

Evaluating the Consistency of Estimation

Pavel Ivanov
HERE, a Nokia business
Tampere, Finland
pavel.ivanov@here.com

Simo Ali-Löytsy
Tampere University of Technology
Tampere, Finland
simo.ali-loytsy@tut.fi

Robert Piché
Tampere University of Technology
Tampere, Finland
robert.piche@tut.fi

Abstract—The error covariance reported by an estimation is said to be consistent if it is a reliable indicator of the actual error. In this paper several types of consistency are defined, and methods for its evaluation are introduced. Mean Squared Deviation consistency is based on the Chebyshev inequality, p equivalence is based on the fact that the concentration ellipse with probability mass p must contain the actual value of the estimated parameter with probability p , and Normalized Deviation Squared (NDS) consistency implies that a concentration ellipse of probability mass p contains the actual value of the estimated parameter with probability at least p . Hypothesis tests for consistency evaluation are presented. The NDS consistency test is applied to WiFi localization system data in order to investigate sources of inconsistencies and adjust parameters of the system. It is shown that underestimated measurement noise is the main cause of inconsistent behavior; however, an incorrect motion model or underestimated process noise might also result in inconsistent estimates.

I. INTRODUCTION

An estimate of an unknown parameter can be represented in different ways: by a point estimate and a confidence region around it, by a point estimate and a mean squared error matrix, or by a probability distribution. It is desirable, especially for statistical sensor fusion, that the actual value of the estimated parameter falls within the confidence region of probability p that is declared by the estimate with probability at least as large as p . An estimate that meets this requirement is considered to be consistent.

Any estimate can be made consistent by ad-hoc raising of its uncertainty, however this would make the estimate less informative. Thus, there is a trade-off between consistency and information content.

Consistency is an important property especially in filtering, where the current estimate is used as a prior distribution for future estimation. If an estimate is not consistent then it is overly "optimistic" about its precision, and new measurements have too little influence. In this case the filter can "get stuck" in an erroneous state.

Several definitions of consistency and ways to evaluate it are presented in the literature. Lefebvre et al. [1] define an estimate as consistent if its covariance matrix is larger than the actual mean estimation error matrix of the estimate. Van der Heijden [2] defines consistency based on the fact that for any univariate continuous random variable X with cumulative distribution function F , $F(X)$ has a standard uniform distribution, and transformation of multivariate random variable (RV) to multivariate standard uniform RV

[3]. Bar-Shalom and Li [4] consider consistency of Gaussian estimates in particular and introduce the corresponding normalized estimation error squared (NEES) and normalized innovation squared (NIS) consistency checks: a Gaussian estimate is considered as consistent if the corresponding normalized estimation error follows a χ^2 distribution. Scalzo et al. [5] extend the NEES/NIS tests to non-linear systems and suboptimal filters (e.g. particle filter) by approximating the posterior distributions with Gaussian distributions. Nurminen et al. [6] define an estimate as 95% consistent if the actual value of the estimated parameter falls in the 95% concentration ellipse.

In this paper, several consistency definitions are considered, and hypothesis tests for consistency evaluation are presented. The rest of the paper is organized as follows. In Section 2 consistency and information content of the estimate are defined. In Section 3 hypothesis tests for consistency evaluation are presented. In section 4 filter consistency, and hypothesis tests for its evaluation are discussed. Practical applications are presented in Section 5.

II. CONSISTENCY AND INFORMATION CONTENT

Two important criteria that must be considered for estimator evaluation are consistency and information content. Consistency of the estimate reflects how well the estimated probability distribution of the parameters agrees with its true distribution. Information content of the estimate reflects its certainty or precision, i.e. how well the probability is concentrated.

Consistency and information content are interdependent properties of the estimate. Improving of consistency might lead to the loss of information content, and conversely, making the estimate more informative might make it inconsistent. Consistency is regarded as the more important since underestimated errors' magnitude might badly affect stability of the filter (it might "get stuck").

A. Consistency

There are several ways to define consistency depending on the format of the estimate uncertainty, e.g. probability distribution, covariance matrix or confidence region. In this section, two consistency definitions that have been proposed in the literature and one novel consistency definition are

presented.

