A Sufficient Comparison of Trackers

David Bizup
University of Virginia
Department of Systems and Information Engineering
P.O. Box 400747
151 Engineer's Way
Charlottesville, VA 22904

Donald E. Brown
University of Virginia
Department of Systems and Information Engineering
P.O. Box 400747
151 Engineer's Way
Charlottesville, VA 22904

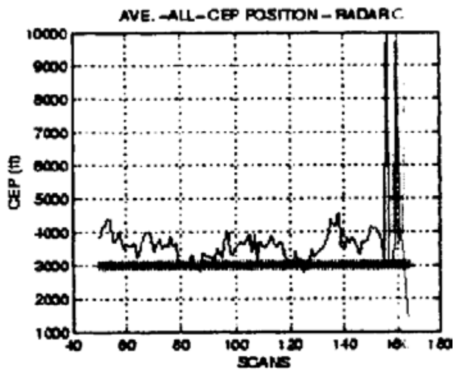
Abstract

Tracking maneuvering targets with radar is a difficult problem, but making a fair comparison between two or more maneuvering target trackers may be even more difficult. At the very least, it is tried less often. In this paper we present a method for comparing trackers based on the probabilistic notion of sufficiency. The advantages of our approach are twofold. First, comparisons are made across tens of thousands of trajectories, not just a few. Second, if one tracker is sufficient for another then it is better no matter how better is defined. We demonstrate the sufficient comparison technique for two trackers; one sets noise levels adaptively based on a statistic of accelerations first introduced at the 7th International Command and Control Research and Technology Symposium, the other is the well known and widely used Interacting Multiple Model.

Introduction

Many papers compare maneuvering target trackers for at most a few different measures of performance and trajectories. Bar-Shalom and Li [Bar-Shalom 1993], de Feo, Graziano, Migliolo, and Farina [deFeo 1997], and Kameda, Tsujimichi, and Kosuge [Kameda 2002] are typical examples. Some report one measure of performance: position accuracy. Some report position and speed accuracies. Averages and interval estimates are sometimes reported, but it is not unusual to see these statistics reported for a single trajectory and maybe even a single simulation run. Other measures of performance like heading and range rate accuracy are rarely seen. The reader wonders why the authors choose particular trajectories, simulation runs, and measure of performance. Or conversely, why they did not choose others. Were they chosen at random, because the results are typical, or because they favor one tracker over the other?

Occasionally, different papers report results for the same tracker. This happens in [Schutz 1997], [Schutz 1999] and [Kirubarajan 1999]. [Schutz 1997] describes what the authors call a Combined Kalman Filter (CKF) tracker. It has one mode, switches process noise levels based on a statistical threshold test of the position measurement residual history, and is intended for a military airborne early warning and control system. [Kirubarajan 1999] compares that tracker to an interacting multiple model (IMM) intended for the same application. Both papers test their trackers against the same 120 target simulation. The papers have some common authors. In fact, the company that



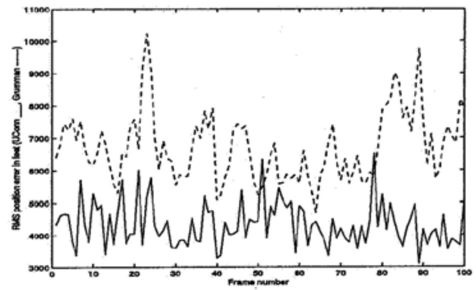


Figure 4: RMS position errors
(— IMM/Assignment, - - - CKF/JVC).

Figure 1. Position error statistics for the same 120 target scenario. (left) from [Schutz 1999], (right) [Kirubarajan 1999]

employs the authors of [Schutz 1997] and [Schutz 1999] partially funded the research in [Kirubarajan 1999].

Figure 1 shows two graphs reporting root mean squared position error. The graph on the left is for the CKF [Shutz 1999], the one on the right for the CKF and IMM [Kirubarajan 1999]. The solid line on the right is IMM error data, the dashed line is CKF data. The IMM's errors average about 4000 ft, the CKF's about 7000 ft. The IMM appears better.

Notice that the left hand graph in figure 1 shows that the CKF errors average about 3500 ft. This is about one half the value reported in [Kirubarajan 1999]. Now, compare these errors to the IMM and the CKF appears better. The same phenomena occur with speed errors. The IMM has lower speed errors according to [Kirubarajan 1999], but higher errors when compared to the original results. The reader wonders not only why the authors choose particular trajectories and statistics, but how the CKF errors double or triple from one day to the next against the same simulation.

