TAMPERE UNIVERSITY OF TECHNOLOGY

PAVEL IVANOV

CONSISTENCY OF ESTIMATION

Master of Science Thesis

Subject approved by the Department Council on 05.12.2012
Examiners: Prof. Robert Piché (TUT)
D.Tech. Simo Ali-Löytty (TUT)

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Information Technology
IVANOV, PAVEL: Consistency of Estimation
Master of Science Thesis, 52 pages, 0 Appendix pages
May 2014
Major: Mathematics
Examiners: Prof. Robert Piché and D.Tech. Simo Ali-Löytty
Keywords: estimation, consistency, consistency test

Besides accuracy, consistency and information content are important properties of the estimation. It is crucial that estimator provide realistic information about possible estimation error, especially in fusion algorithms, where several estimates from different sources are merged into one.

In this thesis consistency is defined in several ways and methods for its evaluation are introduced. Mean Squared Deviation consistency is based on Chebyshev's inequality which defines lower bound of probability mass concentrated around mean of random variable. P consistency implies that concentration ellipse around mean of the estimate with probability mass p must contain actual value of estimated parameter with probability p. Normalized deviation squared consistency implies that concentration ellipse of any probability around mean of the estimate must contain actual value of estimated parameter with probability p. Information content is defined in terms of most informative estimate, estimate which has the highest precision and yet consistent.

In this work statistical hypothesis testing framework is used for consistency and information content evaluation. Hypothesis tests for consistency evaluation are derived for static parameter estimation and state estimation (filtering). Proposed consistency tests are applied to Indoor WiFi localization system in order to investigate sources of inconsistencies and adjust parameters of the system in off-line mode. It is shown that underestimated measurement noise is the main reason of estimates' inconsistency, however, considerably underestimated process noise or motion mis-modeling might also result in inconsistent and abnormal filter's behavior.

Preface

This Master of Science Thesis was written in the Department of Mathematics at Tampere University of Technology. Motivation for the thesis work arose from the research conducted by Personal Positioning Algorithms group of Tampere University of Technology and positioning team of HERE, a Nokia business.
I would like to thank my supervisor Prof. Robert Piché for introducing me to Bayesian inference and its applications in positioning, for support during my studies in TUT and for valuable guidance during MSc thesis writing. I would like to thank D.Tech. Simo Ali-Löytty for valuable comments and feedback related to the thesis work. I also would like to thank D.Tech. Jari Syrjärinne and my colleagues from HERE positioning team and Personal Positioning Algorithms group of TUT for support during my education and work.
And of course I would like to thank my family and friends from Rostov-on-Don who support me during my life.

Tampere, 17th April 2014

Pavel Ivanov

Contents

Symbols and Abbreviations

$\mathbb{R}^N$	set of $N-$dimensional real numbers
$\mathbb{N}$	set of natural numbers
$A \leq B$	matrix difference $B - A$ is a positive semi-definite matrix
$\mathbb{E}(X)$	expectation of random variable X
A^T	matrix transpose
A^{-1}	matrix inverse
$\Pr\{A\}$	probability of A
$\Pr\{A\|B\}$	probability of A given B
$\in$	belong to
$\sim$	distributed according to
$N(\mu, P)$	normal distribution with mean μ and covariance matrix P
$B(n, p)$	binomial distribution with n trials and success probability p
$p_X(x)$	probability density function of X
$p_X(x\|Y = y)$	conditional probability density function of X given $Y = y$
$p_X(x\|Y)$	conditional probability density function of X given Y
H_0	null hypothesis
H_1	alternative hypothesis
$x_{1:k}$	states from time 1 up to time k
$y_{1:k}$	measurements from time 1 up to time k
$\propto$	proportional to
$\approx$	approximately equal to
$\bar{x}_{k\|k}$	posterior mean after update step
$\bar{P}_{k\|k}$	posterior covariance matrix after update step
$\bar{x}_{k\|k-1}$	prior mean after prediction step
$\bar{P}_{k\|k-1}$	prior covariance matrix after prediction step
η_k	innovation at step k
S_k	innovation covariance matrix at step k

K_k	Kalman gain
I	identity matrix
O	zero matrix
$\underset{x}{\operatorname{argmin}} f(x)$	value of x in which $f(x)$ achieves its minimum
$\underset{x}{\operatorname{argmax}} f(x)$	value of x in which $f(x)$ achieves its maximum
$\ll$	much less than
$x^i_{1:k}$	set of particles representing states from time 1 up to time k
$w^i_{1:k}$	weights of particles representing states from time 1 up to time k
δ	Dirac delta function
N_{eff}	number of effective particles
χ^2_M	chi-square distribution with M degrees of freedom
n_x	dimension of x
t_k	k's time moment
$\triangle t_k$	difference between t_k and t_{k-1}
$F_X(x)$	cumulative distribution function of X
$F^{-1}_X(x)$	inverse cumulative distribution function of X
$\hat{F}_X(x)$	empirical cumulative distribution function of X

CDF	Cumulative Distribution Function
CV	Constant Velocity
CDKF	Central Difference Kalman Filter
EKF	Extended Kalman Filter
GPS	Global Positioning System
KF	Kalman Filter
LG	Linear and Gaussian
LRKF	Linear Regression Kalman Filter
MSD	Mean Squared Deviation
NDS	Normalized Deviation Squared
NEES	Normalized Estimation Error Squared
NIS	Normalized Innovation Squared
PF	Particle Filter
RV	Random Variable
SIS	Systematic Importance Sampling
UKF	Unscented Kalman Filter
WiFi	Wireless communication technology

Chapter 1

Introduction

Estimation is a widely used and actively studied scientific and engineering area. In our life we often have to estimate different parameters and quantities that are unknown. The diversity of things to be estimated is huge: geographical location of an object; different environmental parameters such as temperature, air pressure, humidity; future stock prices and many others. In the sequel all such quantities are called parameters of interest or system states.

Estimate of an unknown parameter can be represented in different ways: by a point estimate and confidence region around it, by a point estimate and respective mean squared error matrix, or by a probability distribution from which point estimate as well as confidence regions can be calculated. Uncertainty of the estimate must be consistent with the actual estimation error. Generally it means that actual value of estimated parameter must fall within confidence region of probability p (declared by the estimate) with probability at least as large as p. An estimate that meets this requirement is considered consistent.

In order to make estimate consistent its uncertainty can be increased artificially, but with increasing uncertainty confidence regions become larger and estimate becomes less informative. Thus, there is a trade-off between consistency and information content.

Several definitions of consistency and ways to evaluate it are presented in the literature. Lefebvre et al. [10] define estimate as consistent if its covariance matrix is larger then actual Mean Estimation Error Matrix of the estimate. Ali-Löytty et al. [1] present a Generalized inconsistency test based on consistency of Mean Estimation Error Matrix. Van der Heijden [18] 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 that any continuous multivariate RV can be transformed to multivariate standard uniform RV [15]. Bar-Shalom and Li [2] present NEES and NIS consistency checks for Kalman filter, and define filter estimate as consistent if its normalized estimation error squared (NEES) and corresponding normalized innovation squared (NIS) follow chi-square distribution with number of degrees of freedom equal to dimensions of state and measurement vectors respectively. Gibbs [5] introduces modification of NIS test by using posterior measurement residual as a test statistic. Scalzo et al. [16] extend application of NEES/NIS test to non-linear systems and suboptimal filters (Particle filter) by approximating posterior distributions as Gaussian distributions. Nurminen et al. [12] define an estimate as 95% consistent if actual value of the estimated parameter falls in the 95% concentration ellipse.

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

In this work several consistency concepts presented in the literature are defined and formalized, and one new consistency concept is introduced by the author. Hypothesis tests for consistency evaluation for static and dynamic estimation are derived and applied to simulated and real data sets.

Thesis work is organized as follows. In Chapter 2 Mean Squared Deviation (MSD) consistency, p-consistency/equivalence and Normalized Deviation Squared (NDS) consistency/equivalence are defined and exemplified for static estimation. In Chapter 3 hypothesis theory is briefly introduced and hypothesis tests are derived for consistency evaluation of static estimation. Chapter 4 introduces basics of dynamical systems and Bayesian state estimation (filtering), as well as standard state estimators (filters) such as Kalman filter, Extended Kalman filter and Particle filter. In Chapter 5 filter consistency is defined, and hypothesis tests are derived for evaluation of filter consistency. Practical applications and experiments are presented in Chapter 6. Application area is a geographical localization. Simulated and real positioning data and systems are used for position estimation, and consistency of estimates is evaluated with consistency tests developed throughout the thesis work. Chapter 7 concludes the thesis work.

Chapter 2

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 parameter 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 vice versa, informative estimate might be inconsistent. However, consistency is regarded as more important criterion since it determines validity of the estimator, whereas information content indicates how useful estimation is.

2.1 Consistency

There are several ways to define consistency depending on the format of the estimate's uncertainty, e.g. probability distribution, covariance matrix or confidence region. In this Chapter different consistency definitions that have been proposed in the literature and consistency definition proposed by author are presented.

2.1.1 Consistency in Mean Squared Deviation

Consider N-variate random variable X with existing mean and covariance. Estimate of X defined by $\bar{x} \in \mathbb{R}^N$ and covariance matrix $\bar{P}$, is called consistent [10] if

$$M \leq \bar{P}, \tag{2.1}$$

where $M = \mathbb{E}\left[(X-\bar{x})(X-\bar{x})^T\right]$ is the mean squared deviation (error) matrix of RV X with respect to $\bar{x}$, and symbol "$\leq$" means that matrix difference $\bar{P}-M$ is a positive semi-definite matrix. Both M and $\bar{P}$ are assumed to be non-singular.
In this thesis consistency defined by (2.1) is called mean squared deviation consistency and it has the following meaning. If (2.1) holds then $z^T M^{-1} z \geq z^T \bar{P}^{-1} z$ [see 11, p.586] and if $(z-\bar{x})^T M^{-1}(z-\bar{x}) \leq \epsilon$ then $(z-\bar{x})^T \bar{P}^{-1}(z-\bar{x}) \leq \epsilon$. In other words, requirement (2.1) means that ellipsoid centered in $\bar{x}$ and defined by inequality $(z-\bar{x})^T \bar{P}^{-1}(z-\bar{x}) \leq \epsilon$ contains ellipsoid centered in $\bar{x}$ and defined by inequality $(z-\bar{x})^T M^{-1}(z-\bar{x}) \leq \epsilon$. Therefore probability mass contained in ellipsoid defined by $\bar{P}$ is larger or equal than probability mass contained in the ellipsoid defined by M, i.e.

$$\forall \epsilon > N \,:\, \Pr\left\{(X-\bar{x})^T \bar{P}^{-1}(X-\bar{x}) \leq \epsilon\right\} \geq \Pr\left\{(X-\bar{x})^T M^{-1}(X-\bar{x}) \leq \epsilon\right\} \tag{2.2}$$

Moreover, according to generalization of Chebyshev's inequality [1, 3]

$$\forall \epsilon > N \,:\, \Pr\left\{(X-\bar{x})^T M^{-1}(X-\bar{x}) \leq \epsilon\right\} \geq 1 - \frac{N}{\epsilon}. \tag{2.3}$$

Therefore, consistency in mean squared deviation implies that

$$\forall \epsilon > N \,:\, \Pr\left\{(X-\bar{x})^T \bar{P}^{-1}(X-\bar{x}) \leq \epsilon\right\} \geq 1 - \frac{N}{\epsilon}, \tag{2.4}$$

i.e. probability of deviation of actual parameter from its estimate agrees with theoretical lower bound given by Chebyshev's inequality.
For example consider random variable X with mean m and covariance matrix Σ, and

its estimation defined by point estimate $\bar{m}$ and estimated mean squared deviation matrix $\bar{\Sigma}$. Assume that $\bar{m} = m + e$, then mean squared deviation matrix of X with respect to $\bar{m}$ can be calculated as

$$\begin{aligned}
\mathbb{E}\left[(X-\bar{m})(X-\bar{m})^T\right] &= \mathbb{E}\left[(X-m-e)(X-m-e)^T\right] \\
&= \mathbb{E}\left[(X-m-e)(X-m)^T\right] - \mathbb{E}\left[(X-m-e)e^T\right] \\
&= \mathbb{E}\left[(X-m)(X-m)^T\right] - \mathbb{E}\left[e(X-m)^T\right] - \\
&\quad \mathbb{E}\left[(X-m)e^T\right] + \mathbb{E}\left[ee^T\right] \\
&= \Sigma + ee^T.
\end{aligned}$$

Hence, if $\bar{\Sigma} \geq \Sigma + ee^T$ then estimate defined by $\bar{x}$ and $\bar{P}$ is consistent in mean squared deviation sense.

2.1.2 P-Consistency

N-variate random variable $\tilde{X}$ with mean $\bar{x} \in \mathbb{R}^N$ and covariance matrix $\bar{P}$ is a p consistent estimate of N-variate random variable X if for $p \in [0,1]$

$$\exists \epsilon > 0 \mid \Pr\left\{(X-\bar{x})^T\bar{P}^{-1}(X-\bar{x}) \leq \epsilon\right\} \geq \Pr\left\{(\tilde{X}-\bar{x})^T\bar{P}^{-1}(\tilde{X}-\bar{x}) \leq \epsilon\right\} = p \quad (2.5)$$

Requirement (2.5) means that ellipsoid centered in $\bar{x}$ and containing probability p according to the distribution of $\tilde{X}$ contains at least probability p according to the distribution of X.

For example consider scalar random variable X that is distributed according to normal distribution with mean m and variance σ^2 and its approximation $\tilde{X}$ distributed normally with mean $\bar{m} = m + e$ and variance $\bar{\sigma}^2 = (\sigma + |e|)^2$. Let's show that $\tilde{X}$ is a p consistent estimate of X for $p \geq 0.68$.