1) *Consistency in Mean Squared Deviation*: Consider N -variate random variable X with mean and covariance. The estimate of X defined by $\bar{x} \in R^N$ (which is not necessarily the estimated mean of X) and symmetric semi-definite matrix \bar{P} is called consistent [1] if

$$M \leq \bar{P}, \quad (1)$$

where $M = E[(X - \bar{x})(X - \bar{x})^T]$ is the mean squared deviation matrix of X with respect to \bar{x} , and the symbol \leq means that the matrix difference $\bar{P} - M$ is a positive semi-definite matrix. Both M and \bar{P} are assumed to be nonsingular.

If (1) holds then

$$\begin{aligned} \forall \epsilon > N & : \Pr\{(X - \bar{x})^T \bar{P}^{-1} (X - \bar{x}) \leq \epsilon\} \\ & \geq \Pr\{(X - \bar{x})^T M^{-1} (X - \bar{x}) \leq \epsilon\} \end{aligned}$$

Moreover, according to the modified Chebyshev's inequality [7]

$$\forall \epsilon > N : \Pr\{(X - \bar{x})^T M^{-1} (X - \bar{x}) \leq \epsilon\} \geq 1 - \frac{N}{\epsilon}. \quad (2)$$

Therefore, consistency in mean squared deviation implies that

$$\forall \epsilon > N : \Pr\{(X - \bar{x})^T \bar{P}^{-1} (X - \bar{x}) \leq \epsilon\} \geq 1 - \frac{N}{\epsilon}, \quad (3)$$

i.e. the probability of deviation of the actual parameter from its estimate agrees with the theoretical lower bound given by Chebyshev's inequality.

2) *P Equivalence*: \tilde{X} is a p equivalent estimate of X if

$$\begin{aligned} \exists \epsilon > 0 : \Pr\{(X - \bar{x})^T \bar{P}^{-1} (X - \bar{x}) \leq \epsilon\} \\ = \Pr\{(\tilde{X} - \bar{x})^T \bar{P}^{-1} (\tilde{X} - \bar{x}) \leq \epsilon\} = p \end{aligned} \quad (4)$$

p equivalence is a straightforward generalization of the consistency definition given in [6], where an estimate is said to be 95% consistent if the actual value of the estimated parameter falls inside the 95% concentration ellipsoid. Requirement (4) implies that the concentration ellipse of \tilde{X} containing probability mass p contains the same probability mass of X .

3) *Normalized deviation squared consistency*: An N -variate random variable \tilde{X} with mean $\bar{x} \in R^N$, and covariance matrix \bar{P} is a normalized deviation squared (NDS) consistent estimate of N -variate random variable X if

$$\begin{aligned} \forall \epsilon > 0 : \Pr\{(X - \bar{x})^T \bar{P}^{-1} (X - \bar{x}) \leq \epsilon\} \\ \geq \Pr\{(\tilde{X} - \bar{x})^T \bar{P}^{-1} (\tilde{X} - \bar{x}) \leq \epsilon\} \end{aligned} \quad (5)$$

NDS consistency implies that the concentration ellipse of any probability mass of \tilde{X} contains a larger or equal probability mass of X .

B. Information content

Information content reflects how certain the estimation is about the parameter: the higher the certainty of the estimate, the more valuable is its information. Information content is proportional to the inverse of the covariance matrix of an estimate. In the scalar case, the smaller the posterior's variance is, the more certain is the estimation.

It is possible to make an estimate consistent by increasing its covariance, but this would reduce the informativity of the estimation. It is advantageous to make the estimate as informative as possible and at the same time to preserve consistency. This means that the inequalities (1) and (5) in the consistency definitions should be as close to equalities as possible. In the case of mean squared deviation consistency, an estimate defined by \bar{x} and \bar{P} is most informative and yet consistent in mean squared deviation if $\bar{P} = M$. In the case of NDS consistency, an estimate \tilde{X} is most informative and yet NDS consistent if (5) becomes an equality.

