# Reconstruction of Joint Covariances in Networked Linear Systems

Marc Reinhardt, Benjamin Noack, and Uwe D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory (ISAS),
Karlsruhe Institute of Technology (KIT), Germany.
Email: marc.reinhardt@ieee.org, benjamin.noack@ieee.org, uwe.hanebeck@ieee.org

***Abstract*—We propose a sample representation of estimation errors that is utilized to reconstruct the joint covariance in distributed estimation systems. The key idea is to sample uncorrelated and fully correlated noise according to different techniques at local estimators without knowledge about the processing of other nodes in the network. In this way, the correlation between estimates is inherently linked to the representation of the corresponding sample sets. We discuss the noise processing, derive key attributes, and evaluate the precision of the covariance estimates.**

***Keywords*—Covariance reconstruction, Distributed architectures, Kalman filtering**

## I. Introduction

Over the last few decades, distributed estimation has been a major research topic. Several algorithms that consider different types of communication and models have been developed and applied to various fields [1].

Even for linear models, distributed estimation is still considered to be an active research area. While the Kalman filter (KF) provides an – in several respects – optimal solution for the central linear estimation problem, the main challenges in distributed estimation arise from limited knowledge about the processing at remote nodes. In particular, correlations between estimates that emerge from the same (common) process model [2] and from past data exchanges [3] are challenging to track.

Techniques to calculate or guess these correlations under the unrealistic assumption of global model knowledge (every node knows the measurement models and communication paths of all other nodes) have been derived [4] and optimal methods to fuse estimates based on known joint covariances have been proposed for two [2] and more than two nodes [5], [6].

Indeed, the emergence of the research topic *fusion under unknown correlations* [7], [8] emphasizes the need and the challenge to calculate cross-covariances. The main idea of algorithms from this field is to minimize the worst-case error, i.e., to bound all possible outcomes of fused covariances by an artificial covariance of minimal size.

Largely independent from classical KF methods, sample-based approaches have gained attention in (centralized) nonlinear estimation. Here, weakly nonlinear filters that often rely on some type of Gaussian assumption such as Unscented Kalman filtering [9] are distinguished from general density estimators that are referred to as (sequential) Monte-Carlo approaches or particle filters [10].

A connection between linear filtering and sample-based estimation has been established by the Ensemble KF [11], which aims at reducing the computational complexity of processing high dimensional covariances. One of the main challenges that arises in the ensemble KF and in sample-based linear filtering in general, is to reconstruct covariances based on sample sets.

A variety of methods and heuristics to estimate covariances or inverses of covariances – in particular of high dimension – are available in literature. First and foremost, it is worth mentioning the James-Stein-Estimator [12], which is a biased estimator that dominates usual least squares or ML approaches. The underlying idea is called shrinkage and encompasses approaches that (often linearly) combine the estimate with some additional information. More recently, regression-based methods have been proposed that process the samples iteratively in order to obtain coefficients of a triangular matrix that is used in a modified Cholesky factorization to reconstruct the joint covariance [13] [14]. Especially for sparse matrices, optimization algorithms for $L_1$-norms, so called lasso methods, have been developed [15]. A comprehensive overview over different covariance estimation methods is given in [16]. Quality attributes, in particular bounds, for specific types of estimators and covariances are given in [17].

In this paper, we propose to use sampling in linear distributed estimation. We locally calculate sample sets to represent error terms and demonstrate how the joint covariance is reconstructed from these quantities. The main novelties are:

- sample-based error representation in distributed estimation,
- explicit modeling of uncorrelated and fully correlated noise with sample sets,
- reconstruction of joint covariances based on locally obtained data.

The paper is structured as follows. We motivate and illustrate the challenges that arise in the estimation of joint covariances in Sec. II. In Sec. III, the main techniques and theorems for distributed sample-based covariance estimation are derived in a general context. Sec. IV is concerned with the integration of estimation quantities into the sample set representation, before we finally evaluate and discuss sampling techniques in Sec. V.

## II. Problem Formulation

We consider connected nodes $s \in \mathcal{S} = \{1, \dots, S\}$ with local computation and storage capabilities that collectively estimate the state of a common phenomenon. The local processing of nodes is assumed to be unknown to other nodes. The exchange of information follows an arbitrary pattern.

The state evolution model and the node-specific measurement model are known to the nodes and are given by the linear relations

$$\underline{\boldsymbol{x}}_{k+1} = \mathbf{A}\underline{\boldsymbol{x}}_k + \mathbf{B}\underline{u}_k + \underline{\boldsymbol{w}}_k \text{ and } \underline{z}_k^s = \mathbf{H}_k^s\underline{\boldsymbol{x}}_k + \underline{\boldsymbol{v}}_k^s \,. \quad (1)$$

The noise terms are assumed to be independent of each other and are distributed according to arbitrary probability density

functions with zero mean and covariances $E\{\underline{\boldsymbol{w}}_k(\underline{\boldsymbol{w}}_k)^\top\} = \mathbf{Q}$ and $E\{\underline{\boldsymbol{v}}_k^s(\underline{\boldsymbol{v}}_k^s)^\top\} = \mathbf{R}_k^s$ respectively. The nodes derive local estimates $\underline{\hat{x}}_k^s$ by linearly combining and predicting measurements as well as by exchanging information. Due to the common process noise $\underline{\boldsymbol{w}}_k$ that is modeled in all nodes and due to past data exchanges, correlations arise between estimates.

While for linear filters, the calculation of local covariances is feasible in closed form, the evolution of correlations depends on matrix transformations and gains at remote nodes. When these operations are known to all nodes in the network, the problem is to optimize filter and fusion gains under global model knowledge. However, in realistic environments, the quality of measurements is typically time-variant and state-dependent, and nodes have only knowledge about their direct neighborhood. Hence, global model knowledge is at best an approximation of the true system and often inapplicable.