According to the definition, $\tilde{X}$ is p consistent if

$$\exists \epsilon > 0 \mid \Pr\{(X-\bar{m})^T\frac{1}{\bar{\sigma}^2}(X-\bar{m}) \leq \epsilon\} \geq \Pr\{(\tilde{X}-\bar{m})^T\frac{1}{\bar{\sigma}^2}(\tilde{X}-\bar{m}) \leq \epsilon\} = p \quad (2.6)$$

Since X is a scalar RV, inequality (2.6) can be rewritten as

$$\Pr\left\{-\sqrt{\epsilon}\cdot\bar{\sigma} \leq X-\bar{m} \leq \sqrt{\epsilon}\cdot\bar{\sigma}\right\} \geq \Pr\left\{-\sqrt{\epsilon}\cdot\bar{\sigma} \leq \tilde{X}-\bar{m} \leq \sqrt{\epsilon}\cdot\bar{\sigma}\right\} \quad (2.7)$$

Assume that $\Pr\left\{-\sqrt{\epsilon}\cdot\bar{\sigma} \leq \tilde{X}-\bar{m} \leq \sqrt{\epsilon}\cdot\bar{\sigma}\right\} \geq 0.68$, then $\epsilon > 1$, and $-\sqrt{\epsilon}\cdot\sigma \leq z - m \leq \sqrt{\epsilon}\cdot\sigma \Rightarrow -\sqrt{\epsilon}\cdot(\sigma+|e|) \leq z-\bar{m} \leq \sqrt{\epsilon}\cdot(\sigma+|e|)$. Therefore,

$\Pr\{-\sqrt{\epsilon}\cdot\bar{\sigma} \leq X - \bar{m} \leq \sqrt{\epsilon}\cdot\bar{\sigma}\} \geq \Pr\{-\sqrt{\epsilon}\cdot\sigma \leq X - m \leq \sqrt{\epsilon}\cdot\sigma\}$. And since both X and $\tilde{X}$ are distributed normally

$$\Pr\left\{-\sqrt{\epsilon}\cdot\sigma \leq X - m \leq \sqrt{\epsilon}\cdot\sigma\right\} = \Pr\left\{-\sqrt{\epsilon}\cdot\bar{\sigma} \leq \tilde{X} - \bar{m} \leq \sqrt{\epsilon}\cdot\bar{\sigma}\right\},$$

and

$$\Pr\left\{-\sqrt{\epsilon}\cdot\bar{\sigma} \leq X - \bar{m} \leq \sqrt{\epsilon}\cdot\bar{\sigma}\right\} \geq \Pr\left\{-\sqrt{\epsilon}\cdot\bar{\sigma} \leq \tilde{X} - \bar{m} \leq \sqrt{\epsilon}\cdot\bar{\sigma}\right\},$$

Therefore, if

$$\Pr\left\{-\sqrt{\epsilon}\cdot\bar{\sigma} \leq \tilde{X} - \bar{m} \leq \sqrt{\epsilon}\cdot\bar{\sigma}\right\} = p \geq 0.68,$$

then

$$\Pr\left\{-\sqrt{\epsilon}\cdot\bar{\sigma} \leq X - \bar{m} \leq \sqrt{\epsilon}\cdot\bar{\sigma}\right\} \geq \Pr\left\{-\sqrt{\epsilon}\cdot\bar{\sigma} \leq \tilde{X} - \bar{m} \leq \sqrt{\epsilon}\cdot\bar{\sigma}\right\} = p \geq 0.68,$$

i.e. $\tilde{X}$ is p consistent estimate of X for $p \geq 0.68$.
In the Fig. 2.1 probability distribution function of normal RV X with mean 5 and standard deviation 3 and distribution of its normal estimate $\tilde{X}$ with mean 6 and standard deviation 4 are plotted. $\tilde{X}$ is a p-consistent estimate of X for $p \geq 0.68$.

2.1.3 P-Equivalence

N-variate random variable $\tilde{X}$ with mean $\bar{x} \in \mathbb{R}^N$ and covariance matrix $\bar{P}$ is a p equivalent estimate of N-variate random variable X if for $p \in [0,1]$

$$\exists\epsilon > 0 \mid \Pr\left\{(X-\bar{x})^T\bar{P}^{-1}(X-\bar{x}) \leq \epsilon\right\} = \Pr\left\{(\tilde{X}-\bar{x})^T\bar{P}^{-1}(\tilde{X}-\bar{x}) \leq \epsilon\right\} = p \quad (2.8)$$

Requirement (2.8) means that ellipsoid centered in $\bar{x}$ and containing probability p according to the distribution of $\tilde{X}$ contains probability p according to the distribution of X. As seen from (2.8) p-equivalence implies p-consistency, but not vice versa.
p consistency/equivalence originates from consistency definition used in [12], there estimate is said to be 95% consistent if actual value of estimated parameter falls in 95% concentration ellipsoid.
For example consider standard normal random variable X and random variable $\tilde{X} = 2.9412\cdot U - 1.4706$, where U is a standard uniform RV. $\tilde{X}$ is 0.68 equivalent estimate of X since

$$\Pr\{X \in [-1,1]\} = \Pr\left\{\tilde{X} \in [-1,1]\right\} = 0.68$$

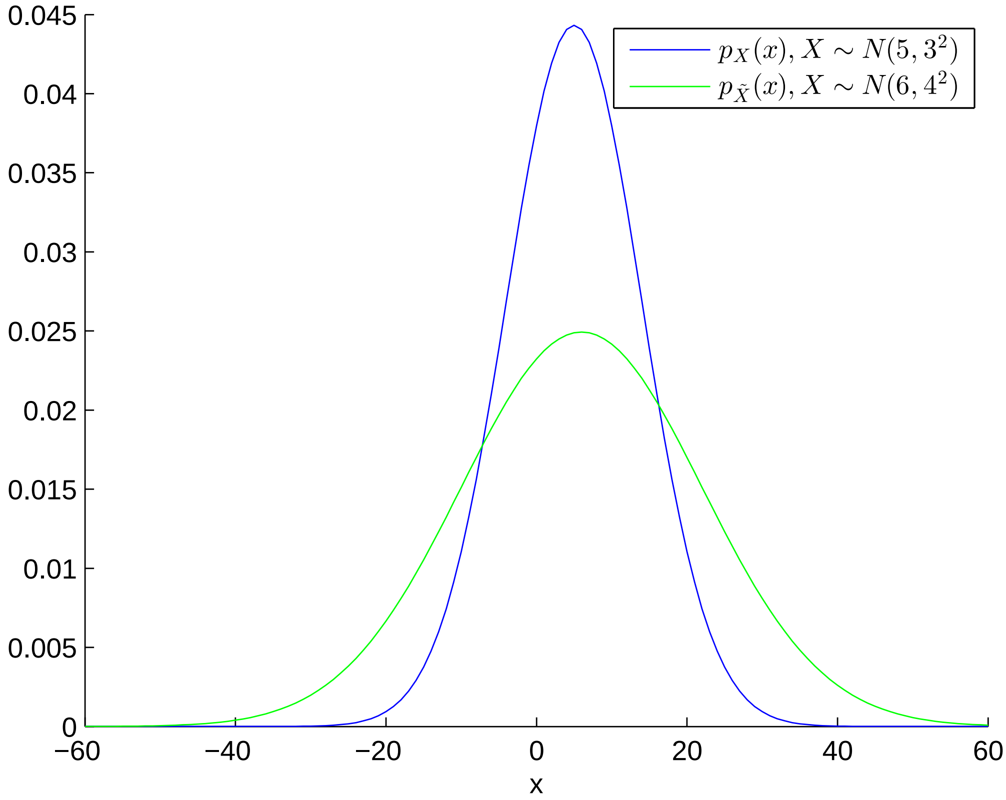


Figure 2.1: Probability density functions of RV $X \sim N(5, 3^2)$ and its p-consistent estimate $\tilde{X} \sim N(6, 4^2)$ for $p \in [0.68, 1]$.

2.1.4 Normalized Deviation Squared consistency

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

$$\forall \epsilon > 0 \,:\, \Pr\left\{(X - \bar{x})^T \bar{P}^{-1}(X - \bar{x}) \leq \epsilon\right\} \geq \Pr\left\{(\tilde{X} - \bar{x})^T \bar{P}^{-1}(\tilde{X} - \bar{x}) \leq \epsilon\right\} \tag{2.9}$$

Inequality (2.9) means that for any $\epsilon > 0$ ellipsoid centered in $\bar{x}$ and defined by inequality $(z - \bar{x})^T \bar{P}^{-1}(z - \bar{x}) \leq \epsilon$ contain probability mass of $\tilde{X}$ that is smaller or equal than probability mass of X contained in the same ellipsoid. I.e. normalized deviation squared consistency implies p consistency for any $p \in [0, 1]$.
For example consider scalar normal random variable X with mean m and variance σ^2

and its estimate $\tilde{X}$ distributed normally with mean $\tilde{m} = m + e$ and standard deviation $\tilde{\sigma} \geq \max(x_0 - \tilde{m}, \sigma + |e|)$, where x_0 is such that

$$\int_{\tilde{m}}^{x_0} p_X(x)dx = \int_{m}^{m+\sigma} p_X(x)dx. \tag{2.10}$$

We assume that $\tilde{m} \geq m$ and that difference $\tilde{m} - m$ is sufficiently small so that there exist $x_0 < +\infty$ satisfying (2.10) (this is possible if $\tilde{m} \in [m, m + 0.4\sigma)$). Let's show that $\tilde{X}$ is an NDS consistent estimate of X.
According to the definition, $\tilde{X}$ is an NDS consistent estimate of X if

$$\forall \epsilon > 0 \,:\, \Pr\left\{(X - \tilde{m})^T \frac{1}{\tilde{\sigma}^2}(X - \tilde{m}) \leq \epsilon\right\} \geq \Pr\left\{(\tilde{X} - \tilde{m})^T \frac{1}{\tilde{\sigma}^2}(\tilde{X} - \tilde{m}) \leq \epsilon\right\}. \tag{2.11}$$

In Section 2.1.2 it was shown that if $\tilde{\sigma} = \sigma + |e|$ then inequality above holds for $\epsilon \geq 1$. Evidently this is also true for $\tilde{\sigma} \geq \sigma + |e|$, therefore it is enough to prove the case when $\epsilon < 1$.
Assume that $\epsilon < 1$ and consider integrals

$$\int_{\tilde{m}}^{\tilde{m}+\tilde{\sigma}} p_{\tilde{X}}(x)dx \text{ and } \int_{\tilde{m}}^{\tilde{m}+\tilde{\sigma}} p_X(x)dx.$$

$\tilde{\sigma}$ is chosen so that $\tilde{\sigma} \geq x_0 - \tilde{m}$, therefore

$$\int_{\tilde{m}}^{\tilde{m}+\tilde{\sigma}} p_X(x)dx \geq \int_{\tilde{m}}^{x_0} p_X(x)dx = \int_{m}^{m+\sigma} p_X(x)dx = \int_{\tilde{m}}^{\tilde{m}+\tilde{\sigma}} p_{\tilde{X}}(x)dx \Rightarrow$$

$$\int_{\tilde{m}}^{\tilde{m}+\tilde{\sigma}} p_{\tilde{X}}(x)dx \leq \int_{\tilde{m}}^{\tilde{m}+\tilde{\sigma}} p_X(x)dx \tag{2.12}$$

Let's now use (2.12) in order to show that for $y \in [\tilde{m}, \tilde{m} + \tilde{\sigma}]$

$$\int_{\tilde{m}}^{y} p_{\tilde{X}}(x)dx \leq \int_{\tilde{m}}^{y} p_X(x)dx \tag{2.13}$$

Due to the nature of normal probability density function two cases are possible.

1. If $p_{\tilde{X}}(x) \leq p_X(x)\ \forall x \in [\tilde{m}, y]$, then

$$\int_{\tilde{m}}^{y} p_{\tilde{X}}(x)dx \leq \int_{\tilde{m}}^{y} p_X(x)dx$$

2. If $p_{\tilde{X}}(x) \geq p_X(x)\ \forall x \in [y_0, y]$ ($\tilde{m} \leq y_0 \leq y$) then $p_{\tilde{X}}(x) \geq p_X(x)$ on $[y_0, \tilde{m} + \tilde{\sigma}]$ and

$$\int_{y}^{\tilde{m}+\tilde{\sigma}} p_{\tilde{X}}(x)dx \geq \int_{y}^{\tilde{m}+\tilde{\sigma}} p_X(x)dx.$$

Hence,

$$\begin{aligned}\int_{\tilde{m}}^{y} p_{\tilde{X}}(x)dx &= \int_{\tilde{m}}^{\tilde{m}+\tilde{\sigma}} p_{\tilde{X}}(x)dx - \int_{y}^{\tilde{m}+\tilde{\sigma}} p_{\tilde{X}}(x)dx \\ &\leq \int_{\tilde{m}}^{\tilde{m}+\tilde{\sigma}} p_X(x)dx - \int_{y}^{\tilde{m}+\tilde{\sigma}} p_X(x)dx = \int_{\tilde{m}}^{y} p_X(x)dx,\end{aligned}$$

and inequality (2.13) holds.
Taking into account that $\tilde{m} > m$ and properties of normal distribution, following is true:

$$\int_{\tilde{m}-(y-\tilde{m})}^{\tilde{m}} p_X(x)dx \geq \int_{\tilde{m}}^{y} p_X(x)dx \tag{2.14}$$

$$\int_{\tilde{m}-(y-\tilde{m})}^{\tilde{m}} p_{\tilde{X}}(x)dx = \int_{\tilde{m}}^{y} p_{\tilde{X}}(x)dx. \tag{2.15}$$

If $0 < \epsilon < 1$ then $\tilde{m} < \tilde{m} + \sqrt{\epsilon} \cdot \tilde{\sigma} \leq \tilde{m} + \tilde{\sigma}$, and (2.13), (2.14), (2.15) hold for $y = \tilde{m} + \sqrt{\epsilon} \cdot \tilde{\sigma}$. Therefore by substituting y in (2.14), (2.15) and (2.13) with $\tilde{m} + \sqrt{\epsilon} \cdot \tilde{\sigma}$ we get the following

$$\begin{aligned}\Pr\left\{(X-\tilde{m})^T \frac{1}{\tilde{\sigma}^2}(X-\tilde{m}) \leq \epsilon\right\} &= \Pr\left\{-\sqrt{\epsilon}\cdot\tilde{\sigma} + \tilde{m} \leq X \leq \sqrt{\epsilon}\cdot\tilde{\sigma} + \tilde{m}\right\} \\ &= \int_{\tilde{m}-\sqrt{\epsilon}\cdot\tilde{\sigma}}^{\tilde{m}} p_X(x)dx + \int_{\tilde{m}}^{\tilde{m}+\sqrt{\epsilon}\cdot\tilde{\sigma}} p_X(x)dx \\ &\overset{(2.14)}{\geq} \int_{\tilde{m}}^{\tilde{m}+\sqrt{\epsilon}\cdot\tilde{\sigma}} p_X(x)dx + \int_{\tilde{m}}^{\tilde{m}+\sqrt{\epsilon}\cdot\tilde{\sigma}} p_X(x)dx \\ &\overset{(2.13)}{\geq} \int_{\tilde{m}}^{\tilde{m}+\sqrt{\epsilon}\cdot\tilde{\sigma}} p_{\tilde{X}}(x)dx + \int_{\tilde{m}}^{\tilde{m}+\sqrt{\epsilon}\cdot\tilde{\sigma}} p_{\tilde{X}}(x)dx \\ &\overset{(2.15)}{=} \int_{\tilde{m}-\sqrt{\epsilon}\cdot\tilde{\sigma}}^{\tilde{m}} p_{\tilde{X}}(x)dx + \int_{\tilde{m}}^{\tilde{m}+\sqrt{\epsilon}\cdot\tilde{\sigma}} p_{\tilde{X}}(x)dx \\ &= \Pr\left\{-\sqrt{\epsilon}\cdot\tilde{\sigma} + \tilde{m} \leq \tilde{X} \leq \sqrt{\epsilon}\cdot\tilde{\sigma} + \tilde{m}\right\} \\ &= \Pr\left\{(\tilde{X}-\tilde{m})^T \frac{1}{\sigma^2}(\tilde{X}-\tilde{m}) \leq \epsilon\right\}.\end{aligned}$$

Thereby for chosen $\tilde{m}$ and $\tilde{\sigma}$

$$\forall\epsilon > 0 \, : \, \Pr\left\{(X-\tilde{m})^T \frac{1}{\sigma^2}(X-\tilde{m}) \leq \epsilon\right\} \geq \Pr\left\{(\tilde{X}-\tilde{m})^T \frac{1}{\sigma^2}(\tilde{X}-\tilde{m} \leq \epsilon\right\},$$

that is $\tilde{X}$ is an NDS consistent estimate of X.
In the Fig. 2.2 the probability distribution function of normal RV X with mean 5 and standard deviation 3 and distribution of its normal estimate $\tilde{X}$ with mean 6 and standard deviation 5 are plotted. It can be checked that $\int_6^{6+5} p_X(x)dx > \int_5^{5+3} p_X(x)dx$ and $\tilde{\sigma} > \sigma + |m - \tilde{m}|$, thus $\tilde{X}$ is an NDS consistent estimate of X.