III. FILTER CONSISTENCY

A filter can be considered as consistent if it provides consistent system state estimates. In order to check whether consistency requirements are met, the actual distributions of the system states are needed. In a practical application the only information about the true distribution is provided by a sample drawn from it, which is the actual state of the system. For example, in a field test setting these samples are obtained using additional high-accuracy reference measurement systems. If only a sample is available, hypothesis testing can be used to evaluate the consistency of the estimate, with the null hypothesis H_0 being "The estimate is consistent", and the test statistic being a function of the sample. In filtering, the estimated state is changing in time, and for each time step it is possible to get a sample whose size rarely exceeds 1. In order to increase the power of the hypothesis test, instead of considering one estimate at a time, several consecutive estimates and their respective samples (actual states) are considered for consistency evaluation.

In the following we consider system states at M (not necessarily different) times modeled as N -variate random variables $[X_1, \dots, X_M]$, state estimates $[\tilde{X}_1, \dots, \tilde{X}_M]$ with known probability distributions, respective means $[\bar{x}_1, \dots, \bar{x}_M]$ and covariance matrices $[\bar{P}_1, \dots, \bar{P}_M]$ provided by the filter, and true states x_1, \dots, x_M which can be considered as realizations of $[X_1, \dots, X_M]$. We assume that x_1, \dots, x_M are independent realizations of $[X_1, \dots, X_M]$ even though they are correlated. This is done in order to increase the power of the hypothesis test without increasing the number of filter runs. Another solution is to use consecutive estimates that are several steps apart so that they are almost uncorrelated or to generate several independent filter runs as proposed in [4].

A. Mean Squared Deviation filter consistency

Let H_0 be the null hypothesis stating that estimates $[\tilde{X}_k, \dots, \tilde{X}_M]$ given by the filter are consistent in mean squared deviation. Choose $\epsilon > N$ and define test statistics U as

$$U = \sum_{k=1}^M U_k,$$

$$U_k = \begin{cases} 1, & \text{if } (X_k - \bar{x}_k)^T \bar{P}_k^{-1} (X_k - \bar{x}_k) \leq \epsilon \\ 0, & \text{otherwise} \end{cases}$$

If H_0 is true then according to (3)

$$\Pr \{ (X_k - \bar{x}_k)^T \bar{P}_k^{-1} (X_k - \bar{x}_k) \leq \epsilon \} \geq 1 - \frac{N}{\epsilon},$$

and U is a sum of M independent Bernoulli random variables with probabilities of success $p_k \geq 1 - \frac{N}{\epsilon}$. It follows that U has a Poisson Binomial distribution since probabilities of success are not equal; this fact complicates estimation of $P(U = k)$. However, taking into account that $p_k \geq 1 - \frac{N}{\epsilon}$, if k is such that

$$1 - \frac{N}{\epsilon} \geq \frac{k}{M},$$

it is intuitively clear that

$$\Pr\{U = k\} \leq \binom{M}{k} \left(1 - \frac{N}{\epsilon}\right)^k \left(\frac{N}{\epsilon}\right)^{M-k}.$$

Thus, for given significance level α the critical region for U can be defined as $r = [0, \dots, K]$, where K is the largest value from $[0, M]$ for which

$$\sum_{k=0}^K \binom{M}{k} \left(1 - \frac{N}{\epsilon}\right)^k \left(\frac{N}{\epsilon}\right)^{M-k} \leq \alpha$$

and

$$1 - \frac{N}{\epsilon} \geq \frac{K}{M}.$$

If U falls inside the critical region r , the hypothesis that estimates are consistent in mean squared deviation can be rejected at significance level α .

1) *p equivalence*: Let H_0 be the null hypothesis stating that estimates $[\tilde{X}_1, \dots, \tilde{X}_M]$ are p equivalent. Choose $\epsilon_k > 0$ | $\Pr \{ (\tilde{X}_k - \bar{x}_k)^T \bar{P}_k^{-1} (\tilde{X}_k - \bar{x}_k) \leq \epsilon_k \} = p, k \in \{1, \dots, M\}$, and define test statistic U as

$$U = \sum_{k=1}^M U_k,$$

$$U_k = \begin{cases} 1, & \text{if } (X_k - \bar{x}_k)^T \bar{P}_k^{-1} (X_k - \bar{x}_k) \leq \epsilon_k \\ 0, & \text{otherwise} \end{cases}$$

