MultiTrajectory Estimation in Noise and Clutter

M.Malyutov and M.Lu
Department of Mathematics
Northeastern University
Boston MA 02115, USA
E-mail:MLTV@NEU.edu,
MILU@lynx.neu.edu

Andrey Nikiforov
BBN Technologies Inc,
10 Moulton Street, Cambridge, MA
02138, E-mail:anikifor@bbn.com

Rostislav Protassov
Statistics Department
Harvard University,
1 Oxford St., Cambridge, MA
02138, e-mail:
protasso@stat.harvard.edu

Abstract – *Suppose that at each frame we observe several points describing several trajectories positions at a given time moment disturbed by noise and clutter. The clutter is a background noise causing deletion of certain points and creation of false ones by a mechanism which is both time and space correlated. The non-random trajectories are assumed which either belong to a parametric family or have sufficient smoothness. We apply either some modifications of the EM-algorithm, or the method of fitting certain symmetric functions of measurements to fit the trajectories robustly. A comparative simulation results of various algorithms' performance is discussed.*

Keywords: fitting multiple trajectories, assignment of measurements to trajectories, symmetric functions of measurements, EM-algorithm

I Introduction

Traditional methods of multiple target tracking (MTT) are based on preliminary data association (PDA) (measurements and the objects are associated at each frame according to some criterion) [1]. The number of computations required by these methods grows exponentially in both the number of frames and the number of targets requiring supercomputers. Additionally, we have not found any investigation of these methods consistency even in the simplest case of deterministic trajectories observed with random noise (no clutter), if the distances between targets are comparable with the standard errors of measurements. As a result, it may be the case that whatever amount of data and computation are available, the algorithm resolution limit may exceed certain nonvanishing limits.

Generally, our problem can be embedded into the following data processing model. The input data consists of a sequence of frames which are the results of observations of some moving objects in noise and clutter; "clutter" is the component of noise which is correlated in time and space, its popular model is dealt with by us in [10], the non-correlated component is the instrumental noise. On the basis of the frames, it is necessary to detect all the objects and to estimate their trajectories. These procedures are based on some prior knowledge about the background, the object's movement equations etc.

We skip here any discussion of the Background Filtration Problem (suppression of the background noise), of the Signal Detection Problem (detection of useful signals including both their localization and an estimate of their amplitude), and the dynamic features of the Tracking Problem (dynamic tracks detection and maintenance of the moving objects).

Here we deal only with a posteriori *fitting* the motions' non-random trajectories (either parametric or sufficiently smooth) in noise and clutter. The first efficient method solving this problem for trajectories decribed by polynomial equations in clutter absence, namely Symmetric Functions of Measurements (SFM), was proposed in seventies for Soviet anti-missile defense, see [2]. We outline here the simulation results of modified SFM for locating fixed targets in clutter presence in section IV, and both simulation and theoretical results of a joint forthcoming paper with A.B. Tsybakov on fitting non-parametrically described multi-trajectories in section V.

We describe also another approach to fitting non-random trajectories based on certain modifications of the EM-algorithm having a wider applicability range than the SFM-method. The EM was previously applied to fitting a non-random trajectory in a particular case of incomplete measurements (observing only one of targets at each frame) typical for radar measurements of a *distributed* target motion which is also known in Econometrics under the title *switching regression models* [7,14]. A simplification of the EM for this case was proposed in [9]. Dr. Roy Streit developed PMHT algorithm based on a modified version of the EM for tracking the trajectories described by the Kalman filter (see [17] for a collection of 42 papers from 1992 to 1998 developing this approach).

Our present application is much simpler and can serve as an introduction for the beginners in MTT. Still, it appears useful in many applications such as tracking ballistic missiles.

In a demo-code implementing our application of the EM-algorithm (available by request from the third author) we simulate noisy non-assigned measurements of several non-random trajectories described by parametric models, and restore the trajectories in real time on a laptop with good precision. An extension of this demo to include clutter and robust version of the EM to analyze the data is to be shown at the conference.

II Model

We assume that there exist k objects, for convenience labeled 1 through k, moving along k distinct trajectories; k is a known integer. T images (or *frames*), indexed by t ($1 \leq t \leq T$), of the objects are made. Each frame is reported to us as a vector $X[t] = (x_1[t], ..., x_k[t])$ with k coordinates, whose (unknown) permutation corresponds to coordinates of the objects 1 through k plus random noise whose distribution is known. More specifically, we assume that for frame t object i, $1 \leq i \leq k$, generates coordinate $z_i[t]$ from the density $f_t(z|\theta_i)$, where $f_{(\cdot)}(\cdot|\cdot)$ is a known function, and $X[t]$ is some (generally unknown) permutation of the elements of the vector $Z[t] = (z_1[t], ..., z_k[t])$. It is convenient to think about the mechanism that generated the data $X[t]$ as consisting of two stages:

Stage 1: Coordinate Generation Mechanism. $z_i[t]$ belonging to trajectory i is generated from density $f_t(z|\theta_i)$, $1 \leq i \leq k$.