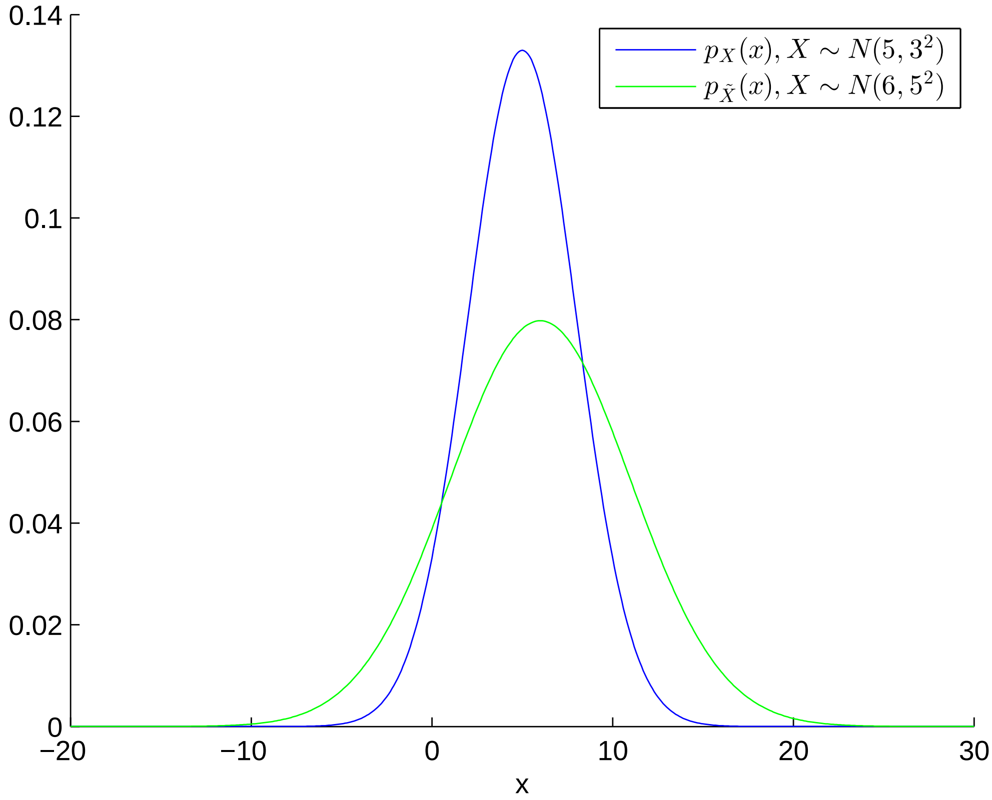


Figure 2.2: Probability density functions of RV $X \sim N(5, 3^2)$ and its NDS consistent estimate $\tilde{X} \sim N(6, 5^2)$.

2.1.5 Normalized Deviation Squared equivalence

N-variate random variable $\tilde{X}$ with mean $\bar{x} \in \mathbb{R}^N$ and covariance $\bar{P}$ is an NDS equivalent estimate of N-variate random variable X if

$$\forall \epsilon > 0 : \ \Pr\left\{(X - \bar{x})^T \bar{P}^{-1}(X - \bar{x}) \le \epsilon\right\} = \Pr\left\{(X - \bar{x})^T \bar{P}^{-1}(X - \bar{x}) \le \epsilon\right\} \quad (2.16)$$

Inequality (2.16) means that probability mass of $\tilde{X}$ in any ellipsoid defined by $(z - \bar{x})^T \bar{P}^{-1}(z - \bar{x}) < \epsilon$ is equal to probability mass of X in the same ellipsoid. I.e. normalized deviation squared equivalence implies p equivalence for any $p \in [0, 1]$.

2.2 Information content

Information content reflects how certain estimation is about the parameter being estimated: the higher the certainty of the estimate, more valuable information it possesses. Information content is directly related to the inverse of the covariance matrix of an estimate. In scalar case, the smaller variance of RV is, the more certain is the estimation.

It is possible to make estimate consistent by increasing its covariance matrix, but this would reduce the informativity of the estimation. The aim is to make the estimate as informative as possible and at the same time to preserve consistency. It means that inequalities in the consistency definitions must be as close to equalities as possible.

In case of mean squared deviation consistency, estimate defined by $\bar{x}$ and $\bar{P}$ is most informative and yet consistent in mean squared deviation if $\bar{P} = M$.

In case of p consistency, estimate $\tilde{X}$ is most informative and yet p consistent if

$$\exists \epsilon > 0 \mid \Pr\left\{(X - \bar{x})^T \bar{P}^{-1}(X - \bar{x}) \leq \epsilon\right\} = \Pr\left\{(\tilde{X} - \bar{x})^T \bar{P}^{-1}(\tilde{X} - \bar{x}) \leq \epsilon\right\} = p,$$

i.e. if $\tilde{X}$ is p equivalent estimate of X.

In case of NDS consistency, estimate $\tilde{X}$ is most informative and yet NDS consistent if

$$\forall \epsilon > 0 : \Pr\left\{(X - \bar{x})^T \bar{P}^{-1}(X - \bar{x}) \leq \epsilon\right\} = \Pr\left\{(\tilde{X} - \bar{x})^T \bar{P}^{-1}(\tilde{X} - \bar{x}) \leq \epsilon\right\},$$

i.e. if $\tilde{X}$ is NDS equivalent estimate of X.

Chapter 3

Hypothesis testing for consistency evaluation

3.1 Concept of hypothesis testing

In statistics, hypothesis is a statement about some phenomena or population whose truthfulness has to be checked. Hypothesis test starts with making the null hypothesis H_0, which is the statement that is rejected or not rejected as a result of the test. In contrast to null hypothesis there exist alternative hypothesis H_1 that is the logical complement of H_0. Hypothesis test assumes two possible outputs:

- H_0 is rejected in favor of H_1
- H_0 is not rejected

Null hypothesis is rejected if after obtaining observations about phenomena/population there is a sufficient evidence that it is false. And hypothesis is not rejected if there is not sufficient evidence that it is false. However, the latter case does not mean that null hypothesis is true.
Hypothesis inference is made based on the test statistic $U \in \mathbb{R}^m$ associated with the hypothesis and critical region $r \subset \mathbb{R}^m$, which is defined based on probability distribution of U and significance level α. Test statistic U is considered as a random variable whose value is calculated according to the observations obtained during the test. Critical region r is defined so that

$$\Pr\{U \in r|H_0\} \leq \alpha,$$

where α is usually small value, e.g. 0.05. Hypothesis is rejected if U falls inside critical region, and it is not rejected otherwise.[19]
The underlying concept of hypothesis testing is that if H_0 is true then probability that U will fall inside the critical region is very small (equal to or less than α), and if U actually falls inside r then one would reject hypothesis rather than accept happening of such a rare event.
Hypothesis tests never produce conclusions about the statement with absolute certainty. There are two types of errors that can be made in hypothesis inference. Type 1 error occurs if the null hypothesis is rejected when it is actually true. Type 2 error occurs if hypothesis is not rejected when it is actually false. Correct decisions and errors of type 1 and 2 are summarized in the table 3.1.

Table 3.1: Possible hypothesis test outcomes

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type 2 error
Reject H_0	Type 1 error	Correct decision

Good hypothesis test must have minimal probabilities of type 1 and type 2 errors so that probabilities of correct decisions are maximized. In some cases probability of errors of second type can not be directly calculated since alternative hypothesis H_1 includes large number of events for which U might fall outside critical region r with high probability. In this case only rejection of the hypothesis is reliable, since rejection excludes the type 2 error, and type 1 error is known.
Summarizing all above, hypothesis test consists of the following steps:

1. Define null hypothesis H_0
2. Define test statistics U
3. Define significance level α and set critical region r so that $\Pr\{U \in r|H_0\} \leq \alpha$
4. Make observations and calculate U
5. Reject H_0 if U falls inside critical region r, do not reject H_0 if U falls outside critical region r.

3.2 Hypothesis tests for consistency evaluation

In order to check whether parameter estimate meet consistency requirements, actual distribution of parameter is needed, which is almost never available. In most of the cases the only information about the true distribution is represented by a sample drawn from it. If only a sample is available, hypothesis test can be used to evaluate consistency of the estimate, with null hypothesis H_0 being "Estimate is consistent", and test statistic being a function of the sample.

3.2.1 Hypothesis test for Mean Squared Deviation consistency

Consider estimate of N-variate RV X given in the form of point estimate $\bar{x} \in \mathbb{R}^N$ and corresponding mean squared deviation matrix $\bar{P}$, and random sample $X_1, \ldots, X_M$ from the distribution of X.
Let H_0 be a null hypothesis stating that estimate is consistent in mean squared deviation. Choose $\epsilon > N$ and define test variable U as

$$U = \sum_{i=1}^{M} U_i,$$

where

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

If H_0 is true then according to the definition of mean squared deviation consistency

$$\Pr\left\{(X - \bar{x})^T \bar{P}^{-1} (X - \bar{x}) \leq \epsilon\right\} \geq 1 - \frac{N}{\epsilon}, \tag{3.1}$$

and since X_i are independent realizations of X, U is a sum of M independent Bernoulli random variables with probability of success $p \geq 1 - \frac{N}{\epsilon}$, i.e. U is a Binomial random variable with probability mass function

$$\Pr\{U = k\} = \binom{M}{k} p^k (1-p)^{M-k}, k \in \{0, \ldots, M\}.$$

In order to define critical region for U, consider function $f(y) = \binom{M}{k} y^k (1-y)^{M-k}$ which has positive derivative on $(0, \frac{k}{M})$ and negative derivative on $(\frac{k}{M}, 1)$. Taking into account that $p \geq 1 - \frac{N}{\epsilon}$, we can guarantee that for k such that

$$1 - \frac{N}{\epsilon} \geq \frac{k}{M}, \tag{3.2}$$

following inequality holds

$$\binom{M}{k} p^k (1-p)^{M-k} \leq \binom{M}{k} \left(1 - \frac{N}{\epsilon}\right)^k \left(\frac{N}{\epsilon}\right)^{M-k}. \tag{3.3}$$

We can choose $K \in \{0, \ldots, M\}$ so that $k \in \{0, \ldots, K\}$ satisfies (3.2), and for which (3.3) holds. Then critical region for U can be set as $r = \{0, \ldots, K\}$ so that

$$\Pr\{U \in r | H_0\} = \sum_{k=0}^{K} \binom{M}{k} p^k (1-p)^{M-k} \leq \sum_{k=0}^{K} \binom{M}{k} \left(1 - \frac{N}{\epsilon}\right)^k \left(\frac{N}{\epsilon}\right)^{M-k} \leq \alpha,$$

where α is the significance level of the hypothesis test. If U falls inside r, hypothesis H_0 can be rejected at significance level α.
Critical region r is dependent on significance level α and value ϵ, and for certain choices critical region might be empty. General rule is to choose α and ϵ so that $|r| > 0$ to make hypothesis test sensible.

3.2.2 Hypothesis test for P-consistency and equivalence

Consider N-variate RV X, its estimate $\tilde{X}$ with known probability distribution, mean $\bar{x}$ and covariance $\bar{P}$, and random sample $X_1, \ldots, X_M$ drawn from the distribution of X.
Let H_0 be a null hypothesis stating that estimate is p-consistent. Choose $\epsilon \geq 0$ such that $\Pr\{(\tilde{X} - \bar{x})^T \bar{P}^{-1} (\tilde{X} - \bar{x}) \leq \epsilon\} = p$ and define test variable U as

$$U = \sum_{i=1}^{M} U_i,$$

where

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

If H_0 is true then according to the definition of p-consistency

$$\Pr\left\{(X_i - \bar{x})^T \bar{P}^{-1} (X_i - \bar{x}) \leq \epsilon\right\} \geq \Pr\left\{(\tilde{X} - \bar{x})^T \bar{P}^{-1} (\tilde{X} - \bar{x}) \leq \epsilon\right\} = p, \tag{3.4}$$

and since X_i are independent realizations of RV X, U is a sum of M independent

Bernoulli random variables with probability of success $s > p$, i.e. U is a Binomial random variable with probability mass function

$$\Pr\{U = k\} = \binom{M}{k} s^k (1-s)^{M-k}.$$

The critical region for U can be defined as in hypothesis test for mean squared deviation consistency (section 3.2.1) by replacing $1 - \frac{N}{\epsilon}$ with p. I.e. we can choose $K \in \{0, \dots, M\}$ so that $k \in \{0, \dots, K\}$ satisfy $p \geq \frac{k}{M}$ and

$$\binom{M}{k} s^k (1-s)^{M-k} \leq \binom{M}{k} (p)^k (1-p)^{M-k}. \tag{3.5}$$

Then critical region for U can be set as $r = \{0, \dots, K\}$ so that

$$\Pr\{U \in r | H_0\} = \sum_{k=0}^{K} \binom{M}{k} s^k (1-s)^{M-k} \leq \sum_{k=0}^{K} \binom{M}{k} p^k (1-p)^{M-k} \leq \alpha. \tag{3.6}$$

If U falls inside r then H_0 can be rejected at significance level α.
In order to check p equivalence, let H_0 be a null hypothesis stating that $\tilde{X}$ is a p equivalent estimate of X and let test statistic U be defined as previously. If H_0 is true then

$$\Pr\left\{(X_i - \bar{x})^T \bar{P}^{-1} (X_i - \bar{x}) \leq \epsilon\right\} = \Pr\left\{(\tilde{X} - \bar{x})^T \bar{P}^{-1} (\tilde{X} - \bar{x}) \leq \epsilon\right\} = p, \tag{3.7}$$

and U is a binomial RV with probability mass function

$$\Pr\{U = k\} = \binom{M}{k} p^k (1-p)^{M-k}. \tag{3.8}$$

For significance level α, two sided critical region for U can be set as $r = \{0, \dots, k_1\} \cup \{k_2, \dots, M\}$, so that

$$\Pr\{U \in r | H_0\} = \sum_{k=0}^{k_1} \binom{M}{k} p^k (1-p)^{M-k} + \sum_{k=k_2}^{M} \binom{M}{k} p^k (1-p)^{M-k} \leq \alpha. \tag{3.9}$$

$P-$equivalence test with two sided critical region for U can be used to check both consistency and information content of the estimate. If U falls into the right part of critical region then estimate is consistent but uninformative, if U falls into the left part of the critical region then estimate is inconsistent.
Consider as an example a scalar random variable X distributed normally with mean $m = 5$ and standard deviation $\sigma = 3$ and its estimate $\tilde{X}$ distributed normally with mean $\tilde{m} = 8$ and standard deviation $\tilde{\sigma} = 4$. Let's apply the hypothesis test to check 0.68 consistency of $\tilde{X}$.

For significance level $\alpha = 0.1$, sample size $M = 20$, and $p = 0.68$, critical region for test statistic U can be defined as $r = [0, 10]$ according to (3.6).
Now consider random sample [6.3091, 3.4869, 5.3063, 8.5888, 5.3608, 1.8895, 2.4287, 4.4904, 4.4250, 2.4026, 5.5420, 8.7996, 4.2465, 4.3863, -1.6046, 2.6765, 0.8202, 3.8413, 6.5768, 9.5698] drawn from the distribution of X. $U = 8$ calculated based on this sample falls inside the critical region $r = [0, 10]$, thus, we can reject hypothesis about 0.68 consistency of $\tilde{X}$ with significance level α.
It can be checked that $\Pr\{\tilde{X} \in [\tilde{m} - \tilde{\sigma}, \tilde{m} + \tilde{\sigma}]\} = 0.68$ but $\Pr\{X \in [\tilde{m} - \tilde{\sigma}, \tilde{m} + \tilde{\sigma}]\} = 0.62$, thus $\tilde{X}$ is not p consistent estimate of X for $p = 0.68$.