If H_0 is true then according to (4)

$$\begin{aligned} & \Pr \{ (X_k - \bar{x}_k)^T \bar{P}_k^{-1} (X_k - \bar{x}_k) \leq \epsilon_k \} \\ & = \Pr \{ (\tilde{X}_k - \bar{x}_k)^T \bar{P}_k^{-1} (\tilde{X}_k - \bar{x}_k) \leq \epsilon_k \} = p \end{aligned} \quad (6)$$

and U is a sum of M independent Bernoulli random variables with probability of success equal to p , i.e. it has a Binomial distribution. Therefore, for given significance level α the two

sided critical region for U can be defined as $r = [0, \dots, K_1] \cup [K_2, \dots, M]$ such that

$$\Pr\{U \in r\} = \sum_{k \in r} \binom{M}{k} p^k (1-p)^{M-k} \leq \alpha.$$

The p equivalence hypothesis test with two sided critical region for U checks both consistency and information content of the estimate. If U falls into the left part of the critical region, estimates are not p consistent; if U falls into the right part of the critical region, estimates are consistent but uninformative.

B. Filter NDS consistency

1) *NDS consistency*: Let H_0 be the null hypothesis stating that estimates $[\tilde{X}_1, \dots, \tilde{X}_M]$ are NDS consistent. Define test statistics U and \tilde{U} as

$$U = \sum_{k=1}^M U_k, \quad \tilde{U} = \sum_{k=1}^M \tilde{U}_k, \quad (7)$$

where

$$\begin{aligned} U_k &= (X_k - \bar{x})^T P_k^{-1} (X_k - \bar{x}), \\ \tilde{U}_k &= (\tilde{X}_k - \bar{x})^T P_k^{-1} (\tilde{X}_k - \bar{x}). \end{aligned}$$

If H_0 is true then from (5) it can be shown (see [8]) that

$$\forall \epsilon_k > 0 \Pr\{U_k \leq \epsilon_k\} \geq \Pr\{\tilde{U}_k \leq \epsilon_k\}, \quad (8)$$

and

$$\forall \epsilon > 0 \Pr\{U \leq \epsilon\} \geq \Pr\{\tilde{U} \leq \epsilon\}.$$

Therefore, for given significance level α , a critical region for U can be defined as $r = [\epsilon_1, +\infty]$, such that

$$\Pr\{U \in r\} \leq \Pr\{\tilde{U} \in r\} \leq \alpha, \quad (9)$$

This r can be calculated based on the distributions of $\tilde{X}_1, \dots, \tilde{X}_M$. If U falls inside r , the null hypothesis H_0 can be rejected at significance level α .

IV. PRACTICAL APPLICATIONS

A. Consistency tests for suboptimal filters

The Kalman filter provides the exact optimal solution of the estimation problem when the system is linear and noises are Gaussian. When linear Gaussian assumptions hold NEES/NIS consistency tests [4] should be used for filter consistency evaluation. However, in practical applications systems are usually non-linear and noises are not necessarily Gaussian. To solve such kinds of problems suboptimal filters are used, e.g. Extended Kalman Filter or Linear Regression Kalman Filter, or non-linear systems are approximated by linear Gaussian system and Kalman Filter algorithm applied to them. Such algorithms estimate the state of the system by Gaussian random variables even though the actual distribution of the state is not necessarily Gaussian. In this case the hypothesis tests for p equivalence or NDS consistency can

be used. When the estimate is provided in the form of a Gaussian random variable, the NDS consistency test is almost the same as the NEES consistency test. However, in the NDS consistency test the true posterior distribution is not assumed to be Gaussian, whereas the NEES test assumes that both estimated and actual distributions are Gaussian. Moreover, the NDS consistency test as well as the mean squared deviation and p equivalence tests make only mild assumptions about the estimated posterior distribution, and require only calculation of test statistics that can be done at least numerically. This makes the proposed tests applicable to different types of distribution and to different filtering algorithms such as particle filters and Grid based filters.

In this section we apply consistency tests to a WiFi based Indoor Localization system. We use constant velocity model as a motion model of a user, and vectors of position derived from wireless signals are used as a measurements of the user position. Process and measurement noises are approximated as Gaussian and a Kalman filter is used to compute the posterior distribution of user position.

