1} \left[ \frac{\eta(t) - \eta_p(t)}{\xi(t) - \xi_p(t)} \right] \quad (53)$$

and of shifted frequency; see (54) at the bottom of the page. The coordinate frame is shown in Fig. 5, and the trajectories are shown in Fig. 6. Although a constant-velocity target is observable via a nonmaneuvering platform via bearing and frequency-shift measurements (it is not observable from bearing-only measurements), a platform maneuver, as here, enhances accuracy.

The range of the measurements is  $[0, 2\pi)$  for the bearing and  $[B_1, B_2]$  for the frequency, where the latter is the bandwidth of the sonar signal processing filter. We identify a subregion

$$V_\theta = [A_1, A_2] \subset [0, 2\pi) \quad (55)$$

$$V_\gamma = [F_1, F_2] \subset [B_1, B_2] \quad (56)$$

for the bearing and frequencies, respectively. We define  $V = V_\theta \times V_\gamma$ , which is our surveillance region.

The set of measurements in  $V_\theta \times V_\gamma$  collected at  $t$  is denoted as

$$Z(t) = \{(\beta_{tj}, f_{tj})\}_{j=1}^{m_t} \quad (57)$$

where  $\beta$  and  $f$  are the bearing and frequency measurements, respectively. There are  $m_t$  measurement pairs collected at time  $t$ , and this set will include the target-originated pair [according to (53) and (54)] with probability  $P_d$  and with a number of false alarms having as prior a Poisson distribution with parameter  $\lambda V$ . The target-originated measurements are corrupted by additional zero-mean and independent noises

$$\beta_{tj_0} = \theta_t(\mathbf{x}) + \epsilon_t; f_{tj_0} = \gamma_t(\mathbf{x}) + \nu_t \quad (58)$$

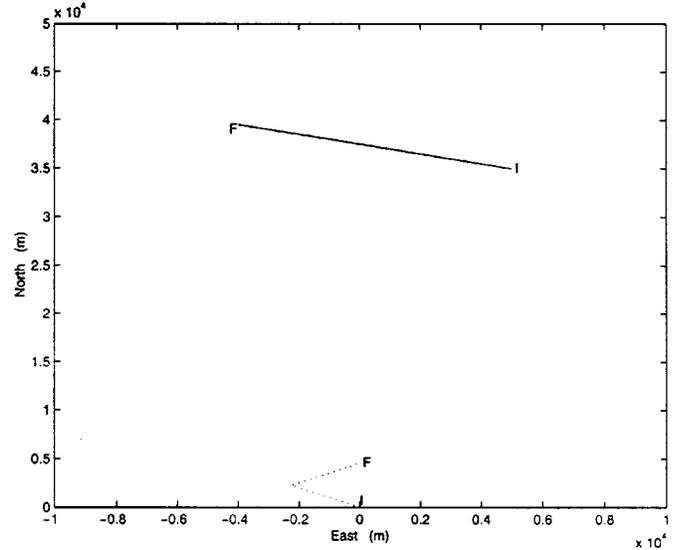


Fig. 6. In corroborative simulation, the trajectories of target (solid) and platform (dotted). “I:” Initial position of trajectories. “F:” final position of trajectories.

where  $j_0$  is the (unknown) index of the “good” measurement, that is

$$p(\epsilon_t, \nu_t) = \frac{1}{\sigma_\theta \sigma_\gamma} p_1 \left( \frac{\epsilon_t}{\sigma_\theta} \right) p_1 \left( \frac{\nu_t}{\sigma_\gamma} \right) \quad (59)$$

where  $p_1(\cdot)$  is the pdf of the measurement noise.<sup>4</sup> Second, the false measurements (clutter) are uniformly distributed

$$p(\beta_{tj}, f_{tj}) = \begin{cases} V^{-1}, & \text{in } V_\theta \times V_\gamma \\ 0, & \text{else} \end{cases} \quad (60)$$

in the rectangle  $V_\theta \times V_\gamma$ .<sup>5</sup> Substituting these and (6) into (12), we get the likelihood function of the measurements at the sampling time  $t$

$$p(\{\mathbf{z}_j(t)\}_{j=1}^{m_t}; \mathbf{x}) = \frac{\lambda^{m_t} e^{-\lambda V}}{m_t!} \left[ (1 - P_d) + \frac{P_d}{\lambda} \sum_{j=1}^{m_t} \frac{1}{\sigma_\theta \sigma_\gamma} p_1 \left( \frac{\beta_{tj} - \theta_t(\mathbf{x})}{\sigma_\theta} \right) p_1 \left( \frac{f_{tj} - \gamma_t(\mathbf{x})}{\sigma_\gamma} \right) \right]. \quad (61)$$