3.2.3 Hypothesis test for Normalized Deviation Squared consistency and equivalence

In order to define hypothesis test for normalized deviation squared consistency we need the following Lemma and Theorem.

Lemma 3.1. *If functions $f(x) \geq 0$, $g(x) \geq 0$, $s(x) \geq 0$ have finite number of discontinuity points, $s(x)$ is monotonically decreasing, and*

$$\forall \epsilon \geq 0 : \int_0^{\epsilon} f(x)dx \geq \int_0^{\epsilon} g(x)dx \tag{3.10}$$

then

$$\forall \epsilon \geq 0 : \int_0^{\epsilon} f(x)s(x)dx \geq \int_0^{\epsilon} g(x)s(x)dx \tag{3.11}$$

Proof. Consider integral difference

$$\int_0^{\epsilon} f(x)s(x)dx - \int_0^{\epsilon} g(x)s(x)dx. \tag{3.12}$$

Denote $r(x) = f(x) - g(x)$, and represent interval $[0, \epsilon]$ as $\bigcup_{i=1}^{N} a_i$, where $a_i \cap a_j = \emptyset$, $r(x)$ is either non-negative or negative on each of the segments and changes sign every time the segment is changed. Due to (3.10), $r(x) \geq 0$ for $x \in a_1$.
When N is even integral difference (3.12) can be rewritten as

$$\int_0^{\epsilon} r(x)s(x)dx = \int_{a_1} r(x)s(x)dx + \ldots + \int_{a_N} r(x)s(x)dx \tag{3.13}$$

Condition (3.10) imply that

$$\forall k : \int_{a_1} r(x)dx + \ldots + \int_{a_k} r(x)dx \geq 0.$$

Since $s(x)$ is monotonically decreasing

$$\min_{a_i} s(x) \geq \max_{a_j} s(x), \min_{a_i} s(x) \geq \min_{a_j} s(x) \text{ for } i < j$$

Therefore

$$\begin{aligned}\int_{a_1} r(x)s(x)dx + \int_{a_2} r(x) + \ldots + \int_{a_N} r(x)s(x)dx &\geq \\ \int_{a_1} r(x)\min_{a_1}(s(x))dx + \ldots + \int_{a_N} r(x)\max_{a_N}(s(x))dx &\geq \\ \max_{a_N}(s(x))\left(\int_{a_1} r(x)dx + \ldots + \int_{a_N} r(x)dx\right) &\geq 0\end{aligned}$$

When N is odd

$$\int_{a_1} r(x)s(x)dx + \int_{a_2} r(x)s(x)dx + \ldots + \int_{a_{N-1}} r(x)s(x)dx \geq 0, \tag{3.14}$$

and since $r(x)$ is non-negative on a_1 and changes its sign whenever segment is changed

$$\int_{a_N} r(x)s(x)dx \geq 0.$$

Thus,

$$\int_{a_1} r(x)s(x)dx + \ldots + \int_{a_N} r(x)s(x)dx \geq 0 \Rightarrow \int_0^{\epsilon} f(x)s(x)dx \geq \int_0^{\epsilon} g(x)s(x)dx.$$

□

Theorem 3.2. *Consider continuous mutually independent RVs* $[U_1, \ldots, U_M]$, $[\tilde{U}_1, \ldots, \tilde{U}_M]$ *defined on* $[0, +\infty)$ *with probability density functions* $p_{U_1}(u_1), \ldots, p_{U_M}(u_M), p_{\tilde{U}_1}(u_1), \ldots, p_{\tilde{U}_M}(u_M)$ *having finite number of discontinuity points, and RVs* $U = \sum_{k=1}^{M} U_k$, $\tilde{U} = \sum_{k=1}^{M} \tilde{U}_k$. *If*

$$\forall \epsilon > 0 \,:\, Pr(U_k \leq \epsilon) \geq Pr(\tilde{U}_k \leq \epsilon), \tag{3.15}$$

then

$$\forall \epsilon > 0 \,:\, Pr(U \leq \epsilon) \geq Pr(\tilde{U} \leq \epsilon). \tag{3.16}$$

Proof. Consider probabilities $\Pr(\tilde{U} \leq \epsilon)$ and $\Pr(U \leq \epsilon)$ which are equal respectively to the following multiple integrals:

$$\int_0^{\epsilon} p_{\tilde{U}_1}(u_1) \ldots \int_0^{\epsilon - u_1 - \ldots - u_{M-1}} p_{\tilde{U}_M}(u_M) du_M \ldots du_1,$$

$$\int_0^{\epsilon} p_{U_1}(u_1)\ldots\int_0^{\epsilon-u_1-\ldots-u_{M-1}} p_{U_M}(u_M)du_M\ldots du_1.$$

Denote

$$\epsilon_i = \epsilon - \sum_{k=1}^{i} u_k,$$

$$s_i = \int_0^{e_i} p_{U_{i+1}}(u_{i+1})\ldots\int_0^{e_{M-1}} p_{U_M}(u_M)du_m\ldots du_{i+1},$$

$$\tilde{s}_i = \int_0^{e_i} p_{\tilde{U}_{i+1}}(u_{i+1})\ldots\int_0^{e_{M-1}} p_{\tilde{U}_M}(u_M)du_m\ldots du_{i+1},$$

$$i \in \{1,\ldots,M-1\}.$$

Since probability density functions $p_{\tilde{U}_1}(u_1),\ldots,p_{\tilde{U}_M}(u_M)$, $p_{U_1}(u_1),\ldots,p_{U_M}(u_M)$ have finite number of discontinuity points, functions $s_1,\ldots,s_{M-1}$ also posses the same property. According to (3.15)

$$\forall\eta>0\ :\ \int_0^{\eta} p_{\tilde{U}_k}(u_k)du_k \le \int_0^{\eta} p_{U_k}(u_k)du_k, k\in\{1,\ldots,M\}.$$

For fixed $u_1,\ldots,u_{i-1}$, s_i is a non-negative function of u_i that monotonically decreases. Therefore, lemma requirements are met and we can apply it as follows.
Since

$$\tilde{s}_{M-1} \le s_{M-1},$$

and according to the lemma,

$$\begin{aligned}\tilde{s}_{M-2}=&\int_0^{\epsilon_{M-2}} p_{\tilde{U}_{M-1}}(u_{M-1})\tilde{s}_{M-1}du_{M-1} \le \int_0^{\epsilon_{M-2}} p_{\tilde{U}_{M-1}}(u_{M-1})s_{M-1}du_{M-1}\\ \le&\int_0^{\epsilon_{M-2}} p_{U_{M-1}}(u_{M-1})s_{M-1}du_{M-1} = s_{M-2}.\end{aligned}$$

Now, since

$$\tilde{s}_{M-2} \le s_{M-2},$$

and according to the lemma,

$$\begin{aligned}\tilde{s}_{M-3}=&\int_0^{\epsilon_{M-3}} p_{\tilde{U}_{M-2}(U_{M-2})}\tilde{s}_{M-2}du_{M-2} \le \int_0^{\epsilon_{M-3}} p_{\tilde{U}_{M-2}}(u_{M-2})s_{M-2}du_{M-2}\\ \le&\int_0^{\epsilon_{M-3}} p_{U_{M-2}}(u_{M-2})s_{M-2}du_{M-2} = s_{M-3}.\end{aligned}$$

Proceeding in this way, it can be shown that

$$\tilde{s}_1 \le s_1,$$

and thereby according to the lemma,

$$\Pr(\tilde{U} \leq \epsilon) = \int_0^{\epsilon} p_{\tilde{U}_1}(u_1)\tilde{s}_1 du_1 \leq \int_0^{\epsilon} p_{\tilde{U}_1}(u_1) s_1 du_1$$
$$\leq \int_0^{\epsilon} p_{U_1}(u_1) s_1 du_1 = \Pr(U \leq \epsilon).$$

□

Now we can define hypothesis test as follows. Consider continuous N-variate RV X, its N-variate estimate $\tilde{X}$ with known probability distribution, mean $\bar{x}$ and covariance $\bar{P}$, and random independent sample $X_1, \ldots, X_M$ from the distribution of X. Probability density functions of X and $\tilde{X}$ are assumed to have finite number of discontinuity points.

Let H_0 be a null hypothesis stating that estimate is NDS consistent. Define test statistics U and $\tilde{U}$ as

$$U = \sum_{i=1}^{M} U_i, \; \tilde{U} = \sum_{i=1}^{M} \tilde{U}_i, \tag{3.17}$$

where

$$U_i = (X_i - \bar{x})^T \bar{P}^{-1} (X_i - \bar{x}),$$
$$\tilde{U}_i = (\tilde{X}_i - \bar{x})^T \bar{P}^{-1} (\tilde{X}_i - \bar{x}), \tag{3.18}$$

$\tilde{X}_i, i \in \{1, \ldots, M\}$ are independent and identically distributed according to distribution of $\tilde{X}$.

If H_0 is true then according to the definition of NDS consistency

$$\forall \epsilon > 0 \,:\, \Pr\left\{(X_i - \bar{x})^T \bar{P}^{-1} (X_i - \bar{x}) \leq \epsilon\right\} \geq \Pr\left\{(\tilde{X}_i - \bar{x})^T \bar{P}^{-1} (\tilde{X}_i - \bar{x}) \leq \epsilon\right\}, \tag{3.19}$$

i.e.

$$\forall \epsilon > 0 \,:\, \Pr\{U_i \leq \epsilon\} \geq \Pr\{\tilde{U}_i \leq \epsilon\} \tag{3.20}$$

Random variables U_i's and $\tilde{U}_i$'s are mutually independent and their probability density functions are defined on $[0, +\infty)$ and have finite number of discontinuity points. Therefore according to the Theorem 3.2

$$\forall \epsilon > 0 \,:\, \Pr\{U \leq \epsilon\} \geq \Pr\{\tilde{U} \leq \epsilon\}. \tag{3.21}$$

Since distribution of $\tilde{X}$ is available, $\Pr\{\tilde{U} \geq \epsilon_1\}$ can be calculated, and for significance level α, critical region for U can be defined as $r = [e_1, +\infty)$, so that

$$\Pr\{\tilde{U} \geq \epsilon_1\} \leq \alpha.$$

Then $\Pr\{U \in r|H_0\} = \Pr\{U \geq \epsilon_1|H_0\} \leq \Pr\{\tilde{U} \geq \epsilon_1\} \leq \alpha$, and if U falls inside r, hypothesis H_0 can be rejected at significance level α.
In order to check NDS equivalence of the estimate, let H_0 be a null hypothesis stating that RV $\tilde{X}$ is an NDS equivalent estimate of X, and let test statistics U and $\tilde{U}$ be defined as previously. If H_0 is true then according to definition of NDS equivalence

$$\forall\epsilon : \Pr\left\{(X_i - \bar{x})^T \bar{P}^{-1}(X_i - \bar{x}) \leq \epsilon\right\} = \Pr\left\{(\tilde{X}_i - \bar{x})^T \bar{P}^{-1}(\tilde{X}_i - \bar{x}) \leq \epsilon\right\}, \quad (3.22)$$

and hence

$$\forall\epsilon : \Pr\{U \leq \epsilon\} = \Pr\{\tilde{U} \leq \epsilon\}. \quad (3.23)$$

Since distribution of $\tilde{X}$ is available, $\Pr\{\tilde{U} \leq \epsilon\}$ can be calculated, and for significance level α, two sided critical region for U can be defined as $r = [0, \epsilon_1] \cup [\epsilon_2, +\infty)$, so that

$$\Pr\{\tilde{U} \in r\} \leq \alpha.$$

Then $\Pr\{U \in r|H_0\} = \Pr\{\tilde{U} \in r\} \leq \alpha$.
NDS equivalence test with two sided critical region for U can be used in order to check both consistency and information content of the estimate. If U falls into the left part of the critical region then estimate is consistent but uninformative, if U falls into the right part of the critical region then estimate is inconsistent.
Consider as an example a scalar random variable X distributed normally with mean $m = 5$ and standard deviation $\sigma = 3$ and its estimate $\tilde{X}$ distributed normally with mean $\tilde{m} = 8$ and standard deviation $\tilde{\sigma} = 4$.
Earlier it was shown that $\tilde{X}$ is not 0.68 consistent, hence it is not NDS consistent. Let's illustrate this by applying hypothesis test for NDS consistency. Since $\tilde{X}$ is normally distributed scalar RV, test statistic U is distributed according to chi-square with M degrees of freedom, where M is the size of the sample. For significance level $\alpha = 0.1$ and $M = 20$, one sided critical region for U is set as $r = [28.4, +\infty]$, according to χ^2_{20} distribution.
Now consider random sample [6.6590, 1.7706, 8.0919, 5.9826, 6.9564, 4.1634, 5.7356, 9.4175, -1.8253, 0.1001, 6.2464, 3.0357, 4.1110, 0.5092, 2.2855, 3.7875, 2.8226, 2.4005, 3.7345, 2.1720] drawn from the distribution of X. Test statistic $U = 54.0276$ calculated based on this sample falls inside the critical region $r = [28.4, +\infty]$, thereby we can reject NDS consistency of $\tilde{X}$ with significance level α.

Chapter 4

Filters

4.1 Dynamical systems

In filtering theory systems of interest are described by their inner states, state transition model, and measurement model. At each time moment system is characterized by a state vector, e.g. for a moving target state vector consists of coordinates in three dimensional space, and vector of velocity. State transition model describes dynamical properties of the system, i.e. how states of the system evolve in time. Measurement model describes relation between states and measurements of the system. Noises represented by a random variables are used in order to take into account modeling inaccuracies.
Dynamical system can be described by the following system of equations:

$$\begin{aligned} x_k &= f_{k-1}(x_{k-1}, w_{k-1}) \\ y_k &= h_k(x_k, v_k) \\ x_0 &\sim p_{x_0}(x) \\ w_{k-1} &\sim p_{w_{k-1}}(w) \\ v_k &\sim p_{v_k}(v) \end{aligned}$$

In the above system of equations x_k is a state vector at time t_k, $k \in \{0, 1, 2, ...\}$, $f_{k-1}(x_{k-1}, w_{k-1})$ is a state transition model of the system, y_k is a measurement vector at time t_k, $h_k(x_k, v_k)$ is a measurement model of the system, initial state x_0 is distributed according to $p_{x_0}(x)$, w_{k-1} and v_k are mutually independent white noises distributed according to $p_{w_{k-1}}(w)$ and $p_{v_k}(v)$ respectively. Initial state x_0 is also independent of process and measurement noises [14].
System might also be controlled, in this case it includes control parameters: control input and control-input model. For simplicity of explanation and without loss of generality non-controlled systems are considered in this thesis.
Given the system above, filter estimates unknown vector x_k at time moment t_k, given measurements $y_{1:k} = \{y_1, \cdots, y_k\}$ and distribution of the initial state $p_{x_0}(x)$, i.e. calculate the posterior distribution $p(x_k|x_0, y_{1:k})$.