B. System model

Let x_k be the state of the system at time t_k , y_k a measurement of the state at time t_k , and Δt_k the time difference between consecutive time moments. In constant velocity (CV) model, the state evolves according to the state transition equation

$$x_k = Fx_{k-1} + w_{k-1}, \quad (10)$$

where

$$F = \begin{bmatrix} 1 & 0 & \Delta t_k & 0 \\ 0 & 1 & 0 & \Delta t_k \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and w_{k-1} is a zero mean Gaussian white noise with covariance matrix

$$Q = \begin{bmatrix} \frac{\Delta t_k^3}{3} Q_c & \frac{\Delta t_k^2}{2} Q_c \\ \frac{\Delta t_k^2}{2} Q_c & \Delta t_k Q_c \end{bmatrix}$$

where

$$Q_c = \begin{bmatrix} \sigma_{\text{Lat}}^2 & 0 \\ 0 & \sigma_{\text{Lon}}^2 \end{bmatrix}$$

is a so-called diffusion of the Brownian motion process [4].

In our system single WiFi-based position fixes are used as measurements of the state; the measurement equation is

$$y_k = Hx_k + v_k \quad (11)$$

where $H = [I_{2 \times 2} \ 0_{2 \times 2}]$ and v_k is a zero mean white Gaussian measurement noise with covariance matrix R . Process and measurement noises are independent.

TABLE I
RATIO OF NDS CONSISTENT ESTIMATES

$r \backslash q$	0.1	0.3	0.5	0.7	0.9	1.1
1	0	0.01	0.01	0.03	0.03	0.05
2	0.07	0.22	0.31	0.38	0.44	0.44
3	0.12	0.62	0.77	0.83	0.87	0.87
4	0.27	0.83	0.90	0.92	0.92	0.94
5	0.42	0.87	0.98	0.98	0.98	0.98
6	0.42	0.92	0.98	0.98	0.98	0.98
7	0.50	0.94	0.98	0.98	0.98	1
8	0.59	0.94	0.98	0.98	1	1
9	0.64	0.96	0.98	1	1	1
10	0.64	0.96	0.98	1	1	1

C. Linear system with non-Gaussian noises (real data)

Consider $T = 804$ real consecutive positions of a user moving indoors, and corresponding position estimates made by a WiFi positioning system. It is reasonable to assume the CV motion model and use a Kalman filter in this case even though the user does not always move with a constant speed and can make abrupt turns or stops, i.e. the actual process noise is not Gaussian. Also measurement noise is not Gaussian due to various factors affecting WiFi signal propagation.

In order to ensure that the Kalman filter is consistent, matrices $Q_c = q^2 \cdot I_{2 \times 2}$ and $R = r^2 \cdot I_{2 \times 2}$ must be appropriately scaled so that process and measurement noises are sufficiently large to compensate for motion modeling and measurement errors. Here we use simple assumptions about noise covariance matrices, i.e. latitude and longitude errors are uncorrelated and have equal variances.

In order to investigate the impact of measurement and process noise on consistency of the estimation, user positions are estimated using different values of parameters q and r , and consistency of the estimates is evaluated with the NDS consistency test. Test statistic U is calculated based on the series of 3 estimates that are 5 time instances apart, in this case estimates are almost uncorrelated and fulfill our assumptions. The reason to take According to III-B.1, the critical region for U can be set based on the χ_{12}^2 distribution. We use significance level $\alpha = 0.1$ and set the critical region to $r = [18.5, +\infty)$. If $U \in r$, the filter is declared to be inconsistent.

In Table I the ratio of NDS consistent estimates for different process and measurement noises is presented. Filters for which the ratio of consistent estimates is larger than $1 - \alpha$ can be considered as NDS consistent in general. As seen from the figure, filter consistency is mainly influenced by the magnitude of the measurement noise, whereas process noise does not have noticeable impact. This indicates that consistency of the estimates largely depends on the correct modelling of the measurement noise. In Fig. 1 the empirical cumulative distribution function of normalized squared estimation error F_{NDS} and $F_{\chi_4^2}$, which is the cumulative distribution function of normalized squared error according to the Kalman filter, are plotted for different measurement noise levels ($q = 0.9, r \in \{2, \dots, 4\}$). If $r \leq 2$, $F_{\text{NDS}}(u)$ is smaller