The problem considered in this paper is to estimate correlations based on locally available information only. Let the estimates $\underline{\hat{x}}_k^s$, $s \in \mathcal{S}$ be given, then we aim to reconstruct the joint covariance $\mathbf{P} = E\{(\underline{\hat{x}} - \mathbf{1}\underline{\boldsymbol{x}})(\underline{\hat{x}} - \mathbf{1}\underline{\boldsymbol{x}})^\top\}$ , with $\underline{\hat{x}} = ((\underline{\hat{x}}^1)^\top, \ldots, (\underline{\hat{x}}^S)^\top)^\top$ and $\mathbf{1} = (\mathbf{I}, \ldots, \mathbf{I})^\top$. Note that the joint covariance reflects the local covariances of the estimates and their correlations. With this information in hand, estimates can be optimally processed [2], [5].

It is worth pointing out that the results derived in this paper are not confined to spatially distributed sensor networks but can be utilized for parallel implementations of estimation algorithms or for networks with specific communication patterns as well.

## III. Distributed Covariance Estimation

First, we introduce the concept of distributed sample-based estimation of joint covariances. For this purpose, we examine an arbitrary estimate $\underline{\hat{x}}$ of a linear system depicted in Sec. II. The error process of the estimate $\underline{\hat{x}}$ is given by a linear combination of noise terms[1]. Consider for example the filter step that consists of a linear combination of estimate and measurement according to

$$\underline{\hat{x}}^+ = \mathbf{L}\underline{\hat{x}} + \mathbf{K}\underline{z} \ .$$

As the measurement $\underline{z}$ is an instance of the random process depicted in (1), we know that for unbiased estimators $\mathbf{L} = (\mathbf{I} - \mathbf{K}\mathbf{H})$ holds. Assume that $\underline{\hat{e}}$ is a linear combination of noise terms. The error process of the filtered estimate is given by

$$\underline{\hat{e}}^+ = \underline{\hat{x}}^+ - \underline{\boldsymbol{x}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\underline{\hat{e}} + \mathbf{K}\underline{\boldsymbol{v}} \ ,$$

which, in turn, is then again a linear combination of noise terms. The same argument holds true for other linear operations such as prediction and fusion so that we consider the error process of estimates as linear combinations of noise terms. In the following, these noise terms are denoted as $\underline{\boldsymbol{\psi}}$.

In linear estimation theory, estimates are usually given as a tuple of estimate and covariance, where the covariance describes the attributes of the error process. In the remainder of this section, we augment this two-quantity approach by a sample set that specifies an alternative way to describe the error process. Note that, at least temporarily, we regard sample sets as additional information. In the next section, we propose methods to integrate the estimate and covariance in the sample representation.

[1]This is not to be confused with the error process of the underlying system.

Remember that the error process is given as the linear combination of noise terms. We collect the indices of all noise terms in $\mathcal{T}$ and those by which the node $s$ is affected in $\mathcal{T}^s \subset \mathcal{T}$. Let $\mathbf{B}_t^s$ denote the linear transformation of the noise term $\underline{\boldsymbol{\psi}}_t$ at node $s$. Then the error process of estimate $\underline{\hat{x}}^s$ is given by $\underline{\boldsymbol{e}}^s = \sum_{t\in\mathcal{T}^s} \mathbf{B}_t^s \underline{\boldsymbol{\psi}}_t^s$. Assumed that we are capable of generating samples from noise distributions, it makes sense to specify a *sample characterization* of an error process.

***Definition 1*** *A sample $\underline{\xi}_i^s$* ***characterizes*** *the error process of an estimate $\underline{\hat{e}}^s \sim \sum_{t\in\mathcal{T}^s} \mathbf{B}_t^s \underline{\boldsymbol{\psi}}_t^s = \underline{\boldsymbol{e}}^s$ when $\underline{\xi}_i^s \sim \underline{\boldsymbol{e}}^s + \underline{b}^s$ with a non-stochastic bias $\underline{b}^s$.*

In the linear system from Sec. II, the noise terms are independent, and therefore, for recursive estimators, the samples can be processed recursively as well. To this end, it is sufficient to apply the same linear transformations to the noise samples to which the error process is subject to and to generate new noise independently of already sampled one.

In order to represent covariances, we extend Def. 1 to sample sets.

***Definition 2*** *A sample set $\boldsymbol{\Xi}^s = [\underline{\xi}_1^s, \ldots, \underline{\xi}_n^s]$ with $\bar{\underline{\xi}}^s = \frac{1}{n}\sum_{i=1}^n \underline{\xi}_i^s$ (with affiliated sampling policy)* ***characterizes*** *the error process of an estimate $(\underline{\hat{x}}^s, \mathbf{P}^s)$ when $\underline{\xi}_i^s$ characterizes $\underline{\hat{e}}^s$, $i \in \{1, \ldots, n\}$ and*

$$\tilde{\mathbf{P}}^s = \frac{1}{n-1}\sum_{i=1}^n (\underline{\xi}_i^s - \bar{\underline{\xi}}^s)(\underline{\xi}_i^s - \bar{\underline{\xi}}^s)^\top \tag{2}$$

*is a consistent estimate of $\mathbf{P}^s$, i.e., $\lim_{n\to\infty} \tilde{\mathbf{P}}^s = \mathbf{P}^s$.*

While the recursive processing of samples is uniquely specified by the error process of the estimate, the noise generation has some degrees of freedom. For the considered system (1) with uncorrelated noise terms, it is meaningful to formalize the noise generation according to the following definition.