4.2 Recursive Bayesian filter

Bayesian filter estimates posterior distribution of the state of the system based on the Bayes' rule:

$$p_X(x|Y = y) = \frac{p_Y(y|X = x)p_X(x)}{p_Y(y)} \tag{4.1}$$

In Bayesian filtering system process is considered as Markov process, which means that current system state depends only on state at previous time moment but not on states that took place earlier.
At each time moment system state can be represented as a random variable conditional on previous states and measurements. Suppose that at some time moment distribution of the state vector is available, i.e. $p(x_{k-1}|y_{1:k-1})$ is known. Because of Chapman-Kolmogorov equation and the fact that system states are from Markov process, following holds

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1} \tag{4.2}$$

To prove this, consider joint distribution $p(x_k, x_{k-1}|y_{1:k-1})$. According to Chapman-Kolmogorov equation

$$p(x_k|y_{1:k-1}) = \int p(x_k, x_{k-1}|y_{1:k-1})dx_{k-1}.$$

Therefore,

$$p(x_k|y_{1:k-1}) = \int p(x_k, x_{k-1}|y_{1:k-1})dx_{k-1} =$$

$$= \int p(x_k|x_{k-1}, y_{1:k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1} =$$

$$= \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1}.$$

The last equation holds because of the Markov property of the system states. Equation (4.2) defines the prediction step of the recursive Bayesian filter.
According to the Bayes rule

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})} \propto p(y_k|x_k)p(x_k|y_{1:k-1}). \tag{4.3}$$

Equation (4.3) defines the update step of the recursive Bayesian filter. Filter is recursive since posterior of the state can be calculated iteratively using only posterior from the previous step and current measurement.[14]

4.3 Kalman Filter

Kalman filter [4] is an optimal Bayesian filter that analytically solves estimation problems for linear systems with additive Gaussian noises. It provides analytical solution of the estimation problem and is also computationally simple, but is only applicable to linear systems. This is why several KF extensions were developed for non-linear systems during the last decades. KF algorithm with some of its non-linear extensions are presented next.

4.3.1 Kalman Filter

In Linear Gaussian (LG) systems, state transition and measurement models are defined by linear operators, initial state is a Gaussian RV, and noises of the system are zero-mean, white, Gaussian, and independent of each other and initial state. Linear Gaussian system can be defined as follows:

$$\begin{aligned} x_0 &\sim N(\bar{x}_0, \bar{P}_0) \\ x_k &= F_{k-1}x_{k-1} + w_{k-1} \\ y_k &= H_k x_k + v_k \\ w_{k-1} &\sim N(0, Q_{k-1}) \\ v_k &\sim N(0, R_k) \end{aligned}$$

Kalman filter calculates optimal posterior distribution of the system state in the form of normal distribution with mean $\bar{x}_{k|k}$ and covariance $\bar{P}_{k|k}$, which are calculated recursively according to the following algorithm [13]:

1. $\bar{x}_{0|0} = \bar{x}_0, \bar{P}_{0|0} = \bar{P}_0$
2. $\bar{x}_{k|k-1} = F_{k-1}\bar{x}_{k-1|k-1}$
3. $\bar{P}_{k|k-1} = F_{k-1}\bar{P}_{k-1|k-1}F_{k-1}^T + Q_{k-1}$
4. $\eta_k = y_k - H_k\bar{x}_{k|k-1}$
5. $S_k = R_k + H_k\bar{P}_{k|k-1}H_k^T$
6. $K_k = \bar{P}_{k|k-1}H_k^T S_k^{-1}$
7. $\bar{x}_{k|k} = \bar{x}_{k|k-1} + K_k\eta_k$
8. $\bar{P}_{k|k} = (\mathrm{I} - K_kH_k)\bar{P}_{k|k-1}$
9. $k = k + 1$, go to step 2.

4.3.2 Extended Kalman filter

In many applications systems have non-linear state transition or measurement equations. In Non-linear Gaussian systems state transition and measurement models can be defined by non-linear functions, initial state is a Gaussian RV, and noises of the system are zero-mean, white, Gaussian, and independent of each other and initial state. Non-linear Gaussian system can be defined as follows:

$$\begin{aligned} x_0 &\sim N(\bar{x}_0, \bar{P}_0) \\ x_k &= f_{k-1}(x_{k-1}) + w_{k-1} \\ y_k &= h_k(x_k) + v_k \\ w_{k-1} &\sim N(0, Q_{k-1}) \\ v_k &\sim N(0, R_k) \end{aligned}$$

For such a systems Kalman filter is not applicable and analytical solution might be intractable in general. In this case Extended Kalman Filter (EKF) [7] can be used instead. EKF approximates non-linear functions $f_{k-1}(x)$ and $h_k(x)$ by first order Taylor polynomials in the neighborhood of posterior mean $\bar{x}_{k-1|k-1}$ at previous time moment and prior mean $\bar{x}_{k|k-1}$ at current time moment. EKF algorithm has the following steps:

1. $\bar{x}_{0|0} = \bar{x}_0, \bar{P}_{0|0} = \bar{P}_0$
2. $\bar{x}_{k|k-1} = f_{k-1}(\bar{x}_{k|k-1})$
3. $F_{k-1} = \frac{\partial f_{k-1}(x_{k-1})}{\partial x_{k-1}}, x_{k-1} = \bar{x}_{k|k-1}$
4. $\bar{P}_{k|k-1} = F_{k-1}\bar{P}_{k-1|k-1}F_{k-1}^T + Q_{k-1}$
5. $H_k = \frac{\partial h_k(x_k)}{\partial x_k}, x_k = \bar{x}_{k|k-1}$
6. $\eta_k = y_k - H_k\bar{x}_{k|k-1}$
7. $S_k = R_k + H_k\bar{P}_{k|k-1}H_k^T$
8. $K_k = \bar{P}_{k|k-1}H_k^T S_k^{-1}$
9. $\bar{x}_{k|k} = \bar{x}_{k|k-1} + K_k\eta_k$
10. $\bar{P}_{k|k} = (\mathrm{I} - K_k H_k)\bar{P}_{k|k-1}$
11. $k = k + 1$, go to step 2.

4.3.3 Linear Regression Kalman Filter

Linear regression Kalman filter (LRKF) [9, 10] is another variant of non-linear KF that is used for non-linear Gaussian systems. It differs from the EKF in the way it linearizes the state transition and measurement functions. Additionally, LRKF estimates linearization errors and takes them into account as additional noises of the system.

LRKF uses function values of r regression points $\chi^i_{k-1|k-1}$ in the region of uncertainty around mean of the previous posterior estimate $\bar{x}_{k-1|k-1}$ to model behavior of non-linear state transition and measurement functions f_{k-1} and h_k. Particular choices of regression points correspond to different variants of LRKF [10], e.g. Centered Deviation Kalman Filter (CDKF) [17] or Unscented Kalman Filter (UKF) [8]. Using values of the function f_{k-1} at the regression points, linearized functions F_{k-1}, H_k and linearization errors Q^*_{k-1}, R^*_k are calculated so that deviation between function values of linearized functions and original ones is minimized in least square sense. Additional linearization noises $w^*_{k-1} \sim N(0, Q^*_{k-1})$, $v^*_k \sim N(0, R^*_k)$ are included in the model. Linearized functions and linearization errors are calculated as follows [10]:

$$\chi^i_{k|k-1} = f_{k-1}(\chi^i_{k-1|k-1})$$

$$e_i = \chi^i_{k|k-1} - F\chi^i_{k-1|k-1}$$

$F_{k-1} = \underset{F}{\operatorname{argmin}} \sum_{i=1}^{r} e_i^T e_i$

$Q_{k-1}^* = \frac{1}{r} \sum_{i=1}^{r} e_i e_i^T$

$e_i = h(\chi_{k|k-1}^i) - H\chi_{k|k-1}^i$

$H_k = \underset{H}{\operatorname{argmin}} \sum_{i=1}^{r} e_i^T e_i$

$R_k^* = \frac{1}{r} \sum_{i=1}^{r} e_i e_i^T$

LRKF requires more computational resources than EKF, however, it explicitly estimates linearization errors which are then added to the system noises, while in the EKF linearization errors are not taken into account.

4.4 Particle Filter

Particle filter (PF) is a Sequential Monte Carlo algorithm based on particle representation of probability densities. PF is used when the system equations are sophisticated and estimation problem is analytically intractable. Moreover, PF accepts more general system models than non-linear Kalman Filters, namely, PF does not require system noises to be Gaussian, and accepts models of the following form

$$
\begin{aligned}
x_0 &\sim p_{x_0}(x) \\
x_k &= f_{k-1}(x_{k-1}, w_{k-1}) \\
y_k &= h_k(x_k, v_k) \\
w_{k-1} &\sim p_{w_{k-1}}(w) \\
v_k &\sim p_{v_k}(v)
\end{aligned}
$$

w_{k-1} and v_k are mutually independent, white noises with known probability distribution functions $p_{w_{k-1}}(w)$ and $p_{v_k}(v)$. Even though PF solves more general problems, in order to make reliable estimates, it must use reasonably large amount of particles, which makes it computationally heavier then EKF or LRKF.

4.4.1 Monte Carlo Integration

Consider an integral

$$I = \int g(x)dx. \tag{4.4}$$

In Monte Carlo Methods $g(x)$ is factorized so that $g(x) = f(x)\pi(x)$, $\pi(x) \geq 0$ and $\int \pi(x)dx = 1$. By drawing large enough sample $\left\{x^1, \ldots, x^N\right\}, 1 \ll N$ from the distribution $\pi(x)$, integral

$$I = \int g(x)dx = \int f(x)\pi(x)dx \tag{4.5}$$

can be approximated by sample mean

$$I_N = \frac{1}{N}\sum_{i=1}^{N} f(x^i). \tag{4.6}$$

If x_i's are independent, I_N is an unbiased estimate of I, and, according to the law of large numbers, I_N will almost surely converge to I. If the variance

$$\sigma^2 = \int (f(x) - I)^2\pi(x)dx \tag{4.7}$$

is finite, then, according to central limit theorem, estimation error converges in distribution:

$$\lim_{N\to\infty} \sqrt{N}(I_N - I) \sim N(0, \sigma^2). \tag{4.8}$$

[14]

4.4.2 Posterior moments approximation

In particle filtering Monte Carlo Integration is used to estimate moments of state posterior distribution $\pi(x)$. If $\pi(x)$ is complicated and sample cannot be drawn from it, another distribution $q(x)$, which is similar to $\pi(x)$ and from which sample can be drawn, can be used to approximate moments of $\pi(x)$. Distribution $q(x)$ is called the proposal distribution and it is similar to $\pi(x)$ if from $\pi(x) > 0$ follows $q(x) > 0$ for any $x \in R^{n_x}$ (i.e. support of q include support of π). If $\frac{\pi(x)}{q(x)}$ is bounded, moments of $\pi(x)$, which are generally equal to $\int f(x)\pi(x)dx$, can be rewritten as

$$I = \int f(x)\pi(x)dx = \int f(x)\frac{\pi(x)}{q(x)}q(x)dx. \tag{4.9}$$

Monte Carlo estimate of the integral can be computed by drawing N independent samples $\{x_i; i = 1, ..., N\}$ from the distribution $q(x)$ and calculating their weighted sum

$$I_N = \frac{1}{N}\sum_{i=1}^{N} f(x^i)\widetilde{w}(x^i), \quad \widetilde{w}(x^i) = \frac{\pi(x^i)}{q(x^i)}. \tag{4.10}$$

If density $\pi(x)$ is not normalized then importance weights must be normalized and integral is estimated as

$$I_N = \frac{\frac{1}{N}\sum_{i=1}^{N} f(x^i)\widetilde{w}(x^i)}{\frac{1}{N}\sum_{j=1}^{N} \widetilde{w}(x^j)} = \sum_{i=1}^{N} f(x^i)w(x^i), \tag{4.11}$$

where $w(x^i) = \frac{\widetilde{w}(x^i)}{\sum_{j=1}^{N}\widetilde{w}(x^j)}$ are normalized weights. [14]

4.4.3 Particle filter algorithm

Consider system states $x_{0:k} = \{x_0, \ldots, x_k\}$ and measurements $y_{0:k} = \{y_0, \ldots, y_k\}$ up to time k, set of support points (particles) $\{x^i_{1:k}, i = 1, \ldots, N\}$, and set of normalized support weights $\{w^i_{1:k} | \sum_{i=1}^{N} w^i_{1:k} = 1$. Joint posterior density of the system states $x_{0:k}$, given measurements $y_{0:k}$ has particle representation $\{w^i_{1:k}, x^i_{1:k}\}, i = 1, \ldots, N$, and its moments can be approximated based on density approximation

$$p(x_{0:k}|y_{0:k}) \approx \sum_{i=1}^{N} w^i_{1:k}\delta(x_{0:k} - x^i_{1:k}). \tag{4.12}$$

Support points and weights are chosen based on proposal distribution and method described in previous section, i.e. if samples $x^i_{1:k}$ are drawn from the proposal density $q(x_{0:k}|y_{0:k})$ then

$$w^i_{1:k} \propto \frac{p(x^i_{1:k}|y_{0:k})}{q(x^i_{1:k}|y_{0:k})}. \tag{4.13}$$

There exist recursive formulas for updating particles and their respective weights which are derived as follows.

Suppose we have the particle representation of the state posterior at previous time $k-1$, i.e. posterior distribution function $p(x_{0:k-1}|y_{0:k-1})$ is represented by support points $x^i_{1:k-1}$ and respective weights w^i_{k-1}. Now we need to estimate posterior distribution of the state $p(x_{0:k}|y_{0:k})$ at time k by updating set of support points $x^i_{1:k}$ and associated weights $w^i_{1:k}$. Let's assume that proposal density is of the following form

$$q(x_{0:k}|y_{0:k}) = q(x_k|x_{0:k-1}, y_{0:k})q(x_{0:k-1}|y_{0:k-1}). \tag{4.14}$$

This assumption is valid since proposal distribution can be chosen arbitrarily, and the only requirement is that it must have same support as original posterior density. If proposal density factorizes in such a way, the set of support points $x^i_{1:k}$ can be obtained by appending existing points $x^i_{1:k-1}$ drawn from distribution $q(x_{0:k-1}|y_{0:k-1})$ with new x^i_k drawn from distribution $q(x_k|x^i_{1:k-1}, y_{0:k})$, i.e. $x^i_{1:k} = [x^i_{1:k-1}\ x^i_k]$. To derive the weights' update equation consider the posterior distribution $p(x_{0:k}|y_{0:k})$, which can be rewritten as

$$\begin{aligned} p(x_{0:k}|y_{0:k}) &= \frac{p(y_k|x_{0:k}, y_{0:k-1})p(x_{0:k}|y_{0:k-1})}{p(y_k|y_{0:k-1})} \\ &= \frac{p(y_k|x_{0:k}, y_{0:k-1})p(x_k|x_{0:k-1}, y_{0:k-1})p(x_{0:k-1}|y_{0:k-1})}{p(y_k|y_{0:k-1})} \\ &= \frac{p(y_k|x_k)p(x_k|x_{k-1})}{p(y_k|y_{0:k-1})}p(x_{0:k-1}|y_{0:k-1}) \\ &\propto p(y_k|x_k)p(x_k|x_{k-1})p(x_{0:k-1}|y_{0:k-1}) \end{aligned} \tag{4.15}$$

The third equality holds because states are from the Markov process, and depend only on the previous state. By substituting (4.14) and (4.15) into (4.13), update weights can be calculated as

$$w^i_{1:k} \propto \frac{p(y_k|x^i_k)p(x^i_k|x^i_{k-1})p(x^i_{1:k-1}|y_{0:k-1})}{q(x_k|x^i_{1:k-1}, y_{0:k})q(x^i_{1:k-1}|y_{0:k-1})} = w^i_{1:k-1}\frac{p(y_k|x^i_k)p(x^i_k|x^i_{k-1})}{q(x^i_k|x^i_{1:k-1}, y_{0:k})}. \tag{4.16}$$

Moreover, if proposal density only depends on x_{k-1} and y_k then $q(x_k|x_{0:k-1}, y_{0:k}) = q(x_k|x_{k-1}, y_k)$, and only x_{k-1} and y_k are needed to estimate posterior of the state $p(x_k|y_{0:k})$, i.e. weights' update formula is

$$w^i_{1:k} \propto w^i_{1:k-1}\frac{p(y_k|x^i_k)p(x^i_k|x^i_{k-1})}{q(x^i_k|x^i_{k-1}, y_k)}. \tag{4.17}$$

This is useful when only current state is of interest.