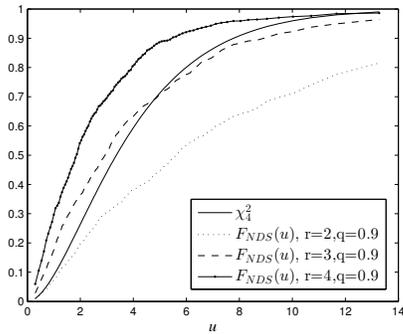


Fig. 1. Cumulative distribution function of NDS statistic for different measurement noise levels.

than $F_{\chi^2_4}(u)$, i.e. the filter does not behave NDS consistently. If $r \geq 4$, $F_{\text{NDS}}(u)$ is greater than $F_{\chi^2_4}(u)$, i.e. the filter is NDS consistent. For $r = 3$, $F_{\text{NDS}}(u)$ is approximately equal to $F_{\chi^2_4}(u)$ and filter can also be accepted as consistent. This complies with the empirical measurement noise covariance approximately equal to $9_{2 \times 2}$. It should also be mentioned that for large r the filter is less informative for confidence regions of probability $p < 0.9$, for example when $r = 4$ the 68% concentration ellipses declared by the filter actually contain about 90% of the estimates, however this still conforms with the definition of NDS consistency.

D. Consistency of predicted measurement

In the course of state estimation most of the filters implicitly predict a future measurement based on predicted state and measurement model of the system. Measurement estimates provided by the filter must be consistent with the actual received measurements.

If linear Gaussian assumptions hold and a Kalman filter is used, NIS test [4] should be used for checking consistency of the predicted measurement. If the system is non-linear and a suboptimal filter is used, the consistency of measurement prediction can be checked with NDS consistency or p equivalence tests.

Consistency of the measurement estimate is highly correlated with consistency of the state estimate and might be a good indicator of abnormal filter behavior.

An important advantage of predictive measurement consistency testing is that it does not require any information about the true state of the system and only requires actual measurements that are always available. Hence, it can be used on-line in order to adjust filter parameters such as uncertainties of process and measurement noises. However, if the assumed measurement noise is not consistent with the actual measurement noise, consistency of predictive measurements does not necessarily indicate consistency of state estimates.

V. CONCLUSION

The proposed NDS consistency test is able to detect inconsistent behavior when the noise intensity parameters are not sufficiently large. It is effective in off-line consistency evaluation and adjustment of system noises. Moreover it makes only minor assumptions about the system compared to the state of the art methods.

ACKNOWLEDGMENT

We thank our colleagues from the Positioning team of HERE (a Nokia Business) and Tampere University of Technology for their support and valuable discussions related to the topic covered in this paper.

REFERENCES

- [1] T. Lefebvre, H. Bruyninckx, and J. de Schutter, "Kalman filters for non-linear systems: a comparison of performance," *International Journal of Control*, vol. 77, pp. 639–653, 2004.
- [2] F. Van der Heijden, "Consistency checks for particle filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 140–145, Jan 2006.
- [3] M. Rosenblatt, "Remarks on a multivariate transformation," *The Annals of Mathematical Statistics*, vol. 23, pp. 470–472, 1952.
- [4] Y. Bar-Shalom and X.-R. Li, *Estimation and Tracking: Principles, Techniques, and Software*. Artech House, 1998.
- [5] M. Scalzo, G. Horvath, E. Jones, A. Bubalo, M. Alford, R. Niu, and P. K. Varshney, "Adaptive filtering for single target tracking," *Proceedings of the SPIE: Defense Security Symposium*, vol. 4336, 2009.
- [6] H. Nurminen, A. Ristimäki, S. Ali-Löytty, and R. Piché, "Particle filter and smoother for indoor localization," in *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN2013)*, Montbéliard-Belfort, France, 28–31 October 2013, pp. 137–146.
- [7] S. Ali-Löytty, N. Sirola, and R. Piché, "Consistency of three Kalman filter extensions in hybrid navigation," in *Proceedings of the European Navigation Conference GNSS 2005, July 19–22, 2005, Munich, 2005*. [Online]. Available: http://math.tut.fi/posgroup/ali-loytty_et_al_enc2005a.pdf
- [8] P. Ivanov, *Consistency of Estimation*. MSc Thesis, Tampere University of Technology, 2014. [Online]. Available: <http://URN.fi/URN:NBN:fi:tyt-201405131158>