***Definition 3*** *A* ***noise sampler*** *with uncertainty $\mathbf{P}^\psi$ generates samples $\underline{\phi}_1, \ldots, \underline{\phi}_n$ with:*

a) *zero mean: $E\{\frac{1}{n}\sum_{i=1}^n \underline{\phi}_i\} = \underline{0}$,*
b) *asymptotically correct covariance: $\lim_{n\to\infty} \frac{1}{n-1}\sum_{i=1}^n (\underline{\phi}_i - \bar{\underline{\phi}})(\underline{\phi}_i - \bar{\underline{\phi}})^\top = \mathbf{P}^\psi$, where $\bar{\underline{\phi}} = \frac{1}{n}\sum_{i=1}^n \underline{\phi}_i$,*
c) *asymptotic independence: $\lim_{n\to\infty} \frac{1}{n-1}\sum_{i=1}^n (\underline{\phi}_i^1 - \bar{\underline{\phi}}^1)(\underline{\phi}_i^2 - \bar{\underline{\phi}}^2)^\top = \mathbf{0}$ for independently generated sample sets $\underline{\phi}_i^1$ and $\underline{\phi}_i^2$, $i \in \{1, \ldots, n\}$ (also for different noise samplers).*

Note that we focus on the second central moment to represent the noise densities. In particular, we have neither specified the distribution from which the samples are drawn (can be different from the noise distribution) nor have we demanded that the samples within a sample set are uncorrelated. Depending on the application, more restrictive definitions of noise samplers are meaningful as the precision of the covariance estimate indeed depends on those attributes.

It is also worth pointing out the difference between noise process $\underline{\boldsymbol{\psi}}$ and error process $\underline{\boldsymbol{e}}$. The latter one is a linear combination of several noise processes and is characterized by

samples $\underline{\xi}$, whereas the samples of noise terms are denoted as $\underline{\phi}$.

For local estimates, the noise errors can be examined independently of each other. Thus, the sample set characterizes the local covariance, if the noise is sampled with a noise sampler from Def. 3 and the noise samples are transformed the same way as the error process. However, for the more general problem of estimating the joint covariance of distributed estimates, we need to sample correlated noise as well. To this end, we define synchronized noise samplers.

***Definition 4*** *Let $\underline{\phi}_1^s, \dots, \underline{\phi}_n^s$ denote samples with means $\bar{\underline{\phi}}^s$ generated by different noise samplers $s \in \mathcal{S}$ with the same uncertainty $\mathbf{P}^\psi$. When $\lim_{n\to\infty} \sum_{i=1}^{n} (\underline{\phi}_i^{s_1} - \bar{\underline{\phi}}^{s_1})(\underline{\phi}_i^{s_2} - \bar{\underline{\phi}}^{s_2})^\top = \mathbf{P}^\psi$ for $s_1, s_2 \in \mathcal{S}$, we refer to them as* ***synchronized noise samplers****.*

Note that the sample sets of synchronized noise samplers are not generated independently of each other as it is mentioned in Def. 3(c). An easy way to realize synchronized noise samplers is to generate the same (ordered) list of noise samples with pseudo-random number generators. For this purpose, pseudo-random number generators at different nodes use network-wide known variables such as the current time step, a synchronized number, etc. as a common seed value.

Having the necessary noise sampler techniques available, we are able to propose a policy for the local processing of noise samples that does not require communication between nodes.

***Definition 5*** *Let noise terms be denoted by $\underline{\boldsymbol{\psi}}_t$ $t \in \mathcal{T}$. A* ***simple covariance-retaining sampling policy*** *consists of the following parts:*

a) *uncorrelated noise, i.e., $E\{\underline{\boldsymbol{\psi}}_{t_1} (\underline{\boldsymbol{\psi}}_{t_2})^\top\} = \mathbf{0}$, $t_1 \neq t_2 \in \mathcal{T}$, is sampled with non-synchronized noise samplers.*
b) *fully correlated noise, i.e., $E\{\underline{\boldsymbol{\psi}}_{t_1} (\underline{\boldsymbol{\psi}}_{t_2})^\top\} = \mathbf{P}_{t_1}^\psi (= \mathbf{P}_{t_2}^\psi)$, $t_1, t_2 \in \mathcal{T}$ is sampled with synchronized noise samplers.*
c) *the same linear transformations and linear combinations are applied to estimates and samples.*

The policy finally enables us to state our main theorems that pertain to the local processing of noise samples (Theorem 1) and to the combination of sample sets (Theorem 2). The proofs are given in the Appendix.

***Theorem 1*** *Consider error processes $\underline{\boldsymbol{e}}^s$, $s \in \mathcal{S}$ that consist of linear combinations of either uncorrelated or fully correlated zero mean noise terms. The joint error estimation process that is given by $\left((\underline{\boldsymbol{e}}^1)^\top \quad \dots \quad (\underline{\boldsymbol{e}}^S)^\top\right)^\top$ with corresponding joint covariance $\mathbf{P}$ is characterized by the sample set $\boldsymbol{\Xi} = \left((\boldsymbol{\Xi}^1)^\top \quad \dots \quad (\boldsymbol{\Xi}^S)^\top\right)^\top$, where $\boldsymbol{\Xi}^s$ is obtained by the local application of the simple covariance-retaining sampling policy from Def. 5. This means that*

$$\tilde{\mathbf{P}} = \frac{1}{n-1} \sum_{i=1}^{n} (\underline{\xi}_i - \bar{\underline{\xi}})(\underline{\xi}_i - \bar{\underline{\xi}})^\top \tag{3}$$

*with $\underline{\xi}_i = ((\underline{\xi}_i^1)^\top \dots (\underline{\xi}_i^S)^\top)^\top$ and $\bar{\underline{\xi}} = ((\bar{\underline{\xi}}^1)^\top \dots (\bar{\underline{\xi}}^S)^\top)^\top$ is a consistent estimator of the joint covariance $\mathbf{P}$.*

***Theorem 2*** *Same assumptions as in Theorem 1. Then, the linear combination of estimates $\hat{\underline{x}}^s$, $s \in \mathcal{S}' \subseteq \mathcal{S}$ according to*

$$\hat{\underline{x}}^{s'} = \sum_{s\in\mathcal{S}'} \mathbf{L}^s \hat{\underline{x}}^s \text{ and } \underline{\Xi}^{s'} = \sum_{s\in\mathcal{S}'} \mathbf{L}^s \underline{\Xi}^s \ , s' \in \mathcal{S} \ ,$$

*yields an estimate with an error process that is a linear combination of either uncorrelated or fully correlated zero mean noise terms as it is obtained by the simple covariance-retaining sampling policy and that is characterized by the sample set $\boldsymbol{\Xi} = ((\boldsymbol{\Xi}^1)^\top \dots (\boldsymbol{\Xi}^{s'})^\top \dots (\boldsymbol{\Xi}^S)^\top)^\top$.*

***Remark 1*** *As stated in Theorem 2, the sample structure is not corrupted due to linear combinations of sample sets. Therefore, consecutive combinations of sample sets and subsequent processing with the simple covariance-retaining sampling policy still provide consistent estimates of covariances.*

In summary, by means of Theorems 1 and 2 it is feasible to estimate the joint covariance (and in particular the cross-covariance terms) in linear filtering based on locally obtained samples. These results can be extended to other consistent estimators, e.g., the Gaussian maximum likelihood covariance estimator, as long as the evaluation of the covariance is adapted to the sampling policy.