The common problem of particle filter algorithm is a degeneracy of particles. Particles' degeneracy occurs when only few of them have weights much larger then majority of other particles. In this case filter estimation is based only on few particles and filter might get "stuck". To resolve this problem particles' re-sampling is used. Re-sampling replaces the particles with low weight and substitute them with new ones that are similar to the particles with high weights.

Most common re-sampling approach is systematic importance sampling (SIS). It re-samples particles when degeneracy is detected. Degeneracy can be detected based on the number of so-called effective particles

$$N_{\text{eff}} = \frac{1}{\sum_{j=1}^{N}(w^{1:k}_j)^2}. \tag{4.18}$$

If N_{eff} falls below some predefined threshold then re-sampling is triggered.

Actual re-sampling is done by taking copies of the existing particles with probabilities proportional to the weights of the particles. This means that new samples are drawn from the distribution

$$p(x|x^i_{1:k}) = \sum_{i=1}^{N} w^i_{1:k-1}\delta(x - x^i_{1:k}). \tag{4.19}$$

[14]

Chapter 5

Filter consistency evaluation

Filter can be considered as consistent if it provides consistent estimates. The main difference between consistency of static and dynamic estimations is that in case of static estimation, estimated parameter is not changing, or changing very slowly, and it is possible to get large sample from its actual distribution, whereas in filtering, estimated state or parameter is changing in time, and for each single estimate it is possible to get only small sample, which size rarely exceeds 1. Moreover, overall consistency of series of estimates provided by a filter is more important than consistency of a single estimate. Thus, instead of considering one estimate and one sample, several consecutive estimates and respective samples (usually of size 1) are considered for filter consistency evaluation.

5.1 Mean Squared Deviation filter consistency

Filter is consistent in mean squared deviation if it provides estimates that are consistent in mean squared deviation. Consider M system states modeled as N-variate random variables $[X_1, \ldots, X_M]$, state estimates $[\bar{x}_1, \ldots, \bar{x}_M]$ and corresponding covari-

ance matrices $[\bar{P}_1, \dots, \bar{P}_M]$ given by the filter, and actual system states $[x_1, \dots, x_M]$ which can be considered as realizations of $[X_1, \dots, X_M]$.
Let H_0 be a null hypothesis stating that estimates are consistent in mean squared deviation. Choose $\epsilon > N$ and define test statistics U as

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

where

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

If H_0 is true then according to (2.4)

$$\Pr\left\{(X_k - \bar{x}_k)^T \bar{P}_k^{-1}(X_k - \bar{x}_k) \leq \epsilon\right\} \geq 1 - \frac{N}{\epsilon}, k = 1, \dots, M. \tag{5.1}$$

Since actual states x_k's are independent samples drawn from distributions of X_k's, U is a sum of M independent Bernoulli random variables with probabilities of success $p_k \geq 1 - \frac{N}{\epsilon}$, i.e. U has Poisson Binomial distribution. Taking into account that $p_k \geq 1 - \frac{N}{\epsilon}$, if l is such that

$$1 - \frac{N}{\epsilon} \geq \frac{l}{M}, \tag{5.2}$$

it is clear that

$$\Pr\{U = l \mid H_0\} \leq \binom{M}{l} \left(1 - \frac{N}{\epsilon}\right)^l \left(\frac{N}{\epsilon}\right)^{M-l}. \tag{5.3}$$

Therefore, for given significance level α, critical region for U can be set as $r = \{0, \dots, L\}$, where L takes maximum value from $\{0, M\}$ for which

$$\sum_{l=0}^{L} \binom{M}{l} \left(1 - \frac{N}{\epsilon}\right)^l \left(\frac{N}{\epsilon}\right)^{M-l} \leq \alpha, \text{ and } 1 - \frac{N}{\epsilon} \leq \frac{L}{M} \tag{5.4}$$

Then $\Pr\{U \in r \mid H_0\} \leq \sum_{l=0}^{L} \binom{M}{l} \left(1 - \frac{N}{\epsilon}\right)^l \left(\frac{N}{\epsilon}\right)^{M-l} \leq \alpha$, and if U falls inside r, H_0 can be rejected at significance level α. Note that hypothesis test does not evaluate consistency of any particular estimate $\tilde{X}_k$, but evaluates "average" consistency of the estimates.

5.2 Filter P-Consistency and P-Equivalence

5.2.1 P-consistency

Filter is p consistent if it provides p consistent estimates. Consider M system states 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 $[X_1, \dots, X_M]$. Let H_0 be a null hypothesis stating that estimates are p consistent. Choose $\epsilon_k > 0 \mid \Pr\left\{(\tilde{X}_k - \bar{x}_k)^T \bar{P}_k^{-1} (\tilde{X}_k - \bar{x}_k) \leq \epsilon_k\right\} = p, k \in \{1, \dots, M\}$ and define test statistics U as

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

where

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

If H_0 is true then according to (2.5)

$$\Pr\left\{(X_k - \bar{x}_k)^T \bar{P}_k^{-1} (X_k - \bar{x}_k) \leq \epsilon_k\right\} \geq \Pr\left\{(\tilde{X}_k - \bar{x}_k)^T \bar{P}_k^{-1} (\tilde{X}_k - \bar{x}_k) \leq \epsilon_k\right\} = p \quad (5.5)$$

Since actual states x_k's are independent realizations of X_k's, U is a sum of M independent Bernoulli random variables with probabilities of success $p_k \geq p$, i.e. it has Poisson binomial distribution. Therefore, for significance level α, critical region for U can be defined as for mean squared deviation consistency test (see Section. 5.1) with $1 - \frac{N}{\epsilon}$ replaced by p. That is r can be set as $r = \{0, \dots, L\}$, where L takes maximum value from $\{0, \dots, M\}$ for which

$$\sum_{l=0}^{L} \binom{M}{l} p^l (1-p)^{M-l} \leq \alpha, \text{ and } p \leq \frac{L}{M} \quad (5.6)$$

Then $\Pr\{U \in r \mid H_0\} \leq \sum_{l=0}^{L} \binom{M}{l} p^l (1-p)^{M-l} \leq \alpha$, and if U falls inside r, H_0 can be rejected at significance level α.

5.2.2 P-equivalence

Filter is p equivalent if it provides p equivalent estimates. Let H_0 be a null hypothesis stating that estimates $[\tilde{X}_1, \ldots, \tilde{X}_M]$ are p equivalent and let test statistic U be defined as previously. If H_0 is true then according to (2.8)

$$\Pr\left\{(X_k - \bar{x}_k)^T \bar{P}_k^{-1}(X_k - \bar{x}_k) \leq \epsilon_k\right\} = \Pr\left\{(\tilde{X}_k - \bar{x}_k)^T \bar{P}_k^{-1}(\tilde{X}_k - \bar{x}_k) \leq \epsilon_k\right\} = p,$$

and U is a sum of M independent Bernoulli random variables with probability of success equal to p, i.e. it has binomial distribution. Therefore, for significance level α, two sided critical region for U can be defined as $r = \{0, \ldots, L_1\} \cup \{L_2, \ldots, M\}$ so that

$$\Pr\{U \in r \mid H_0\} = \sum_{l=0}^{L_1} \binom{M}{l} p^l (1-p)^{M-l} + \sum_{l=L_2}^{M} \binom{M}{l} p^l (1-p)^{M-l} \leq \alpha, \tag{5.7}$$

where α is the significance level of the test.
p-equivalence hypothesis test with two sided critical region for U checks both consistency and information content of the estimates. If U falls into the left part of the critical region, not all the estimates are consistent, if U falls into the right part of the critical region, estimates are consistent but some are uninformative.

5.3 Filter NDS consistency and NDS equivalence

5.3.1 NDS consistency

Filter is NDS consistent if it provides NDS consistent estimates. Consider M system states modeled as N-variate random variables $[X_1, \ldots, X_M]$, state estimates $[\tilde{X}_1, \ldots, \tilde{X}_M]$ with known probability distributions, respective means $[\bar{x}_1, \ldots, \bar{x}_M]$ and covariance matrices $[\bar{P}_1, \ldots, \bar{P}_M]$ provided by the filter, and true states $[x_1, \ldots, x_M]$ which can be considered as realizations of $[X_1, \ldots, X_M]$. $[\tilde{X}_1, \ldots, \tilde{X}_M]$ and $[X_1, \ldots, X_M]$ are assumed to have probability densities with finite number of discontinuity points. Let H_0 be a null hypothesis stating that estimates $[\tilde{X}_1, \ldots, \tilde{X}_M]$ are NDS consistent. Define test statistics U and $\tilde{U}$ as

$$U = \sum_{k=1}^{M} U_k, \ \tilde{U} = \sum_{k=1}^{M} \tilde{U}_k \tag{5.8}$$

where

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

If H_0 is true then according to (2.9)

$$\forall \epsilon_k > 0 \,:\, \Pr\{U_k \leq \epsilon_k\} \geq \Pr\{\tilde{U}_k \leq \epsilon_k\}, \tag{5.9}$$

and according to the Theorem 3.2.

$$\forall \epsilon > 0 \,:\, \Pr\{U \leq \epsilon\} \geq \Pr\{\tilde{U} \leq \epsilon\}. \tag{5.10}$$

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

$$\Pr\{\tilde{U} \in r\} \leq \alpha \tag{5.11}$$

($\Pr\{\tilde{U} \in r\}$ can be calculated based on distributions of $[\tilde{X}_1, \ldots, \tilde{X}_M]$ which are available). Then $\Pr\{U \in r \mid H_0\} \leq \Pr\{\tilde{U} \in r\} \leq \alpha$, and if U falls inside r, hypothesis H_0 can be rejected at significance level α.

5.3.2 NDS equivalence

Filter is NDS equivalent if it provides NDS equivalent estimates. Let H_0 be a null hypothesis stating that estimates $[\tilde{X}_k, \ldots, \tilde{X}_M]$ are NDS equivalent and let test statistics U and $\tilde{U}$ be defined as previously. If H_0 is true then according to (2.16)

$$\forall \epsilon_k > 0 \,:\, \Pr\{U_k \leq \epsilon_k\} = \Pr\{\tilde{U}_k \leq \epsilon_k\} \tag{5.12}$$

and

$$\forall \epsilon > 0 \,:\, \Pr\{U \leq \epsilon\} = \Pr\{\tilde{U} \leq \epsilon\}. \tag{5.13}$$

Therefore, for significance level α, two sided critical region for U can be defined as $r = [0, e_1] \cup [e_2, +\infty]$ so that

$$\Pr\{\tilde{U} \in r\} \leq \alpha \tag{5.14}$$

($\Pr\{\tilde{U} \in r\}$ can be calculated based on distributions of $[\tilde{X}_1, \ldots, \tilde{X}_M]$ which are available). Then $\Pr\{U \in r \mid H_0\} = \Pr\{\tilde{U} \in r\} \leq \alpha$, and if U falls inside r, H_0 can be rejected at significance level α. NDS equivalence test with two sided critical region for U checks both consistency and information content of the estimates. If U falls into the right part of the critical region, not all the estimates are consistent, if U falls into the left part of the critical region, estimates are consistent but some are uninformative.

5.4 NEES and NIS tests for Kalman Filter

Normalized Estimation Error Squared (NEES) and Normalized Innovations Squared (NIS) tests [2] are applied to Kalman Filter. Tests are based on the following filter consistency definition.
Filter is consistent if [2]:

- Estimation errors are acceptable as zero mean, and conform with the corresponding covariances calculated by the filter
- Innovations are acceptable as zero mean, and conform with the corresponding covariances calculated by the filter
- Innovations are acceptable as white

When estimation problem meets linear Gaussian assumptions, Kalman Filter algorithm calculates an exact optimal estimate in the form of Normal RV $\tilde{X}_k$.
Consider state X_k at time t_k and its estimate $\tilde{X}_k$ with corresponding mean $\bar{x}_k$ and covariance matrix $\bar{P}_k$, provided by the Kalman filter, and actual state x_k.
Let H_0 be a hypothesis stating that filter is consistent according to the definition above. Define normalized estimation error squared statistic (NEES) as

$$e_k = (x_k - \bar{x}_k)^T \bar{P}_k^{-1} (x_k - \bar{x}_k). \tag{5.15}$$

If H_0 is true then due to the first consistency criterion e_k is a realization of a $\chi^2_{n_x}$ distribution.
NEES test checks whether e_k can be accepted as a realization of $\chi^2_{n_x}$ using Monte Carlo simulations. For this M independent filter runs are generated to provide sample $e_k^1, \ldots, e_k^M$ calculated as $e_k^i = (x_k - \bar{x}_k^i)^T \bar{P}_k^{i^{-1}} (x_k - \bar{x}_k^i)$. If H_0 is true, then sum

$$\bar{e}_k = \sum_{i=1}^{M} e_k^i \tag{5.16}$$

has a $\chi^2_{M \cdot n_x}$ distribution. Therefore, for given significance level α, critical region for $\bar{e}_k$ can be defined according to $\chi^2_{M \cdot n_x}$ distribution, so that $\Pr\{\bar{e}_k \in r | H_0\} \leq \alpha$. If $\bar{e}_k$ falls inside critical region, H_0 can be rejected by the first criterion of filter consistency.
NIS test is based on a measurement predicted distribution made by a Kalman Filter

algorithm. Under linear Gaussian assumptions, predicted measurement is a multivariate normal RV $\tilde{Y}_k$ with mean $\bar{y}_k = H_k \bar{x}_{k|k-1}$ and covariance $\bar{S}_k$, and innovation $\eta_k = y_k - \tilde{y}_k$ is also a multivariate normal random variable with mean 0 and covariance matrix $\bar{S}_k$ (see section 4.3.1).
Define normalized innovation squared statistic (NIS) as

$$v_k = (y_k - \bar{y}_k)^T \bar{S}_k^{-1} (y_k - \bar{y}_k). \tag{5.17}$$

If H_0 is true then due to the second consistency criterion v_k has a chi-square distribution with n_y degrees of freedom.
NIS test checks whether v_k can be accepted as a realization of $\chi^2_{n_y}$ using Monte Carlo simulations, analogously to NEES test. M independent filter runs are generated to provide sample $v_k^1, \ldots, v_k^M$ calculated as $v_k^i = (y_k - \bar{y}_k^i)^T \bar{S_k^i}^{-1} (y_k - \bar{y}_k^i)$. If H_0 is true then sum