There is obviously a need to compare trackers fairly, across many trajectories, and against more measures of performance than just speed and position. In this paper, we present a methodology for making this comparison and demonstrate it for two different trackers. First, we state and interpret the definitions and theorems necessary to establish the desirability of using the sufficiency relation. Second, we define tens of thousands of maneuvering target trajectories. Third, using simulation, we generated error statistics for two trackers against each trajectory. The data are analyzed and found to conform with the assumptions for a sufficient comparison. Finally, we make a sufficient comparison against seven measures of performance and discover not only how often one tracker is better than the other, but also how much better.

Sufficiency

Blackwell [Blackwell 1953] introduced sufficiency and sufficiency comparisons in 1953. The probabilistic notion of a sufficient estimator is analogous to the more familiar notion of a sufficient statistic. If statistic *X* is a sufficient statistic for an unknown parameter θ, then any other statistic *Y* contains no more information about the parameter than *X* and can be interpreted as *X* plus some noise. In the same way, given probabilistic estimators *A* and *B* for a predictand ω, if *A* is sufficient for *B* then *B* contains no more information about the predictand than *A* and can be interpreted as *A* plus some noise.

The analogy is imperfect. A sufficient statistic summarizes all information about a parameter in the data. A statistic is either sufficient or not, and statistical sufficiency does not depend on any other statistic. Probabilistic sufficiency implies only that *A* contains at least as much information as *B*. There is no such thing as a sufficient estimator because probabilistic sufficiency is a relation between estimators. Probabilistic sufficiency establishes a partial order so it is possible that A is sufficient for *B*, or *B* for *A*, or neither, or both. It is also possible that there exists another estimator *C* that is sufficient for both *A* and *B*.

We now state two definitions and two theorems about sufficiency. We assume that the reader is familiar with Bayesian decision theory. The theorems are stated without proof.

Definition 1 (*Informativeness*) It is said that estimator *A* is more informative than estimator *B*, denoted *AIB*, if $R_A \leq R_B$ for all prior distributions and all loss functions, where R_i is the Bayes Risk of estimator *i*.

If we can show that *AIB* then our Bayes risk of making a decision based on *A* is lower than our risk of making a decision based on *B*, no matter what we already know and no matter how we calculate our losses. In other words, *A* is "better" than *B* however "better" is defined. Unfortunately, this definition does not provide a constructive way to find the relation *I*, or even show that it exists.

Definition 2 (*Sufficiency*) It is said that estimator *A* is sufficient for estimator *B*, denoted *ASB*, if there exists a stochastic transformation ψ:B×A→Re such that the following are satisfied, where ω denotes the predictand or true state of nature.

$$f_B(b|\omega) = \int_A \psi(b|a) f_A(a|\omega)da, \text{ for all } \omega \text{ in } \Omega, \text{ and all b in B}$$
$$\psi(b|a) \geq 0$$
$$\int_B \psi(b|a)db = 1$$

Intuitively, if ASB then *B* equals *A* plus some noise ψ, as diagrammed in figure 2. Now we state two theorems. The first states that sufficiency implies informativeness, the second provides a constructive way to determine sufficiency.

Theorem 1 (*Sufficiency => Informativeness*) If *ASB*, then *AIB*.

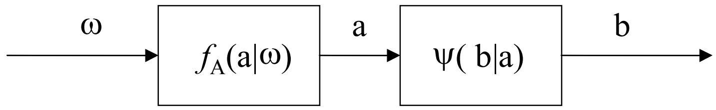


Figure 2. If *ASB*, then interpret *B* as *A* plus noise.

Theorem 2 (*Gaussian Sufficiency*, [Krzysztofowicz 1987]) If the estimators are univariate and have Gaussian likelihoods, that is, $f(x|\omega)=N(\alpha\omega+\beta,\sigma^2)$, then *ASB* if and only if

$$\sigma_B^2 / \alpha_B^2 \geq \sigma_A^2 / \alpha_A^2$$

Notice that the location parameter β has no effect on the association relation. That is, the estimator bias, so long as it is known, is irrelevant to the quality of the estimates. If the parameter α is unity then sufficiency can be determined by a simple comparison of variances. In this case the ratio of their variances indicates the relative quality of the estimators: *ASB* if and only if $\sigma_B^2 / \sigma_A^2 \geq 1$. The larger the ratio, the more improvement *A* offers over *B*.

We now have all the mathematical machinery we need. If the tracker's error distributions for the chosen measures of performance are Gaussian then: by theorem 2 we determine the sufficiency relation by comparing variances; by theorem 1 we know that the sufficient tracker is more informative; and by definition 1 we conclude that the more informative tracker is better.

Defining Trajectories and Generating Error Statistics

We define maneuvering trajectories based on the type of acceleration and the relevant parameters. Simulated air targets fly at constant velocity for three hundred seconds, maneuver for twenty, then fly at constant velocity for another one hundred fifty. We identify three types of maneuvers: coordinated turns, linear accelerations, and altitude changes. Maneuver parameters and their test levels are shown in the following tables. There are 17,280 different coordinated turns, 1,620 linear accelerations, and 1,980 ascents and descents. The total number of different trajectories in the analysis is almost 21,000.