Stage 2: Label Assignment Mechanism. Coordinates $z_i[t]$ are assigned to observations $x_i[t]$ via a one-to-one label assignment mechanism. E.g., for $k = 2$ and independent $z_1[t], z_2[t]$ this mechanism can be deterministic: $x_1[t]$ equals $z_1[t]$, in which case the correct likelihood function for the observation $X[t] = (x_1[t], x_2[t])$ is

$$\mathcal{L}_t(X[t]|\theta_1, \theta_2) = f_t(x_1[t]|\theta_1) f_t(x_2[t]|\theta_2)$$

Alternatively, the mechanism can be stochastic, e.g., assign $z_1[t]$ to $x_1[t]$ with probability 0.7, independently of anything else, in which case the correct likelihood function for the observation $X[t]$ is

$$\mathcal{L}_t(X[t]|\theta_1, \theta_2) = 0.7 f(x_1[t]|\theta_1)_t f_t(x_2[t]|\theta_2) \\ +0.3 f_t(x_2[t]|\theta_1) f_t(x_1[t]|\theta_2).$$

The principal difficulty with the statistical analysis of the data $\{X[t]\}_{t=1}^T$ is that often almost nothing is known a priori about the Label Assignment Mechanism. Depending on the particular device used, for time t this mechanism may depend on t, previuosly made assignments, values of $X[\tau]$ up to and including time t, etc., etc. Therefore it is not possible to write down the likelihood for the data $\{X[t]\}_{t=1}^T$ explicitly. Nevertheless, consistent estimation of the parameters $\theta = (\theta_1, ..., \theta_k)$ is still possible if the following assumptions are met.

Assumptions.

1. Let $z_i[t] \sim f_t(z|\theta_i)$, $1 \leq i \leq k$, indenpendent for all i, t;
2. For a given vector $Z[t] = (z_1[t], ..., z_k[t])$ let $Y[t] = (y_1[t], ..., y_k[t])$ be a random permutation such that any of the $k!$ possible permutations s of indices $1, ..., k$ has the same probability $1/k!$:

$$\Pr[(y_1[t], ..., y_k[t]) = (z_{s(1)}[t], ..., z_{s(k)}[t])] = \frac{1}{k!};$$

Suppose further that $f_t(z|\theta)$ are such that the likelihood function

$$\begin{array}{l} g(Y|\theta) = \prod_{t=1}^T g_t(Y[t]|\theta), \text{ where} \\ g_t(Y[t]|\theta) = 1/k! \sum_{s \in S_k} \prod_{i=1}^k f_t(y_{s(i)}|\theta_i) \end{array} \quad (1)$$

satisfies the standard regularity conditions for the MLE of θ to be consistent and asymptotically normal to hold (see e.g., [15]). Here S_k denotes the set of all permutations of k indices.

Lemma. Under the above assumptions, given the data $\{X[t]\}_{t=1}^T$, argmax $g(X|\theta)$, g being defined in Equation 1, will be a consistent asymptotically normal estimate of θ *under any Label Assignment Mechanism.*

Proof. Suppose for each time frame t, $Y[t]$ satifying assumption 2 were actually available to us. Then, for each t, there would exist a permutation of indices $s_t \in S_k$ such that

$$\begin{array}{l} (x_1[t], ..., x_k[t]) = (y_{s_t(1)}[t], ..., y_{s_t(k)}[t]) \text{ and thus} \\ g_t(X[t]|\theta) = g_t(y_{s_t(1)}[t], ..., y_{s_t(k)}[t]|\theta), \end{array}$$

g_t being defined in Equation 1. By definition, $g_t(y_1[t], ..., y_k[t]|\theta) = g_t(y_{s(1)}[t], ..., y_{s(k)}[t]|\theta)$ for any $s \in S_k$, thus the argmax will be a consistent asymptotically normal estimate of θ.

Remark. Estimating θ via maximization of the objective function $g(X|\theta)$ is thus operationally equivalent to treating the data X as being generated in the same manner as the $\{Y[t]\}_{t=1}^T$ defined in Assumptions 1-2 above and then estimating θ via the maximum likelihood method. That is, we can pretend that the Label Assignment Mechanism for $X[t]$ was equiprobable for the $k!$ possible permutations of indices and was independent of anything else.

With this in mind, we can thus transform our problem into a classical MLE problem in the presense of missing data, the *complete data* being given by $\{(X[t], I[t])\}_{t=1}^T$, where $\{I[t]\}_{t=1}^T$, $I[t]$ being the vector of length k indicating the trajectory from which each $x_i[t]$ came, are the *missing data*. The distribution of the missing data is postulated to be equiprobable over all possible permutations of indices, independently of anything else. By abuse of notation, from now on we shall call the objective function $g(X|\theta)$ the *likelihood function* for the reasons we have just outlined.