From (7), we have

$$p(\mathbf{Z}; \mathbf{x}) = \prod_{t=1}^T p(\mathbf{Z}(t)). \quad (62)$$

Our MLE  $\hat{\mathbf{x}}$  is found by maximizing (62). We use a simplex algorithm to find the maximum—this is not particularly efficient, but the goal of this section is corroboration and not suggestion of a technique.

<sup>4</sup>This is condition C2).

<sup>5</sup>This is condition B1).

$$\gamma_t(\mathbf{x}) = \gamma \left[ 1 - \frac{(\dot{\xi}(t) - \dot{\xi}_p(t)) \cos(\theta_t(\mathbf{x})) + (\dot{\eta}(t) - \dot{\eta}_p(t)) \sin(\theta_t(\mathbf{x}))}{c} \right] \quad (54)$$

Due to the multimodal nature of the log-likelihood ratio, we have to decide whether our estimate  $\hat{x}$  is acceptable or not; if it is not, then incorporation of the data to our estimate of covariance will corrupt the results. In [9], a statistical test is given to distinguish between a local or global maximum. We perform this test to accept or reject our estimate  $\hat{x}$ . Again, multifold integrations are needed to get parameters for the statistical test [9].

### B. Simulation Results

In narrowband sonar signal processing, different bands in the frequency domain are defined by an appropriate cell resolution and a center frequency about which these bands are located. The received signal is sampled and filtered in these bands before applying an FFT and beamforming. The signal processor was assumed to consist of the frequency band [500 Hz, 1000 Hz] with a 2048-point FFT so that the frequency resolution (cell) is

$$C_\gamma = 500/2048 = 0.25 \text{ Hz}. \quad (63)$$

For the bearing measurements, we assume that the sonar has 60 equal beams, resulting in an azimuth cell  $C_\theta$

$$C_\theta = 180^\circ/60 = 3.0^\circ. \quad (64)$$

Assuming a uniform distribution within a cell, the frequency and bearing measurements have standard deviations

$$\begin{aligned} \sigma_\gamma &= 0.25/\sqrt{12} = 0.07 \text{ Hz} \\ \sigma_\theta &= 3.0/\sqrt{12} = 0.87^\circ. \end{aligned} \quad (65)$$

We take the surveillance regions for bearing and frequency as

$$\begin{aligned} V_\theta &= [-20^\circ, 20^\circ] \\ V_\gamma &= [747 \text{ Hz}, 753 \text{ Hz}]. \end{aligned} \quad (66)$$

We restrict the validation gate to  $g = 5$ . In our simulations, we take as ground truth that the target moves at 10 m/s heading west and 5 m/s heading north, starting from (5000 m, 35 000 m). The emitted frequency is 750 Hz; therefore, the true target parameter is  $x = [5000 \ 35\ 000 \ -10 \ 5 \ 750]$ . There are 30 measurements (one each for 30 s), for a total observation interval of 15 min. For the first 15 scans, the platform moves in the northwest direction, and for the next 15, it moves northeast, both at 7.1 m/s. Fig. 6 shows the trajectories of target and platform.

