ROBUST MODIFICATION OF THE EM-ALGORITHM FOR PARAMETRIC MultiTRAJECTORY ESTIMATION IN NOISE AND CLUTTER

M.B. Malyutov
Department of Mathematics, Northeastern University, Boston, U.S.A.

M. Lu
Department of Biostatistics, The University of Iowa, Iowa City, USA

Abstract. A robust family of algorithms generalizing the EM-algorithm for fitting parametric deterministic multi-trajectories observed in Gaussian noise and clutter is proposed. It is based on the M-estimation generalizing the Maximum Likelihood estimation in the M-step of the EM-algorithm. Simulation results of comparative performance of our and traditional EM-algorithm in noise and clutter are described.

Keywords: Parametric deterministic Multi-trajectory estimation, generalized EM-algorithms, M-estimation, simulation of performance

1. Introduction

Traditional methods of multi-trajectory estimation (MTE) are based on preliminary data association (PDA) (measurements and objects are associated at each frame according to some criterion) [1]). The computational complexity of these methods grows exponentially in the number of frames observed requiring supercomputers. We are also unaware of their consistency when the distances between targets are comparable with the standard errors of measurements. Thus it is likely that whatever amount of data and computation is available, the algorithm resolution cannot be made better than certain non-vanishing limits.

The input data is 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 in [12], [16]. The uncorrelated 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 motion equations etc.

Here we deal only with a posteriori *fitting* the motions' parametric trajectories applicable e.g. to the ballistic missiles' observations in noise and clutter. The chronologically first efficient method of solving this problem for trajectories described by polynomial equations in clutter absence, namely Symmetric Functions of Measurements (SFM), was proposed in seventies for the former Soviet anti-missile defense, see [3].

We describe another approach to fitting non-random trajectories based on a robust modification of the EM-algorithm having a wider applicability range and better accuracy than the SFM-method. A class of measurement errors is generated by the clutter. We assume analogously to [16], that a clutter generating noise ϵ_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 $t \to \infty, 0 < b < 1$. The process ϵ_t generates a Poisson process of false targets (with rate (intensity) depending on Λ, 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 ϵ_t. A time interval for a false target presence on a frame is a positive RV with certain density function

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 similar formulas for the rate of missing and false targets with H_2 instead of H_1.

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 $\frac{1}{2}\pi a t e^{-\frac{\pi a^2 t^2}{4}}$, where $a = \frac{H_2 \lambda}{\sqrt{2\pi}}$ for every target. We keep the E-step of the EM-algorithm as described in [13] and replace the weighted Least Squares (LS) estimation of trajectories parameters by its robust weighted M-version on the M-step. The maximal break-down point of this algorithm is attained by the Least Median of Squares (LMS) estimator [17] which is very slow. The Huber kernel for M-estimator outperforms the LS-estimator in the case of strong clutter, and is comparatively fast. We showed our simulation results at the NATO ASI-2003 in Tsakhkadzor, Armenia. The codes are available from the second author.

2. Generalized EM-Algorithms

2.1. Introduction to the EM algorithm

The estimation-maximization (EM) iterative algorithm was created in sixties for mixtures estimation (by several authors independently) and (by L. E. Baum et al) for the parameter estimation of Hidden Markov Models. Dempster et al [5]) in 1977 reformulated it as a general technique for finding maximum likelihood estimates from incomplete data and gave an erroneous proof of its convergence under general conditions to a local maximum of the likelihood function starting from a sufficiently close initial parameter point, which was corrected in [20] and in more theoretical papers of I. Csiszar and G. Tusnady. Our application is a *considerably simplified version of the Hidden Markov model with a deterministic evolution of the hidden state*. We use a justification of generalized EM-algorithms (GEM) proposed in [15] based on representation of the EM-algorithm as an alternating penalized maximization procedure over parameters and posterior distribution of missing values. The EM algorithm formalizes an old idea for handling missing data: 1) replace missing data values by estimated values, 2) estimate parameters, 3) re-estimate the missing values assuming the new parameter estimates are correct, 4) re-estimate parameters, and so forth, iterating until convergence. Such methods are simple for models where the complete data log likelihood $l(\theta|Y_{obs}, Y_{mis}) = lnL(\theta|Y_{obs}, Y_{mis})$ is linear in Y_{mis}; more generally, missing sufficient statistics rather than individual observations need to be estimated, and even more generally, the log likelihood $l(\theta|Y)$ itself needs to be estimated at each iteration of the algorithm.