We deal with missing targets in frames by using marginal distributions obtained via integration over missing variables. As shown in (Nikiforov 1991), this approach keeps the optimality properties of MLE provided that the missing data are ignorable, i.e., the missing data are independent of their unobserved values and the parameters governing the missingness mechanism are distinct from the parameters of the observed data (see [6]). Both of these assumptions are trivially satisfied in the case of completely missing $I[t]$.

The EM-algorithm

The EM-algorithm [5] is the iterative algorithm that converges under rather weak conditions to a local maximum of the likelihood function. One iteration of the EM-algorithm in our problem is described below.

The estimation of parameters for the MTT problem may be interpreted as one involving incomplete data by regarding unlabeled observations as if they missed class (object) indices.

Let $X[t] = \{x_1[t], \ldots, x_k[t]\}$ be the group (pair, if $k = 2$) of observations at time t. We regard $X[t]$ as an observed part of complete observation $X_{complete}[t] = (X[t], I[t])$, where $I[t]$ is some permutation of k integers from 1 to k indicating the origin of each $x_i[t]$ in $X_{complete}[t]$.

We shall use the superscripts "-" and "+" for the current and next iterations respectively.

The general description of the EM-algorithm is as follows:

- $E-$step. Determine $Q(\phi|\phi^-)$, where
$$Q(\phi|\phi^-) = E\left(\log f(x_{complete}|\,\phi)|\, x_{observed}, \phi^-\right). \quad (2)$$
- $M-$step. Determine $\phi^+ = arg\max_\phi Q(\phi|\,\phi^-)$.

Following this general description and the treatment of the model accepted, we shall derive the EM-algorithm for MTT.

$$Q(\phi|\phi^-) = \sum_{i_1,\ldots,i_T} p^-(i_1, \ldots i_T) \log \prod_{t=1}^{T} f(X[t]\,|\, i_t, \phi), \quad (3)$$

where we denote by i_t the permutation within the tth group, and $p^-(i_1, \ldots i_T)$ is the estimate of probability (based on parameter values from the previous iteration) of the permutation sequence $(i_1, \ldots, i_T)$. Here

$$f_t(X[t]|i_t, \phi) = \prod_{j=1}^{k} f_t(x_{i_t(j)}[t]|\theta_i), \quad (4)$$

where $i_t(j)$ is the $j-$th value in the permutation i_t.

Due to the independence of **groups** we have

$$Q(\phi|\phi^-) = \sum_{i_1,\ldots,i_T} p(i_1|X[1], \phi^-) \ldots$$
$$p(i_T|X[T], \phi^-) \sum_{t=1}^{T} \log f(X[t]|i_t, \phi) =$$
$$\sum_{t=1}^{T} \sum_{\{i_t\}} p(i_t|x, \phi^-) \log f_t(X[t]|i_t, \phi)$$

where $\sum_{\{i_t\}}$ is the sum over all permutations within the $t-$th group, and

$$p(i_t|X[t], \phi^-) = \frac{f_t(X[t]|i_t, \phi^-)}{\sum_{\{i_t\}} f_t(X[t]|i_t, \phi^-)} \quad (5)$$

are the posterior probabilities of permutation i_t, calculated using ϕ^-. Hence, the M-step becomes the set of k "usual" weighted MLE procedures. Let $p_i(n|t)$ be the estimate of posterior probability of $x_n[t], n = 1, \ldots, k$, to come from the class i at moment t at the current iteration of the EM. If $k = 2$, we have:

$$p_1(1|t) = p_2(2|t) = \frac{f_t(x_1[t]\,|\,\theta_1^-) \cdot f_t(x_2[t]\,|\,\theta_2^-)}{2h^-(X|\theta)}$$
$$p_2(1|t) = p_1(2|t) = \frac{f_t(x_1[t]\,|\,\theta_2^-) \cdot f_t(x_2[t]\,|\,\theta_1^-)}{2h^-(X|\theta)}.$$

Here h^- means the estimate of h at the current iteration, where

$$h(X|\theta) = 1/k! \sum_{s \in S_k} \prod_{i=1}^{k} f_t(x_{s(i)}|\theta) \quad (6)$$

For the case of $k > 2$ classes, the numerator of $p_i(n|t)$ is the sum over all products $f(x_j[t]\,|\,\theta_i, t) \ldots\ldots$ with $x_n[t]$ corresponding to the class i, and the denominator is $k!\; h(X[t]|\theta)$.

The above equations are independent of particular model $f_i(x|t, \theta)$. To describe further the E-step and M-step in more details, we must specify a model.