We examine the case where the noise corrupting the target-generated measurement is Johnson. For Johnson parameter  $\delta$ , as indicated by its kurtosis, we take 100 Monte Carlo runs. An acceptance test automatically selects reasonable estimates. To verify the theory, we use the normalized estimation error squared (NEES) [2], which is defined as

$$\epsilon_x = (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{J}(\mathbf{x} - \hat{\mathbf{x}}) \quad (67)$$

where  $\hat{\mathbf{x}}$  is the estimate, and  $\mathbf{J}$  is the appropriate Fisher information matrix (20). Assuming the estimation error is approximately Gaussian, the NEES is chi-square distributed with  $n_x$  degrees of freedom, where  $n_x = 5$  is the dimension of the parameter being estimated  $x$ . Our results are listed in Tables I and

TABLE I  
RESULT OF MONTE CARLO RUNS FOR  
JOHNSON NOISE.  $P_d = 0.9$  AND  $\lambda V_g = 0.1$

Kurtosis	Number of accepted estimations	NEES of accepted estimations	99% confidence region of NEES
3.0	91	4.47	[4.19 5.90]
4.7	92	5.10	[4.19 5.89]
6.4	91	4.71	[4.19 5.90]
8.1	97	4.61	[4.21 5.87]
9.8	94	5.03	[4.20 5.88]
11.5	95	5.30	[4.20 5.88]
13.2	97	4.62	[4.21 5.87]
14.9	93	5.33	[4.20 5.88]
16.6	96	4.42	[4.21 5.87]
18.3	87	5.12	[4.17 5.92]
20.0	96	4.86	[4.21 5.87]

TABLE II  
RESULT OF MONTE CARLO RUNS FOR JOHNSON NOISE.  
 $P_d = 0.5$  AND  $\lambda V_g = 1.0$

Kurtosis	Number of accepted estimations	NEES of accepted estimations	99% confidence region of NEES
3.0	79	3.73	[4.13 5.96]
4.7	77	3.74	[4.12 5.98]
6.4	86	4.46	[4.17 5.92]
8.1	85	3.91	[4.16 5.93]
9.8	86	4.57	[4.17 5.92]
11.5	82	4.41	[4.15 5.95]
13.2	77	4.46	[4.12 5.98]
14.9	84	4.40	[4.16 5.93]
16.6	80	4.93	[4.14 5.96]
18.3	82	4.98	[4.15 5.95]
20.0	86	4.46	[4.17 5.92]

II. When  $P_d = 0.9$  and  $\lambda V_g = 1.0$ , corresponding to a *per-resolution-cell* (or per test) signal to noise ratio of 16 dB,<sup>6</sup> the NEES of the accepted estimations all fall within the 99% confidence region. When  $P_d = 0.5$  and  $\lambda V_g = 1.0$ , corresponding to a *per-resolution-cell* SNR of 4.8 dB,<sup>7</sup> eight of NEESs fall in the 99% confidence regions, whereas three are slightly too small. The deviations are minor, however, and are most likely the result of inappropriateness in the Gaussian assumption either in the test for acceptance or the NEES confidence values. In any event, even these small deviations occur for cases that are *closest* to the previously studied Gaussian situation. In Fig. 7, for a particular Johnson parameter value, the results of estimation, along with the theoretically predicted covariances, are shown. These

<sup>6</sup>The SNR numbers are given for reference only, and to calculate them, a Swerling I target model is assumed. Other definitions of SNR include *per-Hertz* (in this case 10 dB) and *broadband* (i.e., over the whole 6-Hz window, in this case 2.4 dB).

<sup>7</sup>Here, the *per Hertz* SNR is  $-1.2$  dB, and the *broadband* SNR is  $-9$  dB.

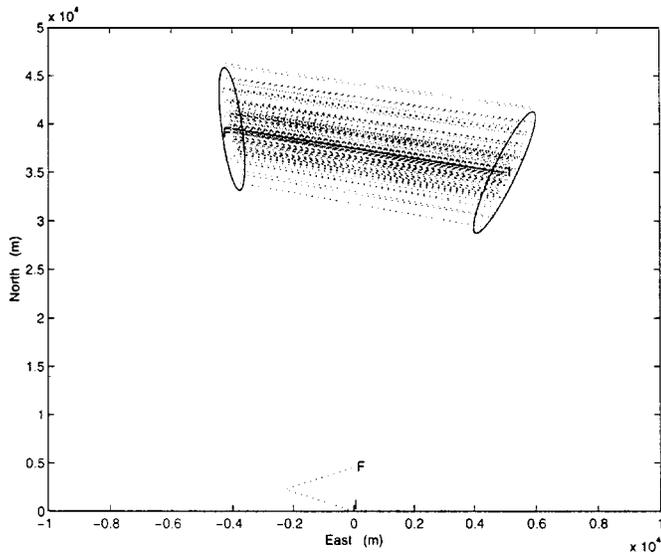


Fig. 7. True and estimated trajectories from 100 Monte Carlo runs for Johnson noise case  $P_d = 0.9$ ,  $\lambda V_g = 0.1$ , and kurtosis = 11.5. Note that in almost all cases, the estimated values lie within their respective ellipses.