Each iteration of the EM consists of an **E** step (expectation step) and a **M** step (maximization step). The M step is simpler to describe : maximize the likelihood for parameter θ given a posterior distribution of the missing data. Thus the **M** step of the EM algorithm uses the weighted ML- estimation estimation. The **E** step fits the conditional expectation of "missing data" given the observed data and current estimated parameters, and then substitutes these expectations for the "missing data". These steps are often easy to construct, to program for calculation, and to fit into computer storage. Also, each step has a direct statistical interpretation. An additional advantage of the algorithm is its steady convergence: under general conditions each iteration increases the log likelihood $l(\theta|Y_{obs})$, and if $l(\theta|Y_{obs})$ is bounded, the sequence $l(\theta|Y_{obs})$ generally converges to a stationary value of $l(\theta|Y_{obs})$. Usually, if the sequence $\theta^{(m)}$ converges, it converges to a local maximum or saddle point of $l(\theta|Y_{obs})$.

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

- $E-$step. Determine $Q(\theta, \theta^{(m)})$, where

$$Q(\theta, \theta^{(m)}) = E\,(\log f(Y_{complete}|\,\theta)|\,Y_{observed}, \theta^{(m)}).$$

- $M-$step. Determine $\theta^{(m+1)} = arg\max_{\theta} Q(\theta|\,\theta^{(m)})$.

An EM algorithm chooses $\theta^{(m+1)}$ to maximize Q($\theta|\theta^{(m)}$) with respect to θ. More generally, a **GEM** (generalized **EM**) chooses $\theta^{(m+1)}$ so that Q($\theta^{(m+1)}|\theta^{(m)}$) is greater than Q($\theta^{(m)}|\theta^{(m)}$). Say, we replace by its robust version. The following is the key straightforward result of [5]:

Every GEM algorithm increases $l(\theta|Y_{obs})$ at each iteration, that is

$$l(\theta^{(m+1)}|Y_{obs}) \geq l(\theta^{(m)}|Y_{obs})$$

with equality if and only if

$$l(\theta^{(m+1)}|Y_{obs}) = l(\theta^{(m)}|Y_{obs})$$

2.2. LMS Estimation for the MTE

We model the motion $x_k[t]$ of k-th object as polynomial $\theta_{\mathbf{k}}[\mathbf{t}]$ of order n, of t, $\theta_{\mathbf{k}}[\mathbf{t}] := (\theta_{\mathbf{k0}}, \ldots, \theta_{\mathbf{kn}}, \mathbf{k} = \mathbf{1}, \ldots, \mathbf{K})$. In the EM algorithm for the MTE, the M-step is reduced to finding $\theta_{j0}, \cdots, \theta_{jn}$ minimizing

$$\sum_{t=1}^{T} \sum_{k=1}^{n_t} w_{kj} (x_k[t] - \theta_j[t])^2, \qquad j = 1, 2, \cdots, K;$$

and we use *sweep* operator described in [2] to find $\theta_{j0}, \theta_{j1}, \cdots, \theta_{jn}$ and σ_j

False targets and missing targets influence the M-step similarly to the contamination dealt with in robust regression techniques. In contrast to [14] we prefer to use robust versions [18] of the M-step instead of fitting a parametric family of the clutter models for their analysis regarding parametric family of clutter not always relevant.

In the LMS estimation for MTE, we find $\theta_{j0}, \theta_{j1}, \cdots, \theta_{jn}$ to minimize

$$\operatorname*{med}_{1 \leq t \leq T} \sum_{k=1}^{n_t} w_{kj} (x_k[t] - \theta_j[t])^2, \qquad j = 1, 2, \cdots, K, \tag{1}$$

To solve (1.2), we use the following scanning algorithm based on consecutive minimizations in the obvious way:

$$\min_{\theta_{jn}} \cdots \min_{\theta_{j0}} \operatorname*{med}_{1 \leq t \leq T} \sum_{k=1}^{n_t} w_{kj} ((x_k[t] - \theta_{jn} t^n) \cdots - \theta_{j0})^2,$$

where $j = 1, 2, \cdots, K$.