For the simple "static" case, when $f_i(x|t, \theta) = f(x|\,\mu_i, \sigma_i)$, where f is the normal density, the estimate of the mean (say) for class $i, i = 1, 2$, at the next iteration is

$$\mu_i^+ = \frac{\sum_0^{T-1} \left(p_i(1|t)\, x_1[t] + p_i(2|t)\, x_2[t]\right)}{\sum_0^{T-1} \left(p_i(1|t) + p_i(2|t)\right)} \quad (7)$$

If $k > 2$, we have for class $i, i = 1, \ldots, k$:

$$\mu_i^+ = \frac{\sum_{t=0}^{T-1} \sum_{j=1}^{k} p_i(j|t)\, x_j[t]}{\sum_0^{T-1} \sum_1^k p_i(j|t)} \quad (8)$$

Algorithm complexity

At one iteration, for each t $(1 \le t \le T)$, $k!$ products of k densities $f_t(.|.)$ for all permutations are calculated. Calculation of sums for μ_i^+, etc is only of order k^2. Therefore, for large k the main computational burden at one iteration is computation of $p_i(n|t)$, since it requires $(k-1)k!T$ multiplications if direct scheme is implemented (see below) and the same number of additions. Fitting assignment weights separately in each iteration as in the version of the EM proposed in [12] and similar to that used in PMHT [17], allows reducing the complexity for large k. For small k (2-4) the time spent is due mostly to computation of $f(\cdot)$ and additive statistics.

The number $(k!)$ in the complexity estimate may be further reduced for large k by simply considering only **close** subgroups of targets, since

$$f_t(x_{i_t(1)}|\theta_1) \ldots f_t(x_{i_t(k)}|\theta_k), \quad (9)$$

where $i_t(k)$ is a permutation of $(1, ..., k)$, is [nearly] zero in other cases. Also saving partial products instead of direct calculation of $f_t(x_{i_t(1)}|\theta_1) \ldots f_t(x_{i_t(k)}|\theta_k)$ for every permutation (and similar schemes) will strongly reduce the time.

EM advantages

(+) consistent and asymptotically normal under mild regularity conditions.
(+) Ability to process various kinds of missing data.
(+) Variables may not be nessesarily euclidian coordinates of objects: we can include size, speed etc. Discrete variables are no exception.
(+) easy to implement
(+) does not require calculation Hessian and its inversion . This is important for multiclass and multivariate problems, especially if estimation of say, covariance matrix, is required.
(+) very flexible. The ability of quick and effective modifications is mostly valuable. The list of potentially available extentions is:

- various types of parametric models,
- effective modelling of error distribution via full, tree or diagonal covariance,
- partial classification can be forced:e.g. if it is known for some measurements at some moments t, that they come from particalar classes, or probabilistic pre-classification or relative pre-ranking (targets are ranked with respect to their chances to came from the class 1, 2....) etc.
- model assumptions, e.g.:
  - equal or general distribution of errors,
  - equal or general traces of covariance matrices (rotation invariant)
  - narrowing the parameter space description (mean $|c| < const, a > b$ etc.)
  - a priori assumption ofequal "time-slopes" (colinear trajectories)
  - a priori assumption of a common origin of trajectories etc

(+/-) Parametric description. If a model of distribution of errors is adequate, we get effective procedures, exploiting this information. If it is not, the EM estimate may be of poor quality. Robust versions may be developed for this case.

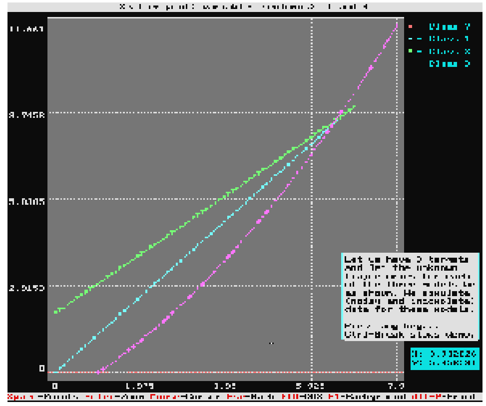

Fig 1: Motion along three parametric trajectories shown used in further simulation.

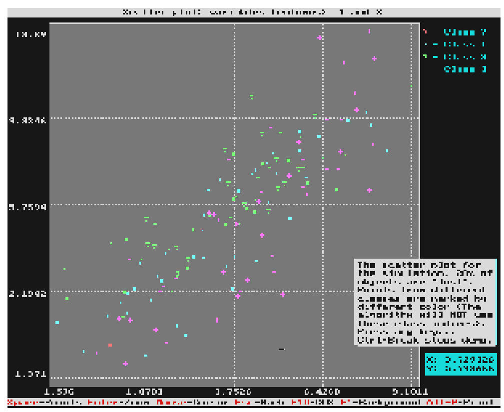

Fig 2: The scatterplot of the simulated unmarked triples. .20 of the positions are randomly lost. Points from different classes are marked by different color (Our algorithm will NOT use these class indices)

III Clutter modeling

Consider m not moving targets located at points $A_i, i = 1, \dots, m$, from the interval $[A, B]$ and $\min_{i \neq j} |A_i - A_j| \geq \delta$. We observe the targets at moments $t_n = n\Delta t$ and measure the target positions with random errors. Let $x_i(n)$ be the measurement of the i-th target on the n-th frame. We model $x_i(n) = A_i + y(i, n)$, where $y(i, n)$ are i.i.d. random variables (RV) with the uniform distribution on the interval $[-\Delta, \Delta]$. The case $\Delta > \delta/2$ is of interest since the measurements assignment to targets cannot be reliably achieved under this condition.