latter are presented as two ellipses referring to the 99% confidence regions of the position estimates at the initial and final sampling instants, based on the CRLB (31). We can see that in 98 out of 100 Monte Carlo runs, the estimated initial and final positions fall into the 99% confidence region.

## V. SUMMARY

Measurements are of *uncertain origin* when, among those collected at each time (each “snapshot”), at most one can be said to be directly relevant to estimation of a quantity of interest. The remaining measurements have nothing to do with that quantity and can be considered “false alarms.” Indeed, at a given instant, the measurement *with* relevance can be absent, denoting a “missed detection.” The remarkable observation was recently made that the Fisher information matrix (and multiparameter CRLB), for the special case of bearings-only tracking and corrupting Gaussian noise, is, for the case *with* uncertainty in measurement origin, simply a multiple (the *information reduction factor*) of that *without* uncertainty. In this paper, we have investigated this result in greater depth and have found sufficient conditions for it to be true.

The interesting sufficient conditions are the following:

- i) the probabilities of detection ( $P_{ds}$ ) and false alarm rates are constant across all observation snapshots; and
- ii) a false (irrelevant) measurement has a uniform distribution; and
- iii) the random element to the true (relevant) observation is additive noise; and either
- iv) this noise is iid in each dimension; or
- v) there exists a linear transformation by which the noise could be made to be iid.

Naturally, this includes all Gaussian situations; the class is reasonably rich but not all-inclusive. We have checked the generalized Gaussian and Johnson cases and confirmed the analysis via simulation. One result is that the information reduction factor

increases—meaning that the degradation in theoretical estimation performance decreases—as the corrupting noise becomes heavier tailed. Presumably, an (optimal) estimator that is robust to large amplitude noise is also more forgiving of spurious measurements.

Finally, let us note that there is a rich and interesting tradition of bounds on estimation performance for dynamic systems (e.g., [3], [5], [8]), culminating in the excellent and explicit CRLB for the nonlinear filtering problem in [14]. In this paper, we have concentrated on results for parameter estimation only (i.e., the case without dynamics, and without need to resort to the MAP-estimation CRLB), but we have explored the case that there is measurement origin uncertainty (which is not treated in [3], [5], and [8]) and have focussed on the *scalar* information reduction factor. However, it is natural to consider the marriage of these two formulations, that is, to find bounds on estimation for a dynamic system with measurement origin uncertainty. This is, for us, current and future work.

## REFERENCES

- [1] J. R. Barra, *Mathematical Basis of Statistics*. New York: Academic, 1981.
- [2] Y. Bar-Shalom and X. Li, *Estimation and Tracking: Principles, Techniques and Software*. Norwell, MA: Artech House, 1993.
- [3] B. Bobrovsky and M. Zakai, “A lower bound on the estimation error for markov processes,” *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 785–788, Dec. 1975.
- [4] Y. Bar-Shalom and X. Li, *Multisensor, Multitarget Tracking: Principles and Techniques*. Storrs, CT: YBS, 1995.
- [5] J. Galdos, “A Cramer–Rao bound for multidimensional discrete-time dynamical systems,” *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 117–119, Feb. 1980.
- [6] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis*. London, U.K.: Chapman & Hall, 1997.
- [7] A. Hero, “Lower bounds on estimator performance for energy-invariant parameters of multidimensional poisson processes,” *IEEE Trans. Inform. Theory*, vol. 35, pp. 843–858, July 1989.
- [8] T. Kerr, “Status of CR-like lower bounds for nonlinear filtering,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, pp. 590–600, Sept. 1989.
- [9] C. Jauffret and Y. Bar-Shalom, “Track formation with bearing and frequency measurements,” *IEEE Trans. Aerosp. Electron. Syst.*, pp. 999–1010, Nov. 1990.
- [10] N. Johnson, “Systems of frequency curves generated by methods of translation,” *Biometrika*, vol. 36, pp. 149–176, 1949.
- [11] T. Kirubarajan and Y. Bar-Shalom, “Low-observable target motion analysis using amplitude information,” *IEEE Trans. Aerosp. Electron. Syst.*, pp. 1367–1384, Oct. 1996.
- [12] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [13] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C*. Cambridge, MA: MIT Press, 1992.
- [14] P. Tichavsky, C. Murachvik, and A. Nehorai, “Posterior Cramer-rao bounds for discrete-time nonlinear filtering,” *IEEE Trans. Signal Processing*, vol. 46, pp. 1386–1396, May 1998.
- [15] H. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968, vol. 1.