For the simplest case n=0, to minimize

$$\operatorname*{med}_{1 \le t \le T} \sum_{k=1}^{n_t} w_{kj}(x_k[t] - \theta_{j0})^2 \tag{2}$$

we use a transformation $\theta_{j0} = \tan\alpha$ and search through all angles α from -1.55 rad to 1.55 rad with a step-size of 0.01 rad to approximate the minimal value in (2), and then we scan with a precision of 0.001 or less rad in the two most promising areas (as determined by the first step).

For n=1, we search in the following way:

$$\min_{\theta_{j1}} \min_{\theta_{j0}} \operatorname*{med}_{1 \le t \le T} \sum_{k=1}^{n_t} w_{kj}((x_k[t] - \theta_{j1}t) - \theta_{j0})^2 \tag{3}$$

Indeed, we can treat the parts in (3) separately. The second portion of the minimization can be solved as n=1 case. Then we have to find $\hat{\theta}_{j1}$ for which

$$m^2(\theta_{j1}) = \min_{\theta_{j0}} \operatorname*{med}_{1 \le t \le T} \sum_{k=1}^{n_t} w_{kj}((x_k[t] - \theta_{j1}t) - \theta_{j0})^2$$

is minimal. This is a minimization of a one-dimensional function $m^2(\theta_{j1})$. In order to find $\hat{\theta}_{j1}$, we use the same techniques as in the case n=1.

For $n > 2$, the scanning would require a tremendous amount of computation.

Finally, we use $\hat{\theta}_{j0}, \hat{\theta}_{j1}, \cdots, \hat{\theta}_{jn}$ to estimate $\hat{\sigma}_j$, $j = 1, 2, \cdots, K$:

$$\hat{\sigma}_j^2 = \operatorname*{med}_{1 \le t \le T} \sum_{k=1}^{n_t} w_{kj}(x_k[t] - \theta_{j0} - \theta_{j0}t - \cdots - \theta_{jn}t^n)^2$$

2.3. M-ESTIMATORS

M-estimators generalize ML-estimators, for the model

$$y_i = g_i(\beta) + \varepsilon_i$$

where Eε_i= 0, Eε_i^2= σ^2, $\beta = (\beta_1, \beta_2, \cdots, \beta_p)$, and $\varepsilon_i' s$ are independent. If $\varepsilon_i \sim$ N(0,σ^2), then $\hat{\beta}_{ML} = \hat{\beta}_{LS}$.

The ML estimator under non-degenerate design

$$\hat{\beta} = \arg\min \frac{1}{2} \sum_{\imath=1}^{n} \left(\frac{y_i - g_i(\beta)}{\sigma} \right)^2$$

is the unique solution of the estimating equations

$$\sum_{i=1}^{n} \frac{\partial g_i(\hat{\beta})}{\partial \beta_j} \left(\frac{y_i - g_i(\hat{\beta})}{\sigma} \right) = 0 \qquad \mathrm{j} = 1, \cdots, \mathrm{p}$$

The problem with the ML estimation is that for objective function $\rho(\mathrm{z}) = \mathrm{z}^2/2$, the least squares minimization weighs heavily large absolute residuals, $|y_i - g_i(\beta)|$. Huber (1964) proposed a more robust approach by using a different ρ- function, which put less weight on extreme residuals. Define $\Psi(\mathrm{z}) = \mathrm{d}\rho(z)/\mathrm{dz}$. The M-estimation

$$\hat{\beta} = \arg\min \sum_{\imath=1}^{n} \rho \left(\frac{y_i - g_i(\beta)}{\sigma} \right)$$

is a solution of estimating equations

$$\sum_{i=1}^{n} \frac{\partial g_i(\hat{\beta})}{\partial \beta_j} \Psi \left(\frac{y_i - g_i(\hat{\beta})}{\sigma} \right) = 0 \qquad \mathrm{j} = 1, \cdots, \mathrm{p}$$