## IV. Combination of Estimation Data Structures

In this section, we propose methods to integrate the estimation vector and the covariance into the sample structure. By this, we improve the precision of the covariance estimation, enhance the rate of convergence, and reduce the amount of data that must be stored and communicated between nodes. The idea is to exploit the concurrent storage and processing of error information in sample sets and covariances.

Basically, there are two concepts of linear filtering with samples. Either, sample sets are used actively in the filtering process, i.e., the gains are chosen based on the samples, or passive functionality is assigned to the sample sets by storing (additional) meta information only – as it was done in Sec. III.

Sample sets play an active role in many state-of-the-art algorithms such as the Unscented KF [9] or the Ensemble KF [11]. The advantage of applying transformations based on samples is that the estimation of nonlinear models is straightforward. However, when the noise terms are Gaussian distributed and the models are linear, it is well-known that the KF is the MMSE estimator. Therefore, only those algorithms that calculate the same gains and transformation matrices as the KF yield the MMSE result. While algorithms with a deterministic sampling strategy might coincide with the KF for linear models, stochastic sampling strategies, which are required for the covariance estimation in Sec. III, yield suboptimal gains in general.

Hence, we focus on a passive integration of samples in this section. Technically speaking, we leave it to the reader to choose the filter gains, e.g., by applying local KFs, ensemble KFs [11], Distributed KFs [18], [19], etc.

Let us assume that a subset of nodes $\mathcal{S}' \subseteq \mathcal{S}$ exchanges their local estimates, covariances and sample sets. As shown in Theorem 1 and Theorem 2, the joint covariance can be estimated by stacking together the samples from the distributed nodes and by calculating the corresponding sample covariance. Additionally, the local covariances, i.e., the block diagonal

entries of the joint covariance, are obtained (exactly) by methods of classic linear filtering theory.

Having these data available, we face the problem of combining the estimated joint covariance with the incomplete block-diagonal matrix consisting of local covariances. The naïve approach is to apply methods from covariance estimation theory that optimize a meaningful error criterion such as the entropy loss or the quadratic loss [20] and that have been discussed in Sec. I.

We pursue a different direction here. By adapting the noise samples generated by a noise sampler from Def. 3, we achieve a sample representation that already comprises the estimate and covariance information.

The first idea is to represent local estimates as sample means. Therefore, the invariant $\underline{\hat{x}} = \underline{\bar{\xi}} = \sum_{i=1}^{n} \underline{\xi}_i$ must hold in all processing steps. Assumed that the invariant is ensured at initialization and all noise terms are additive as described in Sec. II. Then a simple technique is to generate zero mean noise with $\underline{\bar{\psi}} = \frac{1}{n}\sum_{i=1}^{n} \underline{\psi}_i = \underline{0}$ and enforcing the invariant by adding the (same) proper deterministic bias term $\underline{b}$ to all samples. The correctness of this approach directly follows from Def. 1. We impose zero mean in the noise sample generation by a mean correction.

***Proposition 1*** *Let $\underline{\phi}_1, \ldots, \underline{\phi}_n$ be a set of noise samples with $\underline{\bar{\phi}} = \frac{1}{n}\sum_{i=1}^{n} \underline{\phi}_i$, generated by a noise sampler with uncertainty $\mathbf{P}^{\psi}$ from Def. 3. Then, $\underline{\phi}'_i = \underline{\phi}_i - \underline{\bar{\phi}}$ is a noise sampler with uncertainty $\mathbf{P}^{\psi}$ and $\underline{\bar{\phi}}' = \frac{1}{n}\sum_{i=1}^{n} \underline{\phi}'_i = \underline{0}$.*

PROOF. zero mean follows from the definition of the noise sampler. The asymptotic covariance is proven by

$$\lim_{n\to\infty} \frac{1}{n-1} \sum_{i=1}^{n} (\underline{\phi}_i - \underline{\bar{\phi}} - \underline{0})(\underline{\phi}_i - \underline{\bar{\phi}} - \underline{0})^{\top} = \mathbf{P}^{\psi} \ .$$

The asymptotic independence follows from the zero mean attribute of the underlying noise sampler as $\lim_{n\to\infty} \underline{\bar{\phi}} = \underline{0}$ and therefore, the noise sampler coincide for $n \to \infty$. □

Note that even if the values generated by the underlying noise sampler are uncorrelated, the noise samples after the mean correction are in general correlated for $n < \infty$. The considered asymptotic behavior, however, is not affected.

In a second step, we integrate the local covariance into the sample representation. To this end, we must ensure that the local sample covariance coincides with the exactly known local covariance, i.e., $\mathbf{P} = \frac{1}{n-1}\sum_{i=1}^{n}(\underline{\xi}_i - \underline{\bar{\xi}})(\underline{\xi}_i - \underline{\bar{\xi}})^{\top} = \tilde{\mathbf{P}}$. Unfortunately, the correlation between sampled noise and samples is not negligible for $n < \infty$, so that a covariance-enforcing sampling policy depends on the sample instances. We obtain the following noise sampling algorithm.