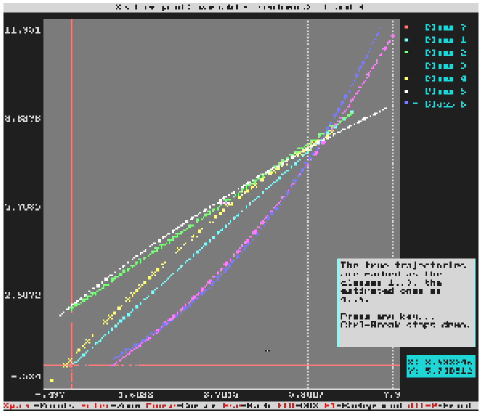

Fig 3: The true trajectories are marked as 1,2 and 3. The estimated ones are marked 4,5 and 6

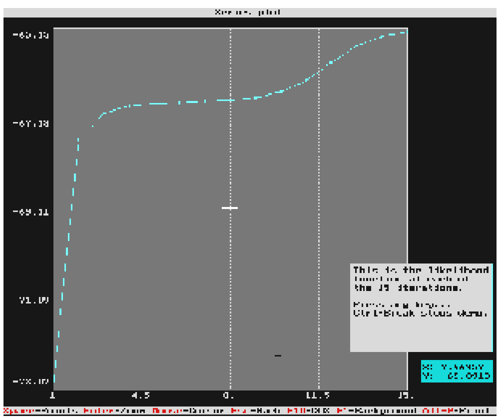

Fig 4: The likelihood function graph during 15 iterations.

The second class of measurement errors is generated by the clutter. We assume analogously to [16], that a clutter generating noise ξ_t is a stationary gaussian process with zero mean and covariance function $f(t) = 1 - \frac{\lambda^2}{2}t^2 + \frac{\lambda_2}{24}t^4 + o(t^4)$ as $t \to 0$ and $f(t) = O(t^{-b})$ as $t \to \infty, 0 < b < 1$. (see [4], (12.1.1)). The process ξ_t generates a Poisson process of false targets (see [4], (12.2.1)) with rate (intensity)

$$\Lambda_{cl} = \Lambda \frac{\lambda}{2\pi} e^{-H_1^2/2}, \qquad (10)$$

where Λ characterizes a clutter intensity on the interval $[A, B]$ and H_1 is specified as the threshold for a signal to appear on the frame. The value H_1 is determined according to the intensity of process ξ_t. A time interval for a false target presence on a frame is a positive RV with a density function

$$(1/2)\pi c t e^{-\pi c^2 t^2/4}, \qquad t > 0, \qquad (11)$$

where

$$c = H_1 \lambda / \sqrt{2\pi} \qquad (12)$$

(see [4] , (12.5.2)).

The third class of measurement errors arises, when some targets are lost on a frame, which is modeled as follows. Let the actual target signal level be $H_2 + H_1$. A target is lost at a frame, if the value of the clutter process is less than $-H_2 < 0$. The sign symmetry of the clutter gaussian process implies that the formulas (10)-(12) with H_2 instead of H_1 are valid for the missing targets.

Starting times for targets becoming missed constitute the Poisson process with rate

$$\Lambda_{tr} = \frac{\lambda}{2\pi} e^{-H_2^2/2},$$

and the time interval the target is lost has a random length with a density function

$$(1/2)\pi d t e^{-\pi d^2 t^2/4},$$

where

$$d = H_2 \lambda / \sqrt{2\pi}$$

for every target.

Algorithm

Let m_n be a number of detected objects on the n-th frame and T_n^k be the $(k-1)$-th degree coefficients of the polynomial $(x - x_1(n)) \cdot \dots \cdot (x - x_{m_n}(n))$, $k = 1, 2, \dots, m_n$. Let us enumerate anew the frames satisfying $m_n \geq K$ from 1 to N_K and calculate $\bar{T}_k = 1/N_K \sum_{n=1}^{N_K} T_n^k$ and $\hat{T}_k$ as the median of a set $\{T_1^k, \dots, T_{N_K}^k, k = 1, \dots, K\}$.

It is evident that $N_k \geq N_{k+1}$ for every k. Let $M = \max\{k : N_k \geq N/2\}$. Based on the statistics $\bar{T}_k$ we construct a polynomial $\bar{P}(x) = x^M + \sum_{k=1}^{M} \bar{T}_k x^{M-k}$. The roots of this polynomial are the estimates for the target positions based on the mean.