P. Huber proposed

$$\rho_H(z) = \begin{cases} \frac{1}{2} z^2 & \text{if} \quad |z| \leq H \\ H|z| - \frac{H^2}{2} & \text{if} \quad |z| > H \end{cases}$$

and

$$\Psi_H(z) = \frac{d\rho_H(z)}{dz} = \begin{cases} -H & \text{if} \quad z < -H \\ z & \text{if} \quad |z| \leq H \\ H & \text{if} \quad z > H \end{cases}$$

Traditionally, H equals to 1.5 $\tilde{\sigma}$, σ can be estimated as $1.483 med_i \left\{ |\mathrm{y}_i - \mathrm{med}_l \left\{ y_l \right\} | \right\}$

Some alternatives to Huber's ρ function were proposed, say, D.F. Andrews' one:

$$\rho_A(z) = \begin{cases} A^2(1 - \cos(z/A)) & \text{if} \quad |z/A| \leq \pi \\ A^2 & \text{if} \quad \text{otherwise} \end{cases}$$

$$\Psi_A(z) = \begin{cases} A \sin(z/A) & \text{if} \quad |z/A| \leq \pi \\ 0 & \text{if} \quad \text{otherwise} \end{cases}$$

and the J.Tukey's Biweight ρ-function

$$\rho_B(z) = \begin{cases} \frac{1}{2} B^2 [1 - (1 - (z/B)^2)^3] & \text{if} \quad |z| \leq B \\ \frac{1}{2} B^2 & \text{if} \quad \text{otherwise} \end{cases}$$

$$\Psi_B(z) = \begin{cases} z[1 - (z/B)^2]^2 & \text{if} \quad |z| \leq B \\ 0 & \text{if} \quad \text{otherwise} \end{cases}$$

For MTE, we use finding in the M-step $\theta_{j0}, \cdots, \theta_{jn}$ minimizing

$$\sum_{t=1}^{T} \sum_{k=1}^{n_t} w_{kj} \rho_H(x_k[t] - \theta_j[t]) \quad \text{j} = 1, 2, \cdots, \text{K}$$

where ρ_H is the Huber's function, and σ_j j =1 ,2, $\cdots$, K are estimated as

$$\left(\frac{\sum_{t=1}^{T} \sum_{k=1}^{n_t} w_{kj} \rho_H(x_k[t] - \hat{\theta}_j[t])}{T - n - 1} \right)^{\frac{1}{2}}$$

3. Simulation Results

In our simulation, the number of frames (observations) is 70 and the number of trajectories is 3. we assume the missing rates for each trajectory are λ_1= 0.4, $\lambda_2 = 0.3$ and $\lambda_3 = 0.3$ and the standard deviation of noise σ is 0.1. The true trajectories are defined as

$$\begin{aligned} x &= 0.1t, \quad y = 0.2t \\ x &= 0.1t, \quad y = 2 + 0.2t + 0.001t^2 \\ x &= 1 + 0.1t, \quad y = 0.13t + 0.001t^2 \end{aligned}$$

1. Clutter Rate = 0.3

In this case, the number of false measurements is 18, the number of missing data points is 50, and the number of observed data (including the false observations) is 178.