***Proposition 2*** *Let $\underline{\phi}_1, \ldots, \underline{\phi}_n$ be a set of noise samples generated by a noise sampler with uncertainty $\mathbf{P}^{\psi}$ from Def. 3 and $\underline{\xi}_1, \ldots, \underline{\xi}_n$ a set of samples with estimated uncertainty $\tilde{\mathbf{P}} = \mathbf{P}$. The cross-covariance between samples and noise samples is denoted by $\mathbf{P}^{\xi\phi} = \frac{1}{n-1}\sum_{i=1}^{n}(\underline{\xi}_i - \underline{\bar{\xi}})(\underline{\phi}_i - \underline{\bar{\phi}})^{\top}$. Assume furthermore that the error process is described by the sample processing*

$$\underline{\xi}_i^{+} = \mathbf{K}\underline{\xi}_i + \underline{\phi}_i \ . \tag{4}$$

*Then,*

$$\underline{\phi}'_i = \mathbf{T}(\underline{\phi}_i - \underline{\bar{\phi}}) \tag{5}$$

*with*

$$\mathbf{T} = \sqrt{\mathbf{P}^{\psi} + \mathbf{K}\mathbf{P}^{\xi\phi}(\mathbf{P}^{\phi})^{-1}\mathbf{P}^{\phi\xi}(\mathbf{K})^{\top}}\left(\sqrt{\mathbf{P}^{\phi}}\right)^{-1} - \mathbf{K}\mathbf{P}^{\xi\phi}\left(\mathbf{P}^{\phi}\right)^{-1}$$

*is a noise sampler that ensures that the sample covariance equals the local covariance*[2].

According to Proposition 2, it is possible to concurrently preserve the covariance representation and the asymptotic consistency in linear combinations of noise terms. Linearly transformed noise, as it appears for example in filter equations, must be transformed by inverting it with the linear transformation. For $\underline{\xi}_i^{+} = \mathbf{K}\underline{\xi}_i + \mathbf{L}\underline{\phi}_i$ we obtain $\mathbf{T} = \sqrt{-\Gamma}\left(\sqrt{\mathbf{P}^{\phi}}\right)^{-1} - (\mathbf{L})^{-1}\mathbf{K}\mathbf{P}^{\xi\phi}\left(\mathbf{P}^{\phi}\right)^{-1}$ with $\mathbf{\Gamma} = -(\mathbf{L})^{-1}(\mathbf{P}^{+} + \mathbf{K}\mathbf{P}^{\xi\phi}\left(\mathbf{P}^{\phi}\right)^{-1}\mathbf{P}^{\phi\xi}(\mathbf{K})^{\top} - \mathbf{K}\tilde{\mathbf{P}}(\mathbf{K})^{\top})(\mathbf{L})^{-\top}$.

In order to represent the true covariance by a set of samples, a necessary condition is $n > \dim\{\mathbf{P}\}$ (one eigenvalue is zero due to the mean correction, so $n = \dim\{\mathbf{P}\}$ is not sufficient), as otherwise the sample matrices are not positive definite and the Cholesky decomposition cannot be calculated. It is also worth mentioning that synchronized noise samplers that employ Proposition 2 generate different noise for $n < \infty$ when the local covariance or the correlation between noise and local covariance is different.

### *Usage of Sampling Techniques*

Depending on the estimation algorithm that is employed at the sensors, measurements are either combined in tracks with exclusively local information or they are comprised in local estimates, which may entail measurements from different sources. The first type of processing is sometimes referred to as scan-to-track and is for example used in the distributed Kalman filter [18] or the hypothesizing Kalman filter [19]. As the processing of purely local information provides exact covariances, the techniques from the last section can be applied directly. In contrast to that, estimate processing that involves fusion of recursively obtained information and the subsequent inclusion of measurements, requires additional thought.

As it has been shown in Sec. III, the asymptotic properties of the sample representation of the error process are not affected by linear combinations of sample sets. However, the proposed technique in Sec. IV aims at representing the local covariances exactly. When sample sets from different nodes are (linearly) combined[3], the resulting covariance depends on the cross-covariances. Since the cross-covariance terms are error-prone estimates of the true covariance, the local covariances of the fused estimates are error-prone as well.

For applications that do not require the covariance to be bounded conservatively, this problem can likely be ignored as the covariance of the fused estimate primarily depends on the well-known local covariances and the estimates of the cross-covariances are supposed to have a small error anyway. This especially applies to sensor networks with a low communication rate as then the local covariances mostly depend on exactly known (local) measurement errors and less on the covariance estimates from past data exchanges. Alternatively, covariance

[2]The square root $\sqrt{\cdot}$ denotes a Cholesky factor with $\sqrt{\mathbf{M}}\sqrt{\mathbf{M}}^{\top} = \mathbf{M}$.

[3]That includes the application of optimal fusion algorithm such as [2], [5]

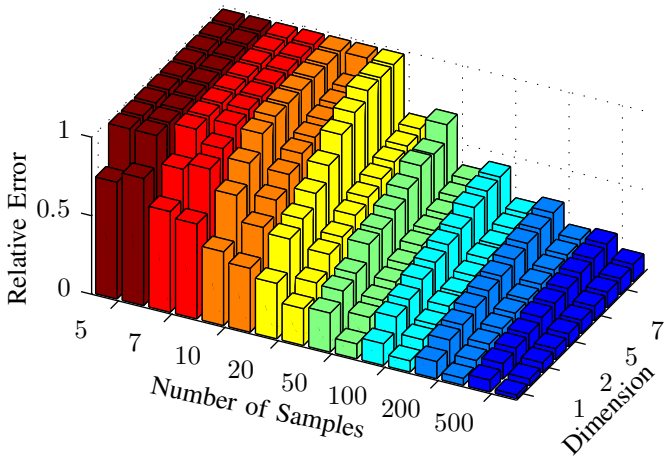


Figure 1. The relative deviation of the estimated covariance from the true covariance. The results of scenarios with the same parameters are denoted in the same color. The left bars illustrate the results of the procedure from Sec. III and the right bars of the covariance-enforcing policy. Error values greater than 1 are clipped.

inflation ideas [7] combined with shrinkage techniques [12] can be used that artificially enlarge local covariances in order to account for uncertainties in the cross-covariance estimate and prevent an underestimation of the local covariance.