Table 1. Parameters for coordinated turns

Parameter	**Levels**	**Number of Levels**
Range from radar at start of maneuver	20, 40, …, 300 miles	15
Initial heading	0, 30, …, 330 degrees	12
Initial speed	100, 200, …, 800 knots	8
Centripetal acceleration	0, 1, …, 5 G	6
Turn direction	Left, Right	2

Table 2. Parameters for linear accelerations

Parameter	Levels	Number of Levels
Range from radar at start of maneuver	20, 40, …, 300 miles	15
Initial heading	0, 30, …, 330 degrees	12
Initial speed	500 knots	1
Linear acceleration	-1, -0.75, ..., 1 G	9

Table 3. Parameters for ascents and descents

Parameter	Levels	Number of Levels
Range from radar at start of maneuver	20, 40, …, 300 miles	15
Initial heading	0, 30, …, 330 degrees	12
Initial speed	500 knots	1
Ascent Rate	-500, -400, …, 500 knots	11

Using the radar model from [Schutz 1997], we generate 100 sets of simulated data for each trajectory and turn them through four different tracking algorithms. We collect error data for seven measures of performance at times just before the maneuver starts, at ten and twenty seconds into the maneuver, and at ten, twenty and thirty seconds after the maneuver ends. Thus, we use more than three billion data points in the anlaysis.

The first tracker is a Kalman filter that switches between nonmaneuvering and maneuvering modes based on the statistical test of the position residuals defined in [Schutz 1997]. The second uses the range rate measurement to calculate a statistic of acceleration, and switches based on a threshold test of that statistic. Bizup [Bizup 2002] introduced the statistic of accelerations at the 7th ICCRTS. The third tracker adaptively sets the noise levels based on the statistic of accelerations. The fourth is a two mode IMM described by Bar-Shalom and Li [Bar-Shalom 1995].

The intent of collecting data just before the maneuver starts is to compare trackers during long periods of constant velocity motion. During the maneuver, the obvious intent is to compare performance while the target is accelerating. After the maneuver, the intent is to see how quickly the trackers converge back to constant velocity performance. We collect error statistics for gross position, speed, heading, range and bearing from the radar to the target, range rate and cross range rate. The intent is to capture all relevant measures, not just one or two.

Results

During the analysis it became clear that the error statistics did not conform to the assumptions of theorem 2. Their variances were not the same for every trajectory. For example, position error variance increased with range from the radar. This is perfectly reasonable. The same bearing errors in degrees equate to larger position errors in miles as range increases. Further analysis showed that within any given trajectory, the error statistics do conform to the assumptions of theorem 2; they are approximately Gaussian with location parameter $\beta=0$ and scale parameter $\alpha=1$.

We could not determine overall sufficiency so we determined the sufficiency for each measure of performance within each trajectory. Then, we counted the number of cases where one tracker was sufficient for another and reported the proportion. If tracker *A* is sufficient for tracker *B* for most trajectories, then we concluded that it should be better most of the time in real environments. Knowing that tracker *A* is sufficient for tracker *B* tells us nothing about how much better *A* is than *B*. So, in addition to the sufficiency relation, we also report the ratio of the average variances.

The following tables summarize the sufficiency relations and variance ratios for the adaptive tracker and the IMM, averaged over all coordinated turns. The data used to generate the statistics in the first row were collected just before the maneuver started, after a long period of constant velocity flights. The second and third rows are based on data collected ten and twenty seconds into the maneuver. The last three rows on data collected ten, twenty, and thirty seconds after the maneuver ends. There is one column per measure of performance.

Table 4. Proportion of coordinated turns where Adaptive Tracker is sufficient for the IMM.

Time	2D Pos	Speed	Heading	Range	Bearing	Rng Rt	X-Rng Rt
0	0.9938	0.9938	0.9750	0.9938	1.0000	0.9938	1.0000
10	0.9875	0.9812	0.9750	0.9063	1.0000	0.9000	1.0000
20	1.0000	0.9938	0.9750	0.9000	1.0000	0.8750	1.0000
30	0.9938	0.9063	0.9437	0.9375	1.0000	0.9000	0.9313
40	0.9375	0.8500	0.8875	0.9812	0.9625	0.9313	0.8500
50	0.8750	0.8500	0.9250	0.9313	0.8625	0.9500	0.8500

