Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

A Novel Information Theoretic Measure Based Sensor Network Design Approach for Steady State Linear Data Reconciliation

Om Prakash * **Mani Bhushan** *

* *Department of Chemical Engineering, Indian Institute of Technology Bombay, India (e-mail: prakash.om@iitb.ac.in, mbhushan@iitb.ac.in)*

Abstract: The current work proposes a novel information theoretic based sensor network design (SND) approach for data reconciliation in a steady state linear process. The proposed approach is based on Kullback-Leibler divergence (KLD), which measures the difference of a density function from a reference density function. In particular, the optimal design is the one that leads to the smallest KLD value of the designed density function of the estimates from a reference density function. This reference density function can be provided by the end-user, and the approach thus enables explicit incorporation of the end-user's preference in the SND procedure. Additionally, the approach does not assume specific forms for the density functions of the estimates and is thus also applicable for cases when the estimates have non-Gaussian density. The significance of the approach is illustrated on a small example. To demonstrate its utility in obtaining optimal sensor networks, it is also applied to a popular case study from SND literature and results are compared with existing approaches.

Keywords: Kullback-Leibler divergence, Gaussian mixture model

1. INTRODUCTION

Given a set of measurements and a process model, obtaining better estimates of the measured variables as well as estimating the unmeasured variables is the problem of data reconciliation, which has been a widely studied problem in literature (Narasimhan and Jordache, 1999). However, the success of any data reconciliation approach depends critically on the choice of variables being measured in the process. Choosing the optimal set of variables to be measured is the sensor network design (SND) problem. In literature, the available SND approaches can be broadly classified as considering the design problem mainly from an individual variable, or a systems perspective.

From the systems perspective, Kretsovalis and Mah (1987) presented an SND approach for a steady state linear process, where the trace of the covariance matrix of estimates of variables is minimized to increase the estimation accuracy of those variables. Joshi and Boyd (2008) minimized the determinant of the covariance matrix of estimates to obtain optimal sensor network. Krause et al. (2008) discussed several ways to quantify the smallness of the covariance matrix, which in turn lead to different approaches for SND based on A-, D-, T-, and E-optimality criteria. In particular, Nabil and Narasimhan (2012) focused on A-optimal design approach for SND in a steady state linear process. They emphasized on the process economics while formulating the SND objective, which involved minimization of the trace of the weighted covariance matrix. Further, Balaji et al. (2018) used A-optimal design approach of SND for optimal selection of reference components in a reaction system, where the measurement model was linear. It should be noted that the A-, D-, T-, and E- optimal designs implicitly assume Gaussianity since they work only with the covariance matrix of the estimates and, as such, cannot directly address scenarios where the estimates are non-Gaussian. Additionally, even for Gaussian cases, these design approaches, in general, lead to different optimal sensor networks, and the choice of the appropriate approach to be used is thus not clear. Further, in such design approaches, there is no way for the end-user to directly specify her preference about the desired performance of the sensor network.

The current work focuses on the SND from systems perspective. In particular, we present a novel approach for SND for data reconciliation in a steady state linear process where an information theoretic framework based on Kullback-Leibler divergence (KLD) is considered. The proposed approach involves the specification of a reference (target) distribution of the estimates by the end-user. The optimal SND is the design that leads to a distribution of the estimates, which is closest to the reference using the KLD measure. Thus, our approach does not assume the estimates to have Gaussian distributions and can work with any arbitrary distribution. Further, the approach provides a direct way for the end-user to specify a desired reference distribution. In the current work, the reference distribution is specified to be the distribution of the estimates for the ideal (but non-implementable) case when all variables in the process are measured. Recently, Jusoh and Ampountolas (2019) have proposed the use of KLD for optimal selection of traffic sensors. However, unlike a physics based linear model considered in our work, their approach is data driven. The presence of a model leads to degenerate density functions for the estimates. This degeneracy is addressed in our work by considering the

density function of the estimates of a carefully chosen subset of variables.