## V. Evaluation

Finally, we examine the precision of sample-based covariance estimation with a finite number of samples. We vary the dimension of the system state from 1 to 10 and the number of samples per node between 5 and 500. The model depends on the dimension of the system $\dim\{\underline{\boldsymbol{x}}\}$ and is given by $\underline{\boldsymbol{x}}_{k+1} = \begin{pmatrix} 1 & \dots & 1 \\ & \ddots & \vdots \\ \mathbf{0} & & 1 \end{pmatrix} \underline{\boldsymbol{x}}_k + \underline{\boldsymbol{w}}_k$ , with $\underline{\boldsymbol{w}}_k \sim \mathcal{N}(\underline{0}, \mathbf{I})$. Two nodes observe the system in each time step according to the models $\underline{z}^i = \underline{\boldsymbol{x}} + \underline{\boldsymbol{v}}^i$, $i \in \{1,2\}$ with $\underline{\boldsymbol{v}}^1 \sim \mathcal{N}(\underline{0}, \mathbf{I})$ and $\underline{\boldsymbol{v}}^2 \sim \mathcal{N}(\underline{0}, 2 \cdot \mathbf{I})$. The initial estimates are uncorrelated with covariances $\mathbf{P}_1^1 = \mathbf{I}$ and $\mathbf{P}_1^2 = 2 \cdot \mathbf{I}$. The covariance is estimated by means of sample sets. The noise is sampled from a Gaussian distribution. After 10 time steps, the samples are exchanged and the reconstructed joint covariance is compared with the actual one.

We examine the Euclidean matrix norm of the difference between estimated and true covariance normalized by the Euclidean matrix norm of the true covariance, i.e., $\|\tilde{\mathbf{P}} - \mathbf{P}\|_2 / \|\mathbf{P}\|_2$, which gives a relative deviation of the covariance estimate from the true covariance. The results of the procedure from Sec. III are compared to the covariance-enforcing strategy from Sec. IV in 50 Monte-Carlo runs.

As it can be seen in Fig. 1, the estimation precision improves with the number of samples and is in general higher when the dimension of the considered system is smaller. The covariance-enforcing procedure is significantly better than the standard approach for almost all dimensions and sample set sizes. Furthermore, the covariance estimates in scenarios with less than 10 samples have a high error attached. For sample sets of sizes 20 and 50, the relative error of the covariance-enforcing algorithm is reduced to $0.2 - 0.6$ and $0.1 - 0.4$, respectively.

Especially for small sample sets, a small increase in the number of samples increases the precision significantly. By comparing for example the error of the covariance-enforcing policy with $n = 10; \dim\{\underline{\boldsymbol{x}}\} = 2$ with $n = 20; \dim\{\underline{\boldsymbol{x}}\} = 10$, which is approximately 0.6 in both cases, one can see that the error remains the same when the dimension is quintupled while the number of samples is only doubled.

Therefore, an evaluation of the theoretic performance of the sampling algorithm should always precede an application of the algorithm. Eventually, the number of samples must be weighed against the estimation performance. Especially in scenarios with low communication costs or when the state dimension is high, the proposed scheme is useful. For other scenarios, it might be worth to evaluate different sampling strategies or other noise transformations.

## VI. Conclusions

We have proposed a sample-based approach for the representation of error processes in linear systems that permits the reconstruction of the joint covariance based on locally obtained data. Although the focus was on the asymptotic properties of the covariance estimation scheme, we also proposed a technique to integrate estimate and covariance into the sample representation and evaluated the different approaches.

Most of the results presented in this paper pertain only asymptotic properties of sample-based cross-covariance estimation. Future research is likely to focus on the derivation of quality attributes for a finite number of samples and the enhancement of sampling strategies to minimize the error. In particular, sampling strategies that take into account higher moments can be considered and policies to sample partial correlations can be derived. Apart from that, the ideas developed here can be easily extended to nonlinear models.

## Appendix

**Proof of Theorem 1.** In a first step, we prove that by applying the simple covariance-retaining sampling policy, the samples characterize the local error process and the sample sets characterize the local covariance.

Consider the sample set $\underline{\xi}_1, \dots, \underline{\xi}_n$. According to Def. 5, the same transformations are applied to estimate and noise samples, which guarantees that the samples characterize the error process. Therefore, we have

$$\underline{\hat{e}}^s \sim \sum_{t \in \mathcal{T}^s} \mathbf{B}_t^s \underline{\boldsymbol{\psi}}_t \text{ and } \underline{\xi}_i^s = \sum_{t \in \mathcal{T}^s} \mathbf{B}_t^s \underline{\phi}_{t,i}^s \ , i \in \{1, \dots, n\} \ .$$

According to the assumptions, the error terms are uncorrelated for $n \to \infty$ and as the noise samples $\underline{\phi}_{t,i}^s$ have been generated independently (c.f., Def. 3(b)), it directly follows $E\{\underline{\boldsymbol{\psi}}_t (\underline{\boldsymbol{\psi}}_t)^\top\} = \lim_{n \to \infty} \frac{1}{n-1} \sum_{i=1}^n (\underline{\phi}_{t,i}^s - \bar{\underline{\phi}}_{t,i}^s)(\underline{\phi}_{t,i}^s - \bar{\underline{\phi}}_{t,i}^s)^\top$.