$$\bar{v}_k = \sum_{M}^{i=1} v_k^i \tag{5.18}$$

has a $\chi^2_{M \cdot n_y}$ distribution. Therefore, for given significance level α, critical region for $\bar{v}_k$ can be defined according to $\chi^2_{M \cdot n_y}$ distribution, so that $\Pr\{\bar{v}_k \in r | H_0\} \leq \alpha$. If $\bar{v}_k$ falls inside critical region r, H_0 can be rejected by the second criterion of filter consistency.
According to the third consistency criterion, innovations of the filter must be white. Whiteness of the innovations can be checked by using innovations' autocorrelation statistic

$$\bar{\rho}(k, l) = \sum_{i=1}^{M} \eta_k^{i\,T} \eta_l^i \left[\sum_{i=1}^{M} \eta_k^{i\,T} \eta_k^i \sum_{i=1}^{M} \eta_l^{iT} \eta_l^i \right]^{-\frac{1}{2}},$$

where k and l indicate different time moments.
For large enough M, $\bar{\rho}(k, l)$ can be approximated as normal, and if H_0 is true and innovations are zero-mean and white then mean of $\bar{\rho}(k, l)$ is 0 and its variance is $\frac{1}{M}$ [2]. Therefore, for given significance level α, critical region for $\bar{\rho}(k, l)$ can be set based on normal distribution with mean 0 and variance $\frac{1}{M}$.
Considered Monte-Carlo based NEES and NIS tests are off-line tests, for which several filter runs must be generated in order to increase the power of hypothesis test. In principle only one simulation (run) can be used, but in this case acceptance interval will be relatively large to reject null hypothesis properly. Additionally, NEES test cannot be used on-line since it requires actual system state, whereas NIS test can be used on-line since it requires only actual measurements which are available.
There are two ways to use NIS test on-line. First one is to use only single filter run, i.e. when M is equal to 1. Alternatively, due to ergodicity of innovations' sequence [2], $M-$run average NIS can be substituted by time average taken over last L time moments. Time average NIS can be calculated as

$$\bar{v}_k = \sum_{i=k-L+1}^{L} v_k. \tag{5.19}$$

Critical region for time average NIS can be set according to $\chi^2_{L \cdot n_y}$ distribution.
For whiteness test, ensemble average autocorrelation of innovations can be substituted by time-average autocorrelation of innovations that are l steps apart

$$\bar{\rho}(l) = \sum_{i=k-L+1}^{k} \eta_i^T \eta_{i+l} \left[\sum_{i=k-L+1}^{k} \eta_i^T \eta_i \sum_{i=k-L+1}^{k} \eta_{i+l}^T \eta_{i+l} \right]^{-\frac{1}{2}}.$$

Critical region for time average autocorrelation $\bar{\rho}(l)$ can be defined based on normal distribution with mean 0 and variance $\frac{1}{L}$ [2].

5.5 Consistency tests for suboptimal filters

Kalman Filter algorithm provides exact optimal solution of the estimation problem, but it is applicable only when linear Gaussian assumptions hold, which is not always the case in practical applications. That is why non-linear systems are approximated as linear Gaussian systems and Kalman Filter applied to them, or suboptimal filters are used instead, e.g. Extended Kalman Filter or Linear Regression Kalman Filter. Such filters estimate state of the system by Gaussian random variables $\tilde{X}_k$, however, its actual distribution is not necessarily Gaussian but might have arbitrary distribution, and assumption about normality of the state estimate might be wrong. In this case hypothesis tests for p-consistency and p-equivalence or NDS consistency and NDS equivalence should be used instead of NEES and NIS tests.
In order to check consistency of the particle filter, its state and measurement estimates can be approximated as Gaussian and NEES/NIS consistency checks can be used [16]. If Gaussian assumptions are not applicable then p-consistency/equivalence or NDS consistency/equivalence tests can be used. Unlike EKF and LRKF, which provide Gaussian estimates, particle filter estimates system states by arbitrary distributions presented in the form of a set of weighted particles. Thus, consistency test statistics cannot be calculated analytically, but it is still possible to accurately approximate them based on the set of weighted particles.
In suboptimal filters, approximation errors are compensated by noises with additional uncertainties. If additional uncertainties are not large enough, filter might become inconsistent, on the other side, if they are excessively large, filter becomes less informative. This in turn might affect accuracy of the estimation. If filter provides inconsistent estimates, it becomes less responsive to the new measurements. If filter provides non-informative estimates, it becomes more responsive to the new measurements and as a side affect to the measurement noise. Consistency tests can be used in order to preserve the optimal trade-off between consistency and information

content of the estimates, and as a result achieve optimal estimates' accuracy.

5.6 Consistency of predicted measurement

Every filter in the course of system state estimation predicts a future measurement. Measurement prediction is based on state prediction and measurement model of the system. Additionally, filter receives an actual measurement which can be regarded as a realization from predicted measurement distribution. Measurement estimate provided by the filter must be consistent and informative as well as state estimate given by the filter. Intuitively its is clear that consistency of the measurement estimate is highly correlated with consistency of the state estimate and might be a good indicator of abnormal behavior of a filter.

Let RV $\tilde{Y}_k$ be a predicted measurement, RV Y_k be a true probabilistic model of a measurement and y_k be an actual measurement that can be considered as a realization of Y_k at time t_k.

In Kalman filter as well as in its non-linear counterparts, predicted measurement $\tilde{Y}_k$ is a normal random variable with mean $\bar{y}_k = H_k \bar{x}_{k|k-1}$ and covariance $\bar{S}_k = H_k \bar{P}_{k|k-1} H_k^T + R_k$, where $\bar{x}_{k|k-1}$ and $\bar{P}_{k|k-1}$ are mean and covariance of the predicted system state, H_k is a state transition function (or its linear approximation), and R_k is the covariance of measurement noise.

If linear Gaussian assumptions hold for the system and Kalman Filter is used, NIS test is used for checking consistency of predicted measurement. If system is non-linear and suboptimal filter is used, estimate $\tilde{Y}_k$ can be checked with p-consistency, p-equivalence, NDS consistency or NDS equivalence tests.

It is not proven that inconsistent measurement estimation results in inconsistent state estimation as well as it is not proven that consistent measurement estimation results in consistent state estimation. However, inconsistent predictive measurement provide good indication of filter abnormal behavior. The source of such an anomalies might be caused by modeling or approximation errors in state transition model or measurement model or both.

Another important advantage of the predictive measurement consistency testing is that it does not require any information about 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 uncertainties of process and measurement noises.

Chapter 6

Practical applications

One application area of filtering is geopositioning. In this chapter some of the applications of consistency testing in positioning systems are presented.
In geopositioning system, state being estimated is a geographical position of a user, e.g. Latitude, Longitude (optionally Altitude), and vector of velocity. In such a system constant state/velocity model or models including velocity of a user derived from sensor measurements are usually used as a state transition model. Pseudoranges and delta pseudoranges between user device and radio beacons (GPS, Cell, WIFI) or position and velocity derived from them are used as measurements of the system.
In the following examples constant velocity (CV) model is used as a state transition model of the system, and user positions calculated based on wireless signals are used as measurements of the state.

6.1 System model

Let x_k be a state of the system at time t_k, y_k is a measurement of the state (or observation) at time t_k, and $\triangle t_k = t_k - t_{k-1}$ is a time difference between consecutive time moments. In constant velocity model, state evolves according to the state transition equation

$$x_k = Fx_{k-1} + w_{k-1}, \tag{6.1}$$

where

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

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

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

where

$$Q_c = \begin{bmatrix} \sigma^2_{\mathrm{Lat}} & 0 \\ 0 & \sigma^2_{\mathrm{Lat}} \end{bmatrix}$$

is a so-called diffusion of Brownian motion process. [1]
In our system single WIFI-based positioning fixes are used as a measurements of the state, thereby measurement equation is of the form

$$y_k = Hx_k + v_k \tag{6.2}$$

where $H = [\mathrm{I}_{2\times2}\ \mathrm{O}_{2\times2}]$ and v_k is a zero mean, white, Gaussian measurement noise with covariance matrix R. Process and measurement noises are independent.

6.2 Linear Gaussian System (simulated data)

Consider constant velocity system where actual measurement and process noises are zero mean and Gaussian. Covariance matrices of process noise and measurement noise are defined respectively as

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

where $Q_c = q^2 \cdot \mathrm{I}_{2\times2}$ and $\triangle t = 1$, and $R = r^2 \cdot \mathrm{I}_{2\times2}$. Here simple assumptions about measurement noise matrix and diffusion matrix are used, i.e. Latitudinal and Longitudinal errors are uncorrelated and have equal variances.
To investigate the impact of process noise's magnitude on the consistency of the filter, system states and measurements are generated with parameters $q = 1$ and $r = 1$ and then estimated with different values of parameter q.
In order to check consistency of the estimates, NEES test based on 10 independent filter runs, significance level $\alpha = 0.1$, and χ^2_{40} distribution is used. In the Fig. 6.1 ratio of estimates accepted as consistent is plotted against value of parameter q.

As seen from the figure, when assumed process noise matches with actual process noise, i.e. when $q = 1$, 90 % of the estimates are accepted as consistent, which agrees with chosen significance level $\alpha = 0.1$. On the other side, when assumed process noise

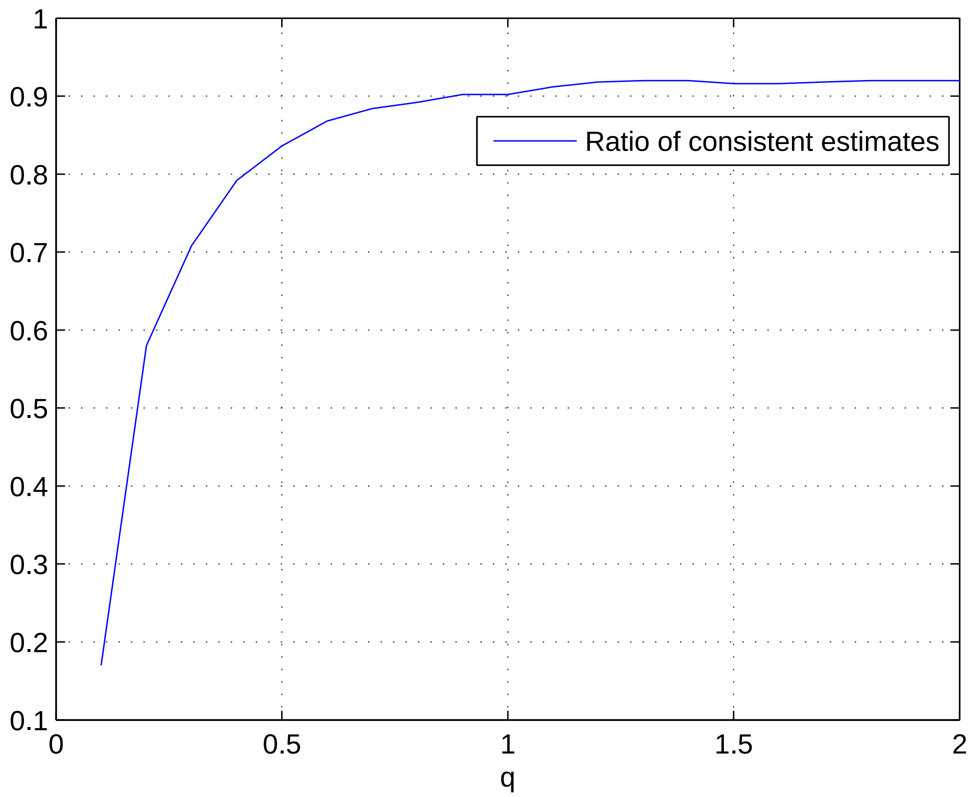


Figure 6.1: Ratio of consistent estimates for different process noises defined by parameter $q \in \{0.1, 0.2, \ldots, 2\}$.

has smaller magnitude than actual process noise, i.e when $q < 1$, ratio of consistent estimates decreases and filter becomes inconsistent.

6.3 Linear system with non-Gaussian noises (real data)

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

Our WIFI positioning system is based on fingerprinting method described in [6] (probabilistic framework and Gaussian likelihood calculation with constant parameters are used). Due to various factors affecting WIFI signal propagation, WIFI positioning

errors are not exactly Gaussian and outliers might occur.
In order to make filter consistent, matrices $Q_c = q^2 \cdot \mathrm{I}_{2\times2}$ and $R = r^2 \cdot \mathrm{I}_{2\times2}$ must be appropriately adapted so that magnitudes of process and measurement noises are sufficiently large to compensate for motion modeling and measurement errors. In order to investigate an impact of measurement and process noises on consistency of the estimation, user positions are estimated with different values of parameters q and r, and consistency of the estimates is evaluated by MSD, p, and NDS consistency tests.
Consistency tests are carried out every fifth time moment, and test statistics are calculated based on $M = 5$ last estimates.
For NDS consistency, test statistic U is calculated as sum of normalized squared deviations of the estimates. According to 5.3.1, critical region for U is determined based on χ^2_{20} $(M \cdot n_x = 20)$ and significance level α. Here $\alpha = 0.1$ is used and critical region is set to $[28.4, +\infty)$. In the Table 6.1 ratio of NDS consistent estimates for different values of parameters q and r is presented.
p consistency is tested for $p = 0.68$. According to 5.2.1, test statistic U is calculated as a number of estimates that fall within 68% concentration ellipse, and critical region for U is determined based on $B(M, 0.68)$ and significance level α. Critical region is set to $\{0, 1\}$, which correponds to the significance level $\alpha = 0.1905$. In the Table 6.2 ratio of 0.68 consistent estimates for different values of parameters q and r is presented.
Mean squared deviation of the estimates is tested for $\epsilon = 8$. According to 5.1, test statistic U is calculated as a number of estimates which normalized squared deviation is less than ϵ, and critical region is determined based on $B(M, 1 - \frac{N}{\epsilon})$ and significance level α. Critical region for U is set to $\{0\}$, which corresponds to the significance level $\alpha = 0.1875$. In the Table 6.3 ratio of mean squared deviation consistent estimates for different values of parameters q and r is presented.
Filter can be considered as consistent in general if ratio of its consistent estimates is larger then $1 - \alpha$. This is because under hypothesis H_0 (filter is consistent) test statistic must fall inside critical region with probability α, i.e. actually consistent

Table 6.1: Ratio of NDS consistent estimates

r\\q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1	0	0.012	0.037	0.043	0.043	0.049	0.062	0.074	0.080	0.080
2	0.074	0.192	0.242	0.304	0.372	0.385	0.416	0.453	0.472	0.490
3	0.173	0.409	0.540	0.621	0.652	0.689	0.714	0.739	0.764	0.776
4	0.304	0.602	0.751	0.850	0.869	0.894	0.925	0.925	0.931	0.931
5	0.403	0.732	0.832	0.937	0.956	0.956	0.962	0.962	0.962	0.962
6	0.453	0.807	0.900	0.962	0.968	0.968	0.975	0.981	0.981	0.987
7	0.484	0.844	0.937	0.962	0.975	0.975	0.981	0.981	0.987	0.987
8	0.546	0.875	0.937	0.975	0.975	0.981	0.981	0.987	0.987	0.993
9	0.565	0.906	0.950	0.981	0.987	0.987	0.993	0.993	0.993	0.993
10	0.596	0.906	0.956	0.981	0.987	0.987	0.993	0.993	0.993	0.993