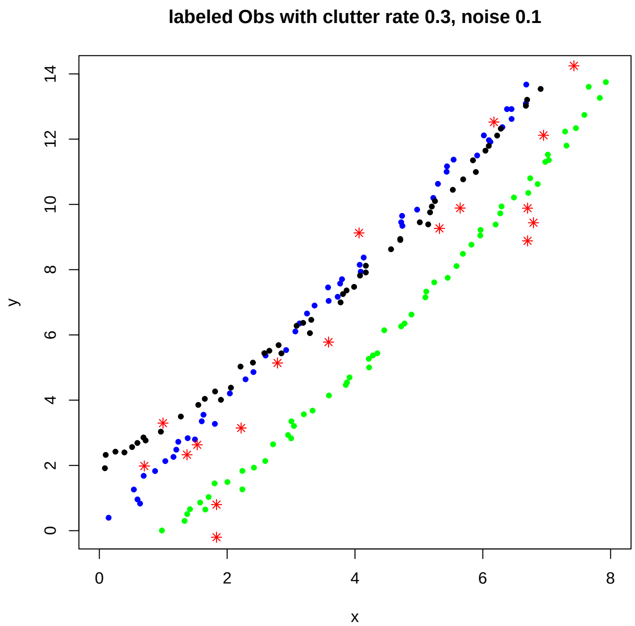

Fig 1a Observations with clutter rate 0.3 and noise 0.1

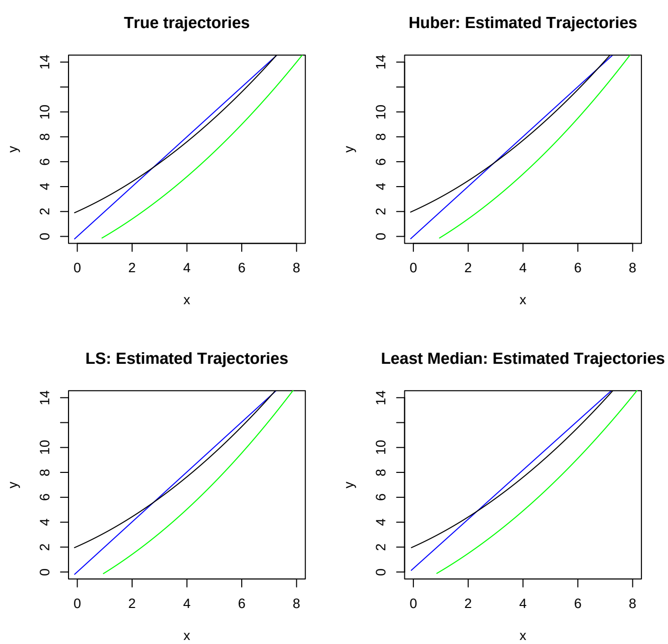

Fig 1b Real and Estimated Trajectories

Feature	True Value	LS	LMS	Huber	missing
x_0	0	0.0017	0.0190	0.0016	19
x_1	0.1	0.0998	0.0973	0.0998	19
x_2	0	0	0	0	19
σ_x	0.1	0.1450	0.0704	0.093	
y_0	0	0.0146	0.3247	0.0152	19
y_1	0.2	0.2002	0.1923	0.2001	19
y_2	0	0	0	0	19
σ_y	0.1	0.0830	0.1062	0.0725	

Table 1a first trajectory

Feature	True Value	LS	LMS	Huber	missing
x_0	0	-0.0056	0.0260	-0.0053	19
x_1	0.1	0.0984	0.0993	0.0983	19
x_2	0	0.0003	0	0	19
σ_x	0.1	0.0876	0.0516	0.0627	
y_0	2	2.0512	2.0554	2.051	19
y_1	0.1	0.0967	0.0993	0.0968	19
y_2	0.001	0.001	0.001	0.001	19
σ_y	0.1	0.1006	0.0558	0.0719	

Table 1b second trajectory

Feature	True Value	LS	LMS	Huber	missing
x_0	1	1.047	0.9543	1.0472	12
x_1	0.1	0.0955	0.1003	0.0955	12
x_2	0	0	0	0	12
σ_x	0.1	0.0774	0.0582	0.0815	
y_0	0	0.0013	-0.02	0.0012	12
y_1	0.13	0.1311	0.1307	0.1311	12
y_2	0.001	0.001	0.001	0.001	12
σ_y	0.1	0.0840	0.0754	0.0719	

Table 1c third trajectory

2. Clutter Rate = 1.1

In this case, the number of false measurements is 79, the number of missing data points is 50, and the number of observed data (including false ones) is 239. The mean-based estimation breaks down, if the missing rate is greater than 0.3