It remains to show that the cross-covariances are obtained correctly. Let $s_1 \neq s_2 \in \mathcal{S}$ denote the indices of two sensors. The local error processes comprise noise terms with indices $\mathcal{T}^{s_i} \subseteq \mathcal{T}^{\mathcal{S}}$, $s_i \in \mathcal{S}$. Then, the cross-covariance between these estimates is given by

$$\mathbf{P}^{s_1 s_2} = \sum_{t_1 \in \mathcal{T}^{s_1}, t_2 \in \mathcal{T}^{s_2}} \mathbf{B}_{t_1}^{s_1} \, E\{\underline{\boldsymbol{\psi}}_{t_1} (\underline{\boldsymbol{\psi}}_{t_2})^\top\} (\mathbf{B}_{t_2}^{s_2})^\top \ .$$

For $t_1 \neq t_2$, the sampled cross-covariance for $n \to \infty$ is $\mathbf{0}$ according to Def. 3(c). Therefore, we obtain

$$\frac{1}{n-1} \sum_{i=1}^{n} \sum_{t \in \mathcal{T}^{s_1} \cap \mathcal{T}^{s_2}} \mathbf{B}_t^{s_1} (\underline{\phi}_{t,i}^{s_1} - \bar{\underline{\phi}}_{t,i}^{s_1})(\underline{\phi}_{t,i}^{s_2} - \bar{\underline{\phi}}_{t,i}^{s_2})^\top (\mathbf{B}_t^{s_2})^\top \ ,$$

and by changing the position of the sum

$$\sum_{t\in\mathcal{T}^{s_1}\cap\mathcal{T}^{s_2}} \mathbf{B}_t^{s_1}\left(\frac{1}{n-1}\sum_{i=1}^{n}(\underline{\phi}_{t,i}^{s_1}-\bar{\underline{\phi}}_{t,i}^{s_1})(\underline{\phi}_{t,i}^{s_2}-\bar{\underline{\phi}}_{t,i}^{s_2})^\top\right)(\mathbf{B}_t^{s_2})^\top .$$

According to the simple covariance-retaining sampling policy from Def. 5, correlated noise is sampled with synchronized noise samplers and thus, from Def. 4 it follows that this term equals $\mathbf{P}^{ij}$. □

PROOF OF THEOREM 2. We consider samples as instances of linear combinations of the true noise terms $\underline{\psi}_t$ with $\bar{\underline{\xi}}^s \sim \sum_{t\in\mathcal{T}^s}\mathbf{B}_t^s\underline{\psi}_t$, $s\in\mathcal{S}$. Then, the fused error process is given by the linear combination of fully correlated and uncorrelated noise terms

$$\hat{\underline{e}}^{s'} \sim \sum_{t\in\bigcup_{s\in\mathcal{S}'}\mathcal{T}^s}\mathbf{B}_t^{s'}\underline{\psi}_t \text{ with } \mathbf{B}_t^{s'} = \sum_{s \text{ with } t\in\mathcal{T}^s}\mathbf{L}^s\mathbf{B}_t^s , \quad (6)$$

where $\mathbf{B}_t^s$ denotes the (unaltered) local transformation of the noise and $\mathbf{L}^s$ is the transformation matrix in the fusion. The same considerations apply for the samples so that they characterize the noise.

It remains to show that the sample-based covariance estimation is consistent. This, however, follows along the lines of the proof of Theorem 1 as the samples from uncorrelated noise are not only uncorrelated from other local noise terms but also from noise of remote noise samplers. Therefore, the sampled covariance for $n\to\infty$ is given by $\sum_{t\in\bigcup_{s\in\mathcal{S}'}\mathcal{T}^s}\mathbf{N}_t$ with

$$\mathbf{N}_t = \sum_{\substack{s_1,s_2 \text{ with}\\ t\in\mathcal{T}^{s_1}\cap\mathcal{T}^{s_2}}}\mathbf{L}^{s_1}\mathbf{B}_t^{s_1}\mathbf{M}_t^{s_1s_2}(\mathbf{B}_t^{s_2})^\top(\mathbf{L}^{s_2})^\top ,$$

where $\mathbf{M}_t^{s_1s_2}$ denotes the sample covariance of correlated noise between nodes $s_1$ and $s_2$. For $n\to\infty$, $\mathbf{M}_t^{s_1s_2} = E\{\underline{\psi}_t(\underline{\psi}_t)^\top\}$ and thus, $\mathbf{N}_t = \mathbf{B}_t^{s'}E\{\underline{\psi}_t(\underline{\psi}_t)^\top\}(\mathbf{B}_t^{s'})^\top$. But with this, $\sum_{t\in\bigcup_{s\in\mathcal{S}'}\mathcal{T}^s}\mathbf{N}_t$ represents the true covariance of the fused estimate that is obtained from (6). □

PROOF OF PROPOSITION 2. First we note that the underlying true covariance process is given by

$$\mathbf{P}^+ = \mathbf{K}\mathbf{P}(\mathbf{K})^\top + \mathbf{P}^\psi .$$

The sample covariance with noise samples transformed by (5) is given by

$$\tilde{\mathbf{P}}^+ = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{K}(\underline{\xi}_i-\bar{\underline{\xi}})-\mathbf{T}(\underline{\phi}_i-\bar{\underline{\phi}}-\underline{0}))\cdot (\mathbf{K}(\underline{\xi}_i-\bar{\underline{\xi}})-\mathbf{T}(\underline{\phi}_i-\bar{\underline{\phi}}-\underline{0}))^\top =$$

$$\mathbf{K}\mathbf{P}(\mathbf{K})^\top - \mathbf{T}\mathbf{P}^{\phi\xi}(\mathbf{K})^\top - \mathbf{K}\mathbf{P}^{\xi\phi}(\mathbf{T})^\top + \mathbf{T}\mathbf{P}^\phi(\mathbf{T})^\top .$$

We transform this equality to the matrix equation

$$(\mathbf{T}\quad \mathbf{I})\underbrace{\begin{pmatrix}\mathbf{P}^\phi & \mathbf{P}^{\phi\xi}(\mathbf{K})^\top\\ \mathbf{K}\mathbf{P}^{\xi\phi} & \mathbf{K}\mathbf{P}(\mathbf{K})^\top-\mathbf{P}^+\end{pmatrix}}_{\mathbf{E}}\begin{pmatrix}(\mathbf{T})^\top\\ \mathbf{I}\end{pmatrix} = \mathbf{0} ,$$

and obtain with $\mathbf{\Gamma} = \mathbf{K}\tilde{\mathbf{P}}(\mathbf{K})^\top - \mathbf{P}^+ - \mathbf{K}\mathbf{P}^{\xi\phi}(\mathbf{P}^\phi)^{-1}\mathbf{P}^{\phi\xi}(\mathbf{K})^\top = -\mathbf{P}^\psi - \mathbf{K}\mathbf{P}^{\xi\phi}(\mathbf{P}^\phi)^{-1}\mathbf{P}^{\phi\xi}(\mathbf{K})^\top$ the lower block triangular matrix