The median based estimates are constructed similarly if we use the statistics $\hat{T}_k$ instead of $\bar{T}_k$.

These algorithms are generalized in a straightforward way for targets moving according to various linear combinations of several smooth base functions. Mean-based estimator becomes a version of the EM, whereas the median-based one is replaced with the least median of squares applied to symmetrical functions of measurements. The simulation results are the subject of our subsequent publication. The codes are available from the authors.

Numerical results

It follows from Table 1 that the algorithm based on the mean values has a wider interval of applicability under the noise parameters than the one based on medians.

Δ	Target positions estimates					
	Median based			Mean based		
0.000	2.00	2.25	3.00	2.00	2.25	3.00
0.188	1.96	2.37	2.95	2.02	2.25	2.99
0.375	1.94	2.19	3.07	1.99	2.23	3.01
0.563	1.75	2.34	3.11	1.94	2.21	3.05
0.750	3.26	N/f	N/f	1.92	2.23	3.05
0.938	1.88	2.02	3.32	3.08	N/f	N/f
1.125	1.45	N/f	N/f	2.06	2.20	3.05
1.313	1.14	N/f	N/f	2.86	N/f	N/f
1.500	1.12	N/f	N/f	1.73	2.25	2.95
1.688	1.15	N/f	N/f	2.10	N/f	N/f

Table 1. Results of numerical simulations: no clutter and no measurements missed.
Input parameters:
number of targets $m = 3$, $\delta = 0.25$,
number of observations $N = 100$,
real targets positions $A_1 = 2, A_2 = 2.25, A_3 = 3$.

N	H_1	H_2	Target positions estimates		
			Median based		
1	3.742	3.850	1.95	2.27	3.03
2	3.606	3.700	1.97	2.31	2.97
3	3.464	3.550	1.97	2.27	2.99
4	3.317	3.400	1.90	2.43	2.96
5	3.162	3.250	1.85	2.55	2.85
6	3.000	3.100	2.02	2.27	3.01
7	2.828	2.950	3.05	N/f	N/f
8	2.646	2.800	3.10	N/f	N/f
9	2.449	2.650	2.05	2.32	3.13
10	2.236	2.500	4.59	N/f	N/f

N	Target positions estimates			Errors in %	
	Mean based			Lost	Add
1	1.99	2.25	3.00	0	0
2	1.97	2.28	2.99	0	1
3	3.21	N/f	N/f	0	3
4	3.56	N/f	N/f	0	4
5	3.62	N/f	N/f	1	4
6	3.84	N/f	N/f	1	8
7	4.02	N/f	N/f	2	9
8	4.28	N/f	N/f	5	15
9	4.90	N/f	N/f	6	32
10	5.72	N/f	N/f	7	55

Table 2.Input parameters:
$m = 3$, $\Delta = 0.3$, $N = 100$, $\lambda = 1$, $\Delta t = 0.1$,
$A_1 = 2, A_2 = 2.25, A_3 = 3$.

N	H_1	H_2	Target positions estimates		
			Median based		
1	3.742	3.850	1.97	2.27	3.00
2	3.606	3.700	1.97	2.31	2.98
3	3.464	3.550	2.01	2.23	3.00
4	3.317	3.400	1.94	2.36	2.97
5	3.162	3.250	1.94	2.36	2.96
6	3.000	3.100	2.00	2.27	3.00
7	2.828	2.950	2.01	2.24	3.01
8	2.646	2.800	2.01	2.25	3.02
9	2.449	2.650	2.02	2.27	3.05
10	2.236	2.500	4.61	N/f	N/f

N	Target positions estimates			Errors in %	
	Mean based			Lost	Add
1	2.00	2.25	3.00	0	0
2	2.00	2.23	3.01	0	1
3	3.22	N/f	N/f	0	3
4	3.57	N/f	N/f	0	4
5	3.62	N/f	N/f	1	4
6	3.83	N/f	N/f	1	8
7	4.01	N/f	N/f	2	9
8	4.27	N/f	N/f	5	15
9	4.90	N/f	N/f	6	32
10	5.73	N/f	N/f	7	55

Table 3. Input parameters:
$m = 3$, $\Delta = 0.1$, $N = 100$, $\lambda = 1$, $\Delta t = 0.1$,
$A_1 = 2$, $A_2 = 2.25$, $A_3 = 3$.

It follows from Table 2 that a small level of clutter violates the applicability of the mean based algorithm, whereas the median based one is still working reliably for a larger range of clutter parameters.

The smaller is location error, the wider is the range of the clutter parameters permitting the applicability of the algorithms.