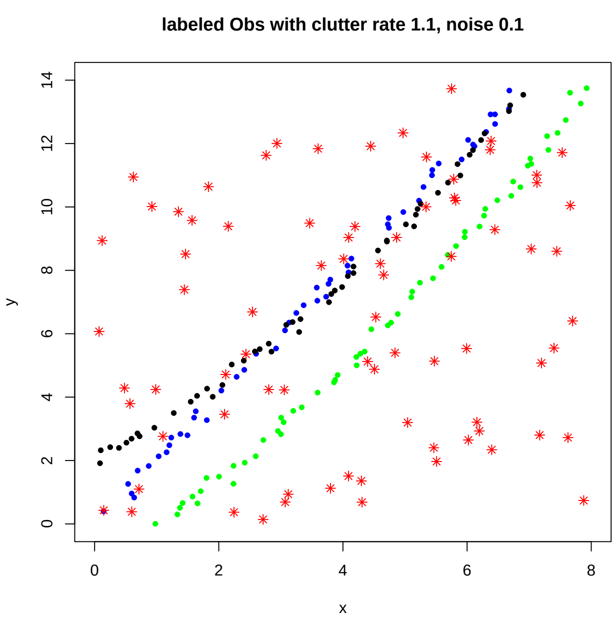

Fig 2a Observations with clutter rate 1.1 and noise 0.1

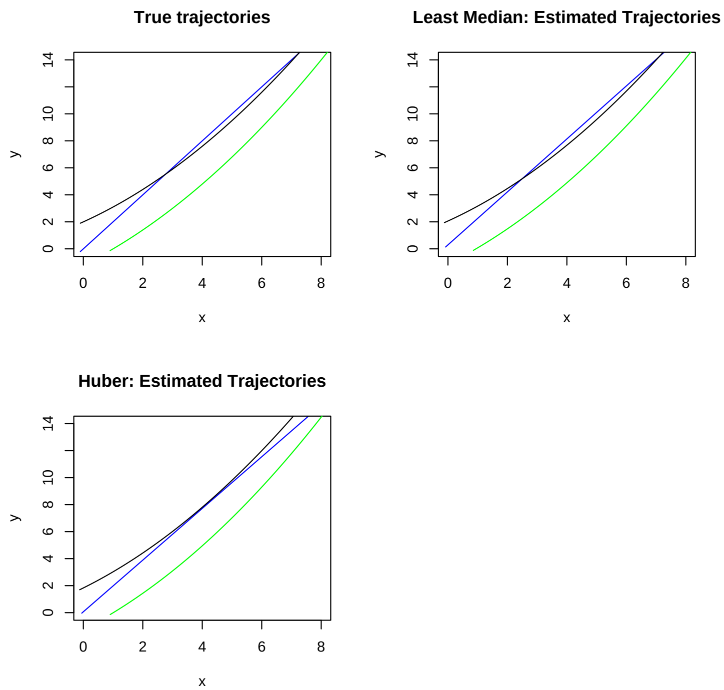

Fig 2b Real and Estimated Trajectories

Feature	True Value	LS	LMS	Huber	missing
x_0	0	N/A	-0.007	0.048	19
x_1	0.1	N/A	0.1003	0.098	19
x_2	0	N/A	0	0	19
σ_x	0.1	N/A	0.1158	0.1901	
y_0	0	N/A	0	0.1572	19
y_1	0.2	N/A	0.2006	0.1914	19
y_2	0	N/A	0	0	19
σ_y	0.1	N/A	0.0777	0.2897	

Table 1a first trajectory

Feature	True Value	LS	LMS	Huber	missing
x_0	0	N/A	-0.010	-0.03	19
x_1	0.1	N/A	0.1003	0.1012	19
x_2	0	N/A	0	0	19
σ_x	0.1	N/A	0.0737	0.1001	
y_0	2	N/A	1.9794	1.807	19
y_1	0.1	N/A	0.1003	0.109	19
y_2	0.001	N/A	0.001	0.001	19
σ_y	0.1	N/A	0.0963	0.238	

Table 1b second trajectory

Feature	True Value	LS	LMS	Huber	missing
x_0	1	N/A	0.9741	1.002	12
x_1	0.1	N/A	0.0993	0.098	12
x_2	0	N/A	0	0	12
σ_x	0.1	N/A	0.0741	0.079	
y_0	0	N/A	0.0340	-0.0298	12
y_1	0.13	N/A	0.1287	0.1315	12
y_2	0.001	N/A	0.001	0.001	12
σ_y	0.1	N/A	0.1042	0.096	

Table 1c third trajectory

The following table shows the running time for the three versions of the M-step under different rates of clutters

Clutter Rate	Clutter Num	LS	M-estimation	LMS
0.3	18	7s	8s	24m 32s
1.1	79	N/A	12s	26m 50s

In the simulation, the LS algorithm breaks down if the the rate of clutter is more than 0.3, that is, 18 clutters in the observations. The M-estimation won't be good enough if the rate of the clutter is more than 1.1. However, the LMS still works at the rate of 1.3.