$$\mathbf{D} = \begin{pmatrix}\sqrt{\mathbf{P}^\phi} & \mathbf{0}\\ \mathbf{K}\mathbf{P}^{\xi\phi}(\sqrt{\mathbf{P}^\phi})^{-\top} & \sqrt{\mathbf{\Gamma}}\end{pmatrix}$$

as square root of $\mathbf{E}$. From this, we get the condition $(\mathbf{T}\sqrt{\mathbf{P}^\phi}+\mathbf{K}\mathbf{P}^{\xi\phi}(\sqrt{\mathbf{P}^\phi})^{-\top})^2+\mathbf{\Gamma}=\mathbf{0}$. This is equivalent to $\mathbf{T}\sqrt{\mathbf{P}^\phi}+\mathbf{K}\mathbf{P}^{\xi\phi}(\sqrt{\mathbf{P}^\phi})^{-\top} = \sqrt{-\mathbf{\Gamma}} \Leftrightarrow \mathbf{T} = (\sqrt{-\mathbf{\Gamma}} - \mathbf{K}\mathbf{P}^{\xi\phi}(\sqrt{\mathbf{P}^\phi})^{-\top})(\sqrt{\mathbf{P}^\phi})^{-1} = \sqrt{-\mathbf{\Gamma}}(\sqrt{\mathbf{P}^\phi})^{-1} - \mathbf{K}\mathbf{P}^{\xi\phi}(\mathbf{P}^\phi)^{-1}$.

zero mean follows from $E\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{T}(\underline{\phi}_i-\bar{\underline{\phi}})\} = E\{\mathbf{T}\frac{1}{n}(-n\bar{\underline{\phi}}+\sum_{i=1}^{n}\underline{\phi}_i)\} = \mathbf{T}\,E\{\underline{0}\}$. The asymptotic attributes of the transformation follow from Proposition 2 as with $\lim_{n\to\infty}\mathbf{P}^{\xi\phi}=\mathbf{0}$ and $\lim_{n\to\infty}\mathbf{P}^\phi=\mathbf{P}^\psi$, the solution is given by $\mathbf{T}=\mathbf{I}$. □

## REFERENCES


[1] M. E. Liggins, D. L. Hall, and J. Llinas, *Handbook of Multisensor Data Fusion: Theory and Practice*, 2nd ed., C. R. C. Press, Ed., 2008.

[2] Y. Bar-Shalom and L. Campo, "The Effect of the Common Process Noise on the Two-Sensor Fused-Track Covariance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 6, pp. 803–805, Nov. 1986.

[3] S. Grime and H. F. Durrant-Whyte, "Data Fusion in Decentralized Sensor Networks," *Control Engineering Practice*, vol. 2, no. 5, pp. 849–863, 1994.

[4] K. H. Kim, "Development of Track to Track Fusion Algorithms," in *Proceedings of the American Control Conference 1994*, vol. 1. IEEE, 1994, pp. 1037–1041.

[5] S.-l. Sun, "Multi-sensor optimal information fusion kalman filters with applications," *Aerospace Science and Technology*, vol. 8, no. 1, pp. 57–62, 2004.

[6] V. Shin, Y. Lee, and T.-S. Choi, "Generalized Millman's Formula and Its Application for Estimation Problems," *Signal Processing*, vol. 86, no. 2, pp. 257–266, 2006.

[7] S. J. Julier and J. K. Uhlmann, "A Non-divergent Estimation Algorithm in the Presence of Unknown Correlations," in *Proceedings of the American Control Conference 1997*, vol. 4, 1997, pp. 2369–2373.

[8] J. Sijs, M. Lazar, and P. P. J. v. d. Bosch, "State Fusion with Unknown Correlation: Ellipsoidal Intersection," in *Proceedings of the American Control Conference 2010*, 2010, pp. 3992–3997.

[9] S. J. Julier and J. K. Uhlmann, "New Extension of the Kalman Filter to Nonlinear Systems," vol. 3068, no. 1. SPIE, 1997, pp. 182–193. [Online]. Available: http://dx.doi.org/doi/10.1117/12.280797

[10] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle Filters for Positioning, Navigation, and Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 425–437, 2002.

[11] G. Evensen, "The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation," *Ocean dynamics*, vol. 53, no. 4, pp. 343–367, 2003.

[12] W. James and C. Stein, "Estimation with Quadratic Loss," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 1961, 1961, pp. 361–379.

[13] M. Pourahmadi, "Joint Mean-covariance Models with Applications to Longitudinal Data: Unconstrained Parameterisation," *Biometrika*, vol. 86, no. 3, pp. 677–690, 1999.

[14] A. J. Rothman, E. Levina, and J. Zhu, "A new Approach to Cholesky-based Covariance Regularization in High Dimensions," *Biometrika*, vol. 97, no. 3, pp. 539–550, 2010.

[15] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse Inverse Covariance Estimation with the Graphical Lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[16] M. Pourahmadi, "Covariance Estimation: The GLM and Regularization Perspectives," *Statistical Science*, vol. 26, no. 3, pp. 369–387, 2011.

[17] R. Vershynin, "How Close is the Sample Covariance Matrix to the Actual Covariance Matrix?" *Journal of Theoretical Probability*, vol. 25, no. 3, pp. 655–686, 2012.

[18] F. Govaers and W. Koch, "An Exact Solution to Track-to-Track-Fusion at Arbitrary Communication Rates," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, 2012.

[19] M. Reinhardt, B. Noack, and U. D. Hanebeck, "The Hypothesizing Distributed Kalman Filter," in *Proceedings of the 2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2012)*, Hamburg, Germany, Sep. 2012.

[20] R. Yang and J. O. Berger, "Estimation of a covariance matrix using the reference prior," *The Annals of Statistics*, pp. 1195–1211, 1994.