Table 5. Error variance ratio, IMM : Adaptive tracker

Time	2D Pos	Speed	Heading	Range	Bearing	Rng Rt	X-Rng Rt
0	2.1389	2.0861	1.3792	2.7807	2.2701	2.3414	2.5977
10	1.8145	1.7185	1.3794	1.0231	1.9893	0.8957	2.3956
20	1.9385	1.6507	1.3863	1.0990	2.5884	0.7674	2.2729
30	1.5389	1.1635	1.1375	1.2705	1.9991	0.9436	1.2503
40	1.1658	0.8484	1.0844	1.3650	1.2327	1.1546	0.8440
50	0.9768	0.8024	1.1548	1.1683	0.9796	1.2984	0.8053

Consider the number 0.9938 in table 4., located in the cell for 2D position error at time 0. This means that the single mode adaptive tracker position estimates are sufficient for the IMM position estimates 99.4% of the time that the target is not maneuvering. The number in the next row, 0.9875, means that the adaptive tracker's position estimates are sufficient for the IMM position estimates ten seconds into 98.8% of all simulated coordinated turns. As the target continues to maneuver and then resumes constant velocity motion, the numbers in the following rows fluctuate. But, at all times, the adaptive tracker's position estimates are sufficient for the IMM's for at least 87.5% of the coordinated turns.

Now consider the same cells in table 5. These are ratios of the IMM error variances to adaptive tracker error variances. Just before the maneuver starts, IMM position error variances are about twice as large as the adaptive tracker's. They are always larger, on average, during the maneuver. Twenty and thirty seconds after the maneuver they are about equal. Eventually the adaptive tracker will start to dominate the IMM again.

Similar tables for other maneuvers, or conditional on the trajectory parameters were generated with similar results. The only time the IMM was usually sufficient for the adaptive tracker rather than the other way around was when the target was heading directly toward the radar. For all other trajectories, the adaptive tracker was sufficient for the IMM 80% of the time or more.

Conclusion

We observe a deficiency in the tracking literature: most comparative studies report only a few error statistics for only a few different trajectories. We propose a new methodology in which tens of thousands of trajectories are defined. Simulation is used to generate error statistics for several trackers for every trajectory. Trackers are compared using the probabilistic notion of sufficiency. Sufficient comparisons are a powerful, mathematically rigorous way to show that one tracker is better than another however we define better. We demonstrate the new methodology for an adaptive tracker and an interacting multiple model against coordinated turns, and find that the adaptive tracker is almost always better.

References

[Bar-Shalom 1993] Bar-Shalom, Y. and Li, X., *Design of an Interacting Multiple Model Algorithm for Air Traffic Control*, IEEE Transactions on Control Systems Technology, vol. 1, no. 2, pp186-194 (Sep 1993)

[Bar-Shalom 1995] Bar-Shalom, Y. and Li, X., *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishers (1995)

[Bizup 2002] Bizup, D., *A Centripetal Acceleration Statistic for Tracking Maneuvering Targets with Radar*, Proceedings of the 7th International Command and Control Research and Technology Symposium (2002)

[Blackwell 1953] Blackwell, D., *Equivalent Comparisons of Experiments*, Annals of Mathematical Statistics, vol. 24, pp265-272 (1953) [deFeo 1997] de Feo, M., Graziano, A., Migliolo, R. and Farina, A., *IMMJPDA versus MHT and Kalman Filter with NN correlation: performance comparison*, IEEE Proceedings - Radar, Sonar and Navigation, vol. 144, no. 2, pp49-56 (April 1997)

[Kameda 2002] Kameda, H., Tsujimichi, S. and Kosuge, Y., *Target Tracking Using Range Rate Measurement Under Dense Environments*, Electronics and Communications in Japan, Part I, Communications, vol. 85, pp19-29 (2002)

[Kirubarajan 1999] Kirubarajan, R., Bar-Shalom, Y., McAllister, R., Schutz, R. and Engleberg, B., *Multitarget Tracking Using an IMM Estimator with Debiased Measurements for AEW Systems*, Proceedings of the Second International Conference On Information Fusion, vol. 1, pp262-269 (1999)

[Krzysztofowicz 1987] Krzysztofowicz, R., *Markovian Forecast Processes*, Journal of the American Statistical Association, vol. 82, no. 397, pp31-37 (1987)

[Schutz 1997] Schutz, R., McAllister, R., Engleberg., B., Maone, V., Helm, R., Kats, V., Dennean, C., Soper, W., and Moran, L., *Combined Kalman Filter and JVC Algorithms for AEW Target Tracking Applications*, SPIE Conference on Signal and Data Processing of Small Targets vol. 3163, pp164-175 (July 1997)

[Schutz 1999] Schutz, R., Engleberg., B., Soper, W., and McAllister, R., *ManeuverTrackingAlgorithmsforAEWTargetTrackingApplications*, SPIE Conference on Signal and Data Processing of Small Targets vol. 3809, pp308-319 (July 1999)