**Ruixin Niu** was born on August 29, 1972. He received the B.S. degree from Xi’an Jiaotong University, Xi’an, China, in 1994 and the M.S. degree from the institute of Electronics, Academia Sinica, Beijing, China, in 1997, both in electrical engineering. He is a Graduate Student/Research Assistant with the University of Connecticut, Storrs, where is pursuing the Ph.D. degree.

His research interests lie in signal processing, detection, and estimation theory.



**Peter Willett** (SM'97) received the B.A.S. degree in engineering science from the University of Toronto, Toronto, ON, Canada, in 1982 and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1986.

He is Professor with the University of Connecticut, Storrs, where he has been since 1986. His interests are generally in detection theory, target tracking, and signal processing.

Dr. Willett is an Associate Editor for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS and the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS.



**Yaakov Bar-Shalom** (S'63–M'66–SM'80–F'84) was born on May 11, 1941. He received the B.S. and M.S. degrees from the Technion—Israel Institute of Technology, Haifa, in 1963 and 1967, respectively, and the Ph.D. degree in from Princeton University, Princeton, NJ, in 1970, all in electrical engineering.

From 1970 to 1976, he was with Systems Control, Inc., Palo Alto, CA. Currently, he is the School of Engineering Distinguished Professor with the Department of Electrical and Systems Engineering and Director of the Estimation and Signal Processing (ESP)

Laboratory, University of Connecticut, Storrs. His research interests are in estimation theory and stochastic adaptive control. He has published more than 250 papers and book chapters in these areas. He coauthored the monograph *Tracking and Data Association* (New York: Academic, 1988), the graduate text *Estimation and Tracking, Principles, Techniques, and Software* (Boston, MA: Artech House, 1993), the text *Multitarget-Multisensor Tracking: Principles and Techniques* (Storrs, CT: YBS, 1995), and edited the books *Multitarget-Multisensor Tracking: Applications and Advances* (Boston, MA: Artech House, vol. I, 1990; vol. II, 1992; vol. III, 2000). He has been a consultant to numerous companies and originated the series of multitarget-multisensor tracking short courses offered via UCLA Extension, at Government Laboratories, at private companies, and overseas. He has also developed the commercially available interactive software packages MULTIDAT™ for automatic track formation and tracking of maneuvering or splitting targets in clutter, VARDAT™ for data association from multiple passive sensors, BEARDAT™ for target localization from bearing and frequency measurements in clutter, IMDAT™ for image segmentation and target centroid tracking, and FUSEDAT for fusion of possibly heterogeneous multisensor data for tracking. From 1978 to 1981, he was an Associate Editor of *Automatica*.

Dr. Bar-Shalom served as Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL during 1976 and 1977. He was program Chairman of the 1982 American Control Conference, General Chairman of the 1985 ACC, and Co-Chairman of the 1989 IEE International Conference on Control and Applications. From 1983 to 1987, he served as Chairman of the Conference Activities Board of the IEEE Control Systems Society, and from 1987 to 1989, he was a member of the Board of Governors of the IEEE CSS. He is a member of the Board of Directors of the International Society of Information Fusion (1999–2001) and Y2K President of the ISIF. In 1987, he received the IEEE CSS Distinguished Member Award. Since 1995, he has been a Distinguished Lecturer of the IEEE AESS and has given several keynote addresses at major national and international conferences. He was co-recipient of the M. Barry Carlton Award for the best paper in the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS in 1995 and the 1998 University of Connecticut AAUP Excellence Award for Research.