IV Non-parametric tracks

Consider the model

$$Y_{i1} = f_1(t_i) + \epsilon_{i1},$$
$$Y_{i2} = f_2(t_i) + \epsilon_{i2}, \quad i = 1, \ldots, n.$$

Here $f_1(\cdot) : [0,1] \to \mathbf{R}$, $f_2(\cdot) : [0,1] \to \mathbf{R}$, are unknown functions, $t_i = i/n$, and $\epsilon_{i1}, \epsilon_{i2}$ are independent random variables such that $\epsilon_{11}, \cdots, \epsilon_{n1}$ and $\epsilon_{12}, \cdots, \epsilon_{n2}$ are i.i.d., and $(\epsilon_{11}, \cdots, \epsilon_{n1}) \perp (\epsilon_{12}, \cdots, \epsilon_{n2})$.

We are given the observations $(Y_{11}, Y_{12}), \cdots, (Y_{n1}, Y_{n2})$, but for each pair of values(Y_{i1}, Y_{i2}), we do not know which value is Y_{i1} and which is Y_{i2}. Our problem is to estimate the functions $f_1(\cdot), f_2(\cdot)$.

Assumption 1. $f_1(\cdot), f_2(\cdot) \in \sum(\beta, L)$ where $\sum(\beta, L)$ is the class of all functions such that their $(\lfloor\beta\rfloor)$-derivatives satisfy $\beta - \lfloor\beta\rfloor$-Holder condition with constant L (see e.g. [8]). Denote by $f_{n1}(x), f_{n2}(x)$ the roots of the quadratic equation

$$Z^2 - a_{n1}(x)Z + a_{n2}(x) = 0, \tag{13}$$

if these roots are real, and set $f_{n1}(x) = f_{n2}(x) = 0$ otherwise. Here $a_{ni}(x), i = 1, 2$, are symmetric functions of measurements introduced in (14) such that, under appropriate conditions

$$a_{n1}(x) \xrightarrow{P} f_1(x) + f_2(x),$$
$$a_{n2}(x) \xrightarrow{P} f_1(x) f_2(x) as n \to \infty.$$

Hence the quadratic function in (13) converges in probability to the function

$$F(Z) = Z^2 - (f_1(x) + f_2(x))Z + f_1(x)f_2(x)$$

(uniformly in Z on every bounded interval). It is well-known that the equation F(Z)=0 has the roots $f_1(x)$ and $f_2(x)$. Hence, it can be shown [3] that the roots of (2) converge to $f_1(x), f_2(x)$ in probability as $n \to \infty$.

A more accurate result will be formulated at the end of this section.

Assumption 2. $E(\epsilon_{i1}) = E(\epsilon_{i2}) = 0$, $\sigma_1^2 = E(\epsilon_{i1}^2) < \infty$,$\sigma_2^2 = E(\epsilon_{i2}^2) < \infty$, $E(\epsilon_{i1}^4) < \infty$,$E(\epsilon_{i2}^4) < \infty$, i=1,..., n.

Consider the estimation of f_1, f_2 at arbitrary fixed point $x \in (0,1)$. Define the statistics

$$a_1(x) = \sum (Y_{i1} + Y_{i2}) W_{ni}(x),$$
$$a_2(x) = \sum Y_{i1} Y_{i2} W_{ni}(x), \qquad (14)$$

where $W_{ni}(x)$ is a weight function such that $\sum W_{ni}(x) = 1$ (or $\sum W_{ni}(x)$=1+o(1) as $n \to \infty$). For example, one may take the weights $W_{ni}(x)$=$\frac{K(\frac{t_i - x}{h})}{\sum K(\frac{t_i - x}{h})}$. Here K:$\mathbf{R} \to \mathbf{R}$ is a kernel and $h > 0$. We can use also $W_{ni}(x) = \frac{1}{nh} K(\frac{t_i - x}{h})$.

Theorem. Let Assuptions 1-2 be satisfied, and let $K(\cdot)$ be a kernel of order $l = \lfloor \beta \rfloor$, i.e.

$$\int u^m K(u) du = 0, m = 1, \ldots, l,$$
$$\int K(\mu) d\mu = 1. \qquad (15)$$

Set $h = Cn^{-\frac{1}{2\beta+1}}$ for some $C > 0$. Let supp $K(\cdot)$ be compact. Then, if the kernels satisfy (15) and $|f_1 - f_2|(x) \geq t_0 > 0$, then

$$\sup_{f_j \in \sum} E_{f_1, f_2} (f_{nj}(x) - f_j(x))^2 \leq Cn^{-\frac{2\beta}{2\beta+1}}, j = 1, 2,$$

where $\sum_0(\beta, L) = \sum(\beta, L) \cap \{$ f: $|f(x)| \leq C_0\}$, $C_0 > 0$ is arbitrary.

Fig. 5 and 6 illustrate the simulation of non-parametric fitting of two parallel parabolic trajectories by the method just outlined which can be run in real time using jave-applet at http://www.coe.neu.edu/ milu/ with different parameters of the model (window size h and the noise standard deviation).