4. Conclusion

Our simulation results confirmed our theoretical expectation that the median-based estimation and M-estimation are more robust than the mean-based estimation. With a small clutter rate, the mean-based,median-based estimation and M-estimation all work pretty well. However, the mean-based estimation breaks down, if a large number of false measurements appears in the data. The median-based estimation and M-estimation continues to work robustly under high clutter rate. The median-based estimation method is very time-consuming, thus to work on line it must be replaced with a faster implemented M-estimation block, say, the one based on Huber kernel.

The future work for the median-based GEM is to improve the "scan" algorithm to reduce the running time of the algorithm.

References

1. Bar Shalom Y. and X.R. Li. 1995. *Multitarget-Multisensor Tracking. Principles and Techniques*, YBS, Storrs, Connecticut.

2. Beaton,A.E. The use of special matrix operations in statistical calculus, *Educational Testing Service Research Bulletin*, RB-64-51, 1964.

3. Bernstein, A.V. Group regression parameter estimation. *Izvestia of USSR Academy of Sciences, Technical Cybernetics*: 137-141, 1973.

4. Cramer H. and M.R.Leadbetter. 1967. *Stationary and related stochastic processes. Sample function properties and their applications.* John Wiley, New-York, London, 1973.

5. Dempster, A.P; N.M. Laird; and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J.Royal Statist. Soc. Ser. B (methodological)* 39: 1-38, 1977.

6. Donoho, D.L. and Huber, P.J. The notion of breakdown point, in *Festschrift for Erich Lehmann*, edited by P. Bickel, K. Doksum, and J.L. Hodges, Jr.,Wadsworth, Belmont, CA, 1983.

7. Frank R. Hampel, Elvezio M, Ronchetti, Peter J. Rousseeuw and Werner A. Stahel. *Robust Statistics The Approach Based on Influence Functions*. John Wiley & Sons, 1986.

8. Huber, P.J.(1964) Robust estimation of a location parameter. *Ann. Math. Statist.* 35, 73-101

9. Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. Preceedings of Fifth Berkley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 221-233

10. Huber, P.J.(1972) Robust statistics : A review. *Ann. Math. Statist.* 43, 1041-1067

11. Huber, P.J.(1977c) Robust methods of estimation of regression coeffcients. *Math. Operationsforschung Statist. Ser. Statist* 8, 41-53

12. Malyutov, M. and I.Tsitovich. 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, 2000.

13. M. B. Malyutov, M. Lu, A. Nikiforov and R. Protassov . Multi-Target estimation in noise and clutter, *Proceedings, International Conference Fusion 2001, August 2001, Montreal, Canada*, Fr.C, 17-25.

14. Molnar K.I. and J.W. Modestino. Application of the EM algorithm for the Multitarget/Multisensor Tracking Problem. *IEEE Transactions on Signal Processing.***46**: 115-129, 1998.

15. Neal, R. M. and Hinton, G. E. 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, 1998.

16. Streit R. 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, 1995.

17. Rousseeuw, P.J., and Yohai,V. Robust regression by means of S-estimator, in *Robust and nonlinear Time Series Analysis*, edited by J.Franke, W.Hardle, and R.D.Martin, Lecture Notes in Statistics No.26. Springer Verlag, New York, pp 256-272, 1984.

18. Peter J.Rousseeuw, and Annick M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, 1987.

19. Streit, R., ed. Studies in probabilistic Multi-Hypothesis Tracking and Related Topics, *Naval Undersea Warfare Center Division*, Newport, RI, 1998.

20. Wu, C.F.J.(1983). On the convergence properties of the EM algorithm. *Ann. Statistics.***11**, 95-103.