The rest of the paper is organized as follows. KLD, and relevant approach to data reconciliation and SND are summarized in Section 2. Section 3 presents the SND formulation proposed in this work and illustrates it with an example. The proposed formulation is applied to a case study in Section 4, and paper is concluded in Section 5.

2. RELEVANT BACKGROUND

2.1 Introduction to Kullback-Leibler divergence

For a continuous random variable $\mathbf{x}$, Kullback-Leibler divergence (KLD) is an asymmetric measure of how far a probability density function (pdf) $g(\mathbf{x})$ is from a reference pdf $f(\mathbf{x})$. Denoted as D_{KL}, it is defined as (Hershey and Olsen, 2007):

$$D_{KL}(f(\mathbf{x})||g(\mathbf{x})) := \int_{-\infty}^{\infty} f(\mathbf{x}) \ln\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) d\mathbf{x} \quad (1)$$

For multivariate Gaussian distributions, namely $f(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu_f}, \boldsymbol{\Sigma_f})$, and $g(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu_g}, \boldsymbol{\Sigma_g})$, KLD can be analytically obtained as (Hershey and Olsen, 2007):

$$D_{KL}\left(f(\mathbf{x})||g(\mathbf{x})\right) = \frac{1}{2}\left[tr\left(\boldsymbol{\Sigma_f}\boldsymbol{\Sigma_g}^{-1}\right) + \ln\left(\frac{det(\boldsymbol{\Sigma_g})}{det(\boldsymbol{\Sigma_f})}\right) + (\boldsymbol{\mu_g} - \boldsymbol{\mu_f})^T\boldsymbol{\Sigma_g}^{-1}(\boldsymbol{\mu_g} - \boldsymbol{\mu_f}) - n\right] \quad (2)$$

where $tr(\cdot)$, $det(\cdot)$, and $(\cdot)^T$ is the trace, the determinant, and the transpose of a matrix $(\cdot)$, respectively, and n is the dimension of random variable $\mathbf{x}$.

2.2 Data Reconciliation

A general measurement model is:

$$\mathbf{y} = \mathbf{x} + \mathbf{v} \quad (3)$$

where $\mathbf{y} \in \mathbb{R}^n$ is the measurement vector, $\mathbf{x} \in \mathbb{R}^n$ is the corresponding vector of true values of the variables, and $\mathbf{v} \in \mathbb{R}^n$ is the measurement noise vector. The measurement noise is considered to be zero mean random vector since it is assumed that the gross errors in the measurements, if any, are pre-treated. In the current work, the true variables are assumed to be linearly related to each other by a known process model given as follows:

$$\mathbf{Cx} = \mathbf{0} \quad (4)$$

where $\mathbf{C} \in \mathbb{R}^{m\times n}$ represents the constraint matrix of the process model, and m is the number of independent constraints in the process.

Typically all the variables in a process are not measured. The general approach to solve the steady state data reconciliation problem in such cases is to rewrite the process model so that it involves only the measured variables, and minimize the least square residuals of the measurements, subject to the model constraints (4) but now expressed only in terms of the measured variables. The estimates of the unmeasured variables are subsequently obtained using the reconciled estimates of the measured variables. The model equations thus used in the two steps are dependent on the measured variables (Narasimhan and Jordache, 1999). Chmielewski et al. (2002) proposed an equivalent formulation which integrated the measurement model in (3) and the model constraints in (4) into one model. Towards this end, $\mathbf{x}$ is partitioned into two disjoint sets: primary variables ($\mathbf{x_p}$) and secondary variables ($\mathbf{x_s}$), such that $\mathbf{x_p}$ contains the minimum number of variables satisfying the observability criteria, which is defined as the ability to estimate all variables using either direct measurements or estimates using measurements of other variables and the process model (Ali and Narasimhan, 1993). Then, (4) can be written as:

$$\mathbf{C_p x_p} + \mathbf{C_s x_s} = \mathbf{0} \quad (5)$$

where $\mathbf{C_p}$ and $\mathbf{C_s}$ are appropriately partitioned submatrices of $\mathbf{C}$. Further, it should be noted that the choice of primary variables ($\mathbf{x_p}$) is not unique, but every set of primary variables will consist of $p = n - m$ variables. To ensure observability $\mathbf{x_p}$ should be chosen such that $\mathbf{C_s}$ is a full rank matrix in (5), and thus $\mathbf{x_s}$ can be written as:

$$\mathbf{x_s} = -\mathbf{C_s}^{-1}\mathbf{C_p x_p} \quad (6)$$

Therefore, the measurement and process model in (3) and (4) can be integrated into a single model as:

$$\mathbf{y} = \mathbf{A x_p} + \mathbf{v} \;\text{ with, }\; \mathbf{A} = \begin{bmatrix} \mathbf{I_p} \\ -\mathbf{C_s}^{-1}\mathbf{C_p} \end{bmatrix} \quad (7)$$

where $\mathbf{I_p}$ is an identity matrix of size $p \times p$.

2.3 Existing SND Approach

Existing SND formulations from the systems perspective for data reconciliation in a steady state linear process involve A-, D-, T-, or E- optimal designs. These design approaches assume that the measurement noise $\mathbf{v}$ (in (3) and (7)) has a Gaussian distribution, i.e., $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma_v})$. For the specific case when the measurement noises are uncorrelated, the solution to the data reconciliation problem is given as (Chmielewski et al., 2002):

$$\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{x}}_\mathbf{p} = \mathbf{A}(\mathbf{A}^T\boldsymbol{\Sigma_v}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Sigma_v}^{-1}\mathbf{y}, \quad (8a)$$

$$\text{cov}(\hat{\mathbf{y}}) = \boldsymbol{\Sigma}_{\hat{\mathbf{y}}} = \mathbf{A}\boldsymbol{\Sigma}_{\hat{\mathbf{x}}_\mathbf{p}}\mathbf{A}^T \quad (8b)$$

$$\text{where, } \boldsymbol{\Sigma}_{\hat{\mathbf{x}}_\mathbf{p}} = \text{cov}(\hat{\mathbf{x}}_\mathbf{p}) = (\mathbf{A}^T\boldsymbol{\Sigma_v}^{-1}\mathbf{A})^{-1}, \quad (8c)$$

$$\boldsymbol{\Sigma_v}^{-1} = diag(q_i/\sigma_i^2), \quad i = 1, 2, \ldots, n, \quad (8d)$$
$$\text{and } q_i \in \{0, 1\}, \quad i = 1, 2, \ldots, n$$

where $\hat{(\cdot)}$ represents the estimate of the variable $(\cdot)$, $diag(\cdot)$ represents a diagonal matrix with i^{th} diagonal entry given in $(\cdot)$, σ_i^2 is the variance of measurement noise of sensor used to measure variable x_i, and q_i is the binary variable to indicate the presence ($q_i = 1$) or absence ($q_i = 0$) of measurement y_i. Note that invertibility of $\mathbf{A}^T\boldsymbol{\Sigma_v}^{-1}\mathbf{A}$ in (8c) is ensured when observability criteria is satisfied.

The generic optimization formulation for SND is as follows:
Formulation I:

$$\min_{\mathbf{q}} \;\; J(\mathbf{q}) \;\text{ subject to, }\; \sum_{i=1}^{n} q_i = N^* \quad (9)$$

where $\mathbf{q} = \{q_i : q_i \in \{0, 1\},\ \forall i \in \{1, \ldots, n\}\}$ is the decision variable for the SND problem, and N^* is the user specified number of sensors to be placed in the process. The objective $J(\mathbf{q})$ in (9), depending on various existing design approaches, is given as follows (Krause et al., 2008):

$$\text{A-opt.: } trace(\boldsymbol{\Sigma}_{\hat{\mathbf{y}}}),\ \text{D-opt.: } det(\boldsymbol{\Sigma}_{\hat{\mathbf{y}}}) \quad (10)$$

$$\text{T-opt.: } -trace(\boldsymbol{\Sigma}_{\hat{\mathbf{y}}}^{-1}),\ \text{E-opt.: } -\min eig(\boldsymbol{\Sigma}_{\hat{\mathbf{y}}}^{-1}) \quad (11)$$