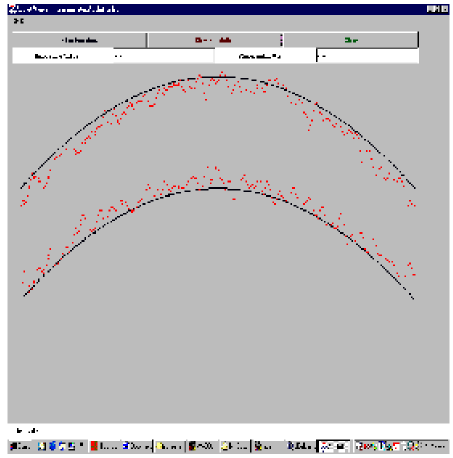

Fig 5: Standard Error: 0.5 Window Size : 0.01

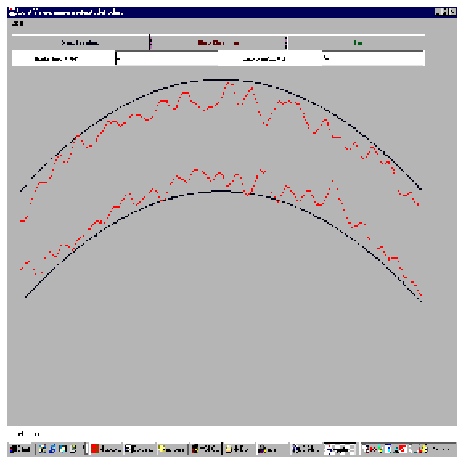

Fig 6: Standard Error: 0.7 Window Size : 0.02

Acknowledgement. The authors are grateful to A.B. Tsybakov and I.I. Tsitovich for giving their permission to describe our joint results in this paper.

References

1. Bar Shalom Y. and X.R. Li. 1995. *Multitarget-multisensor tracking. Principles and techniques*, YBS, Storrs, Connecticut.
2. Bernstein, A.V. 1973. "Group regression parameter estimation." *Izvestia of USSR Academy of Sciences, Technical Cybernetics*: 137-141.
3. Bernstein, A.V. 1998. Statistical analysis of random polynomials. *M*athematical Methods of Statistics, **7**, 274-295.
4. Cramer H. and M.R.Leadbetter. 1967. *Stationary and related stochastic processes. Sample function properties and their applications.* John Wiley, New-York, London, Sydney.
5. Dempster, A.P; N.M. Laird; and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J.Royal Statist. Soc. Ser. B (methodological)* 39: 1-38.

6. Heitjan, D. H. and D. B. Rubin (1991). Ignorability and coarse data, *T*he Annals of Statistics, **19**, 2244–2253.

7. Kiefer, N.M. 1978. Discrete parameter variation: Efficient estimation of a switching regression model, *Econometrica* 46: 427-434.

8. Korostelev, A.P. and A.B. Tsybakov. (1993). *M*inimax theory of image reconstruction, Lecture Notes in Statistics, **82**, Springer, N.Y.

9. Malyutov, M. and M. Gavasheli. 1997. Estimation of parameters of a mixture of (hyper)planes. In: Abstracts, *International conference on Combinatorics, Information Theory and Statistics.* Portland, Portland, Maine,p. 57.

10. Malyutov, M. and I.Tsitovich. 2000. "Modeling multi-target estimation in noise and clutter." Proceedings, 12th European Simulation Symposium, ESS 2000 (Simulation in Industry) Sept. 28-30, Hamburg Germany, Soc. for Computer Simulation, Delft, Netherlands, 598-600.

11. Molnar K.I. and J.W. Modestino. 1998. "Application of the EM algorithm for the multitarget/multisensor tracking problem." *IEEE Transactions on Signal Processing.***46**, 115-129.

12. Neal, R. M. and Hinton, G. E. (1998) "A view of the EM algorithm that justifies incremental, sparse, and other variants", in M. I. Jordan (editor) Learning in Graphical Models, Kluwer Academic Publishers,Dordrecht, 355-368.

13. Nikiforov, A.M. 1991. "Statistical analysis of incomplete data: theory, techniques and software". Supplement to the Russian edition of *Analysis With Missing Data*, R.J.A. Little and D.B. Rubin. Moscow, Finance and Statistics: 284-332 (in Russian).

14. Quandt, R.E. and J.B. Ramsey. 1978. "Estimating mixtures of normal distributions and switching regressions." *Journal of American Statistical Association* 73: 730-738.

15. Serfling, Robert, J. 1980, Approximation theorems of mathematical statistics, New York: Wiley & Sons.

16. Streit R. 1995. Track initialization sensitivity in clutter, In *Proceedings of the Conference on Signal and Data Processing of Small Targets, SPIE International Symposium on Aerospace*, San Diego, CA, SPIE, 460-471.

17. Streit, R., ed. 1998. Studies in probabilistic multi-hypothesis tracking and related topics, Naval Undersea Warfare Center Division, Newport, RI.