Table 6.2: Ratio of p consistent estimates

r\\q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1	0.043	0.087	0.112	0.124	0.168	0.174	0.174	0.186	0.193	0.193
2	0.106	0.273	0.354	0.422	0.460	0.472	0.509	0.528	0.553	0.578
3	0.248	0.491	0.621	0.702	0.720	0.752	0.770	0.776	0.795	0.814
4	0.379	0.646	0.795	0.857	0.882	0.894	0.913	0.919	0.925	0.938
5	0.472	0.770	0.882	0.925	0.950	0.957	0.963	0.969	0.975	0.975
6	0.516	0.826	0.938	0.963	0.963	0.963	0.975	0.981	0.988	0.988
7	0.571	0.863	0.957	0.969	0.969	0.969	0.988	0.994	0.994	0.994
8	0.602	0.888	0.963	0.975	0.981	0.994	0.994	0.994	0.994	0.994
9	0.621	0.901	0.963	0.981	0.981	0.994	0.994	0.994	0.994	0.994
10	0.634	0.907	0.969	0.981	0.988	0.994	0.994	0.994	0.994	0.994

Table 6.3: Ratio of MSD consistent estimates

r\\q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1	0.143	0.298	0.348	0.404	0.435	0.429	0.453	0.478	0.497	0.516
2	0.404	0.615	0.702	0.714	0.758	0.820	0.839	0.845	0.851	0.870
3	0.596	0.820	0.901	0.913	0.944	0.957	0.963	0.963	0.975	0.975
4	0.720	0.901	0.969	0.975	0.975	0.975	0.975	0.981	0.981	0.988
5	0.776	0.950	0.981	0.994	0.994	0.994	0.994	0.994	0.994	0.994
6	0.820	0.969	0.988	0.994	0.994	0.994	0.994	0.994	0.994	0.994
7	0.832	0.975	0.988	0.994	0.994	0.994	0.994	1.000	1.000	1.000
8	0.839	0.975	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
9	0.857	0.981	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0.863	0.981	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

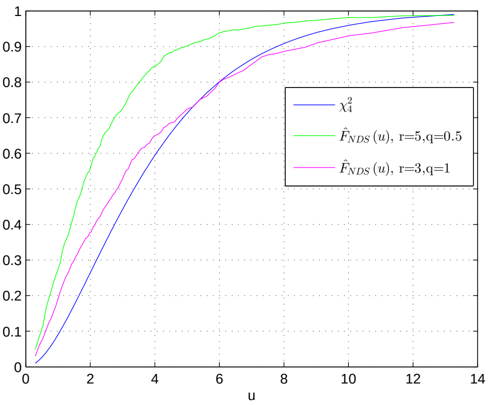
filter might be evaluated as inconsistent with probability α. As seen from the tests' results, filter consistency is mainly influenced by magnitude of the measurement noise, whereas process noise does not have noticeable impact unless it is extremely small. It means that consistency of the estimates largely depends on the correct modeling of the measurement noise.

Compared to NDS and P consistency tests, MSD consistency is more "forgiving", meaning that normalized squared estimation error must be very large in order to exceed threshold prescribed by Chebyshev's inequality, and reject MSD consistency, i.e. MSD test is not effective in filter fault detection, or detection of system mis-modeling. In contrast, NDS and p consistent tests use critical regions prescribed by probability laws (particularly Gaussian) for normalized squared deviation. Such critical regions are much narrower and allow more efficient filter inconsistency and failure detection. This is justified by plots on figures 6.2a, 6.2b. Additionally, NDS, and P consistency tests are highly correlated for $p = 0.68$. This means that 0.68 consistent filter is most probably NDS consistent. Correlation is expected to be even higher for larger values of p, e.g. 0.95.

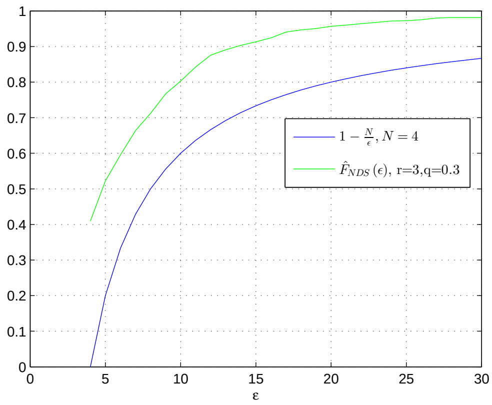
In order to show that generally consistent filters actually provide estimates that are consistent according to the definition, let's consider empirical cumulative distribution of normalized squared deviation (estimation error) of three filters with noises defined by parameters $r = 5, q = 0.5$; $r = 3, q = 1$; and $r = 3, q = 0.3$ respectively. First filter is NDS consistent, second filter is p consistent, and third filter is MSD consistent according to consistency tests.

Empirical cumulative distribution function of normalized squared estimation error for first and second filters is plotted in the Fig. 6.2a along with cumulative distribution of χ^2_4 random variable, which is the theoretical distribution of normalized squared error declared by the filter (since Kalman Filter estimate state by Gaussian distribution, and $n_x = 4$). As seen from the figure, empirical CDF of normalized squared deviation of the estimates provided by the NDS consistent filter is larger then CDF of χ^2_4, this complies with the definition of NDS consistency. For p consistent filter, $\hat{F}_{NDS}(F^{-1}_{\chi^2_4}(0.68)) \geq F_{\chi^2_4}(F^{-1}_{\chi^2_4}(0.68))$, which agrees with the definition of p consistency. Empirical distribution of normalized squared estimation error for third filter along with its theoretical lower bound $1 - \frac{N}{\epsilon}$, provided by Chebyshev's inequality, is plotted in the Fig. 6.2b. As seen from the figure, $\hat{F}_{NDS}$ is larger then $1 - \frac{N}{\epsilon}$, which agrees with MSD consistency criteria.

In the Fig. 6.3 and Fig. 6.4, exemplary user track, estimated by NDS consistent and NDS inconsistent filters respectively, is presented. For NDS consistent filter, more than 50% of the estimates are inside the 50% concentration ellipse, whereas for inconsistent filter fewer than 50% of the estimates are within corresponding ellipses. As also seen from the Fig. 6.4, inconsistent filter is quite responsive to the new measurements, which contradicts with assumption about lagging of inconsistent filter, this is because measurement noise has small uncertainty (which actually cause filter



(a) Cumulative distribution of theoretical and actual normalized squared estimation error.



(b) Empirical cumulative distribution normalized squared estimation error with its lower bound according to Chebyshev's inequality.

Figure 6.2: Empirical cumulative distribution function of normalized squared estimation error.

inconsistency). This means that if measurements uncertainty is underestimated, but measurements are accurate, filter will not get stuck, and will provide accurate point estimates, however, it will not provide realistic information about estimation error. On the other hand, even though consistency of the estimates is not sensitive to moderately underestimated process noises, extremely small process noise magnitudes might cause abnormal behavior of a filter and degradation of both consistency and accuracy of the estimates.

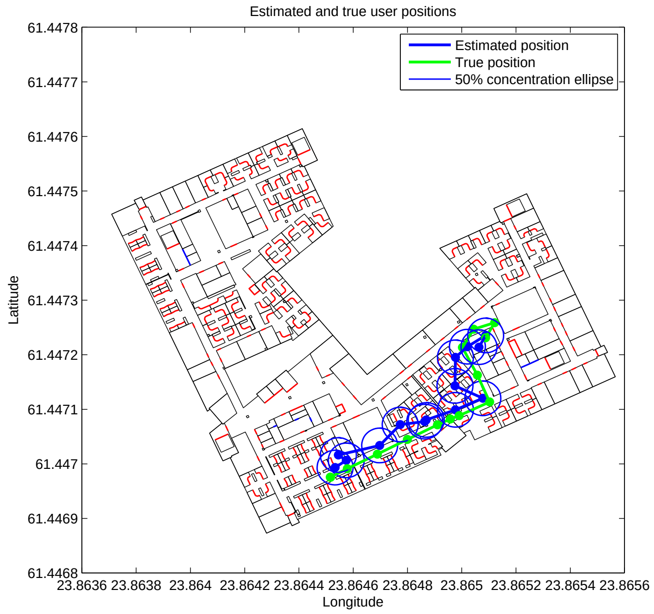


Figure 6.3: Exemplary user track estimated by NDS consistent filter with noises' parameters $q = 0.5$ and $r = 5$.

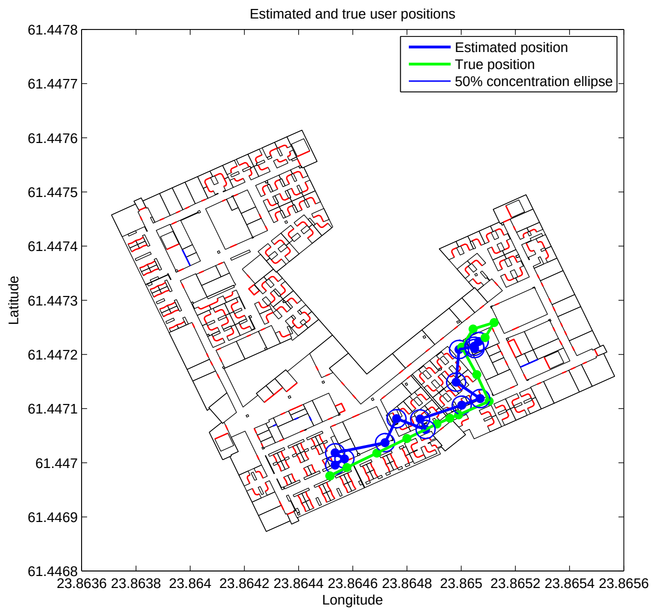


Figure 6.4: Exemplary user track estimated by NDS inconsistent filter with noises' parameters $q = 1$ and $r = 2$.

Chapter 7

Conclusions

The aim of the current work was to study the concept of estimation consistency, and methods of its evaluation.

In Chapter 2. estimation consistency was defined in several ways. Mean Squared Deviation consistency and p consistency are based on the consistency definitions already presented in the literature, and NDS consistency is defined by the author. Basic idea that lies in the core of all three definitions is that normalized squared estimation error must be within the limits implied by the probability distribution associated with the estimate or covariance matrix of the estimate. In case estimate is specified by point and covariance, estimation error must conform with the upper bound provided by a Chebyshev theorem. If estimate is defined by a probability distribution then actual value of estimated parameter must conform with concentration ellipses declared by the distribution (at least one). Even though this ideas are quite intuitive, they were defined in a formal way, exemplified with standard distributions, and used as a basis for the thesis work.

In this thesis, statistical hypothesis framework is used for consistency evaluation. Hypothesis tests are derived for evaluation of consistency of a static estimation, and exemplified with the simulated Gaussian data sets. Static parameter estimation consistency is further extended by definitions of filter consistency, and hypothesis tests are derived for filter consistency evaluation.

Filter consistency evaluation was illustrated by using simulated and real data sets. For linear Gaussian systems NEES test is used. For practical systems with linear model and non-Gaussian noises, NDS-consistency, P-consistency and MSD consistency tests are used.

Experiments revealed that consistency of the estimation is mostly influenced by measurement noise, whereas it is less sensitive to underestimated process noise. However, it was also shown that when assumed process noise is extremely small compared to actual process noise of the system, consistency degrades a lot.

Compared to NEES and NIS tests, which are the most popular techniques for consistency evaluation, proposed MSD, p, and NDS consistency tests are applicable to larger class of filters. This is because in these tests only few assumptions are made about actual and estimated distributions of the parameter. However, it should be mentioned that for non-linear Kalman Filters (EKF, LRKF) NDS consistency test is exactly the same as NEES test with the only difference that NDS test does not make any assumption about actual distribution of estimated parameter. Among MSD, p, and NDS consistency tests, p and NDS tests are preferable for filter failure/mis-modeling detection over MSD consistency test.

Proposed consistency definitions conform with the mathematical and intuitive understanding of estimate uncertainty. If estimate meets requirements of consistency definition then it does not provide over-optimistic information about estimation error i.e. does not underestimate error magnitude. Proposed hypothesis tests are able to check whether estimates meet these requirements or not. It was shown that consistency tests are useful for off-line tuning of the system noises' parameters. Methods for on-line adjustment of system noises are covered only theoretically, and practical studies are left for the future research.

Bibliography

[1] Simo Ali-Löytty, Niilo Sirola, and Robert Piché. Consistency of three Kalman filter extensions in hybrid navigation. *Proceedings of the European Navigation Conference GNSS*, 2005.

[2] Yaakov Bar-Shalom and Xiao-Rong Li. *Estimation and Tracking: Principles, Techniques, and Software*. Artech House, 1998.

[3] Xinjia Chen. A new generalization of Chebyshev's inequality for random vectors. 2011. URL `http://arxiv.org/abs/0707.0805`.

[4] Kalman Rudolph Emil. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering*, 82:35–45, 1960.

[5] Richard G. Gibbs. New Kalman filter and smoother consistency tests. *Automatica*, 49(10):3141–3144, 2013.

[6] Ville Honkavirta, Tommi Perälä, Simo Ali-Löytty, and Robert Piché. A comparative survey of WLAN location fingerprinting methods. In *Proceedings of the 6th Workshop on Positioning, Navigation and Communication 2009 (WPNC'09)*, pages 243–251, March 2009. URL `http://math.tut.fi/posgroup/honkavirta_et_al_wpnc09a.pdf`.

[7] Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory.* Mathematics in Science and Engineering, Vol. 64. Academic Press. 378 p., 1970.

[8] Simon Julier and Jeffrey K. Uhlmann. A general method for approximating nonlinear transformations of probability distributions. Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford, November 1996.

[9] Tine Lefebvre, Herman Bruyninckx, and Joris De Schutter. Comment on "A new method for the nonlinear transformation of means and covariances in filters and estimators" [with authors' reply]. *IEEE Trans. Automat. Contr.*, 47(8):1406–1409, 2002. URL `http://dblp.uni-trier.de/db/journals/tac/tac47.html`.

[10] Tine Lefebvre, Herman Bruyninckx, and Joris de Schutter. Kalman filters for nonlinear systems: a comparison of performance. *International Journal of Control*, 77:639–653, 2004.

[11] Robb J. Muirhead. *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, Inc., 1982.

[12] Henri Nurminen, Anssi Ristimäki, Simo Ali-Löytty, and Robert Piché. Particle filter and smoother for indoor localization. In *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN2013)*, pages 137–146, Montbéliard-Belfort, France, 28-31 October 2013. URL `http://URN.fi/URN:NBN:fi:tty-201403051121`.

[13] Robert Piché. *Stochastic Processes. Lecture notes.* August 2012. URL `http://URN.fi/URN:NBN:fi:tty-201012021377`.

[14] Branko Ristic, Sanjeev Arulampalam, and Neil Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications.* Artech House, 2004.

[15] Murray Rosenblat. Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23:470–472, 1952.

[16] Maria Scalzo, Gregory Horvath, Eric Jones, Adnan Bubalo, Mark Alford, Ruixin Niu, and Pramod K. Varshney. Adaptive filtering for single target tracking. *Proceedings of the SPIE: Defense & Security Symposium*, 4336, 2009.

[17] Tor Steinar Shei. A finite-difference method for linearisation in nonlinear estimation algorithms. *Modeling, Identification and Control*, 19:141–152, 1998.

[18] F. Van der Heijden. Consistency checks for particle filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):140–145, Jan 2006. ISSN 0162-8828. doi: 10.1109/TPAMI.2006.5.

[19] Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye. *Probability and Statistics for engineers and scientists.* Prentice Hall, 2007.