where $eig(\cdot)$ represents the eigen value of a matrix $(\cdot)$. It is to be noted that the covariance matrix $\mathbf{\Sigma_{\hat{y}}}$ is a rank deficient matrix since matrix $\mathbf{A} \in \mathbb{R}^{n \times p}$ in (7), is a tall matrix. Equivalently, this can be understood by noting that the process variables are related by the process model (4), and hence the covariance matrix of the estimates of all the variables will be rank deficient. This rank deficiency of $\mathbf{\Sigma_{\hat{y}}}$ precludes use of D-, T-, and E- optimal design approaches to obtain optimal sensor networks. Hence, to enable the applicability of these approaches in the current work, the objectives in A-, D-, T-, and E-optimal designs are defined using the covariance matrix $\mathbf{\Sigma_{\hat{x}_p}}$ (in (8c)) of the estimates of the primary variables only since $\mathbf{\Sigma_{\hat{x}_p}}$ is a full rank matrix. With this modification, these design approaches can be applied to obtain optimal sensor networks. However, in general, they will result in different optimal sensor networks since these existing design approaches measure the smallness of the covariance matrix in different ways.

Remark 1. The assumption of uncorrelated measurement noise was just to simplify the presentation. The existing objectives for SND, mentioned in (10-11) (with $\mathbf{\Sigma_{\hat{y}}}$ replaced by $\mathbf{\Sigma_{\hat{x}_p}}$), are valid for correlated measurement noise case as well. For such a scenario, the estimates and the covariance matrix of the estimates can be obtained using the two-step data reconciliation approach discussed in Section 2.2.

3. PROPOSED FORMULATION

The existing design approaches for SND (Formulation I, Section 2.3) implicitly assume that the measurement noise has a Gaussian distribution since they involve minimizing an objective function which depends only on the covariance matrix of the estimates. Hence, these approaches are not directly applicable for the case when measurement noises have non-Gaussian distributions. These existing approaches also do not provide a mechanism for the end-user to specify her specifications for the quality of estimates. In the current section, we propose an information theoretic formulation based on Kullback-Leibler divergence for SND in a steady state linear process. This formulation does not assume a specific form of pdf for the estimates. Further, in the proposed formulation, the end-user can specify the reference (target) pdf of the estimates, and the optimal design is the one which is as close to the reference pdf as possible. KLD can be used to quantify the distance between the pdf obtained by the designed sensor network from the reference pdf. Thus, the proposed optimization formulation for SND could be given as:

$$\min_{\mathbf{q}} \; D_{KL}(f(\hat{\mathbf{x}}^{\mathbf{ref}})||g(\hat{\mathbf{x}})) \text{ subject to, } \sum_{i=1}^{n} q_i = N^* \quad (12)$$

The estimates $\hat{\mathbf{x}}$ and its pdf $g(\hat{\mathbf{x}})$ depend on the selected sensor network $\mathbf{q}$. In the formulation in (12), $f(\hat{\mathbf{x}}^{\mathbf{ref}})$ is the reference pdf of the estimates, which can be provided by the end-user. In the current work, $f(\hat{\mathbf{x}}^{\mathbf{ref}})$ is taken to be the pdf of the estimates when all the variables are measured ($q_i = 1, \; \forall i \in \{1, \ldots, n\}$) in the process. This case (all variables measured) can be considered to be the ideal case, and thus the optimal sensor network design is the one which leads to the least distance of the designed pdf from this reference pdf.

Note that the process model in (4) needs to be incorporated while estimating $\mathbf{x}$ using any estimation approach. However, presence of this model leads to dependence between the estimates $\hat{\mathbf{x}}$. As a result, pdf $f(\hat{\mathbf{x}}_{\mathbf{ref}})$ and $g(\hat{\mathbf{x}})$ of the estimates of all variables become degenerate. In particular, since the process model provides m linearly independent constraints between the n process variables, the support of $f(\hat{\mathbf{x}}^{\mathbf{ref}})$ and $g(\hat{\mathbf{x}})$ lies in an $n-m$ dimensional subspace of $\mathbb{R}^n$. Thus, to avoid degeneracy, we propose to use KLD of designed pdf of the estimates $(\hat{\mathbf{x}}_{\mathbf{p}})$ of the primary variables from the corresponding reference pdf of the estimates $(\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}})$ of the primary variables in the sensor network design procedure. Thus, the proposed SND formulation is:

Formulation II:

$$\min_{\mathbf{q}} D_{KL}(f(\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}})||g(\hat{\mathbf{x}}_{\mathbf{p}})) \text{ subject to, } \sum_{i=1}^{n} q_i = N^* \quad (13)$$

where $f(\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}})$ is the reference pdf of the estimates of primary variables when all variables are measured in the process, while $g(\hat{\mathbf{x}}_{\mathbf{p}})$ is the pdf of the estimates of primary variables $\mathbf{x_p}$ obtained with the designed sensor network. These pdfs will depend on the estimation procedure used to estimate the primary variables given the measurements in the process. This issue is discussed in the next subsection. It should be noted that the proposed sensor network design formulation (Formulation II) is an integer nonlinear programming problem.

Remark 2. In this work, we have not considered hardware redundancy ($q_i > 1$). The ideas presented in this work can be easily extended to the case when more than one sensor can be placed on a variable.

3.1 Computation of Kullback-Leibler divergence

In the current work, maximum likelihood (ML) estimation procedure is proposed to obtain the estimate $\hat{\mathbf{x}}_{\mathbf{p}}$, and pdfs $f(\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}})$ and $g(\hat{\mathbf{x}}_{\mathbf{p}})$. For the case when the measurement noises are Gaussian, this leads to analytical forms for the reference $(f(\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}}))$ and the designed $(g(\hat{\mathbf{x}}_{\mathbf{p}}))$ pdfs as well as for the KLD of these pdfs. However, analytical forms are not available in general for non-Gaussian cases. These two scenarios are discussed below:

(I) *Gaussian measurement noise case*:

For the sake of simplicity, consider the uncorrelated Gaussian measurement noise case, i.e., $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_v})$ with $\mathbf{\Sigma_v}$ being a diagonal matrix. Consider the computation of the reference pdf $f(\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}})$ which corresponds to the measurement of all variables in the process. In this case, the estimate $(\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}})$ of primary variables can be obtained by solving the data reconciliation problem and is given as (Chmielewski et al., 2002):

$$\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}} = (\mathbf{A}^T[\mathbf{\Sigma_v^{ref}}]^{-1}\mathbf{A})^{-1}\mathbf{A}^T[\mathbf{\Sigma_v^{ref}}]^{-1}\mathbf{y} \quad (14)$$

where $[\mathbf{\Sigma_v^{ref}}]^{-1}$ is given in (8d) with $q_i = 1, \forall i \in \{1, \ldots, n\}$. The covariance matrix of the estimate $\hat{\mathbf{x}}_{\mathbf{p}}$ is:

$$\mathbf{\Sigma}_{\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}}} = (\mathbf{A}^T[\mathbf{\Sigma_v^{ref}}]^{-1}\mathbf{A})^{-1} \quad (15)$$

Further, the reference pdf $f(\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}})$ is Gaussian with mean $\mathbf{x_p}$ (the true values of the variables) and covariance matrix $\mathbf{\Sigma}_{\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}}}$ (referred as $\mathbf{\Sigma_f}$), i.e., $f(\hat{\mathbf{x}}_{\mathbf{p}}^{\mathbf{ref}}) = \mathcal{N}(\mathbf{x_p}, \mathbf{\Sigma_f})$. In a similar manner, the designed pdf $g(\hat{\mathbf{x}}_{\mathbf{p}})$ can be obtained to be $g(\hat{\mathbf{x}}_{\mathbf{p}}) = \mathcal{N}(\mathbf{x_p}, \mathbf{\Sigma_g})$, where $\mathbf{\Sigma_g} = \mathbf{\Sigma}_{\hat{\mathbf{x}}_{\mathbf{p}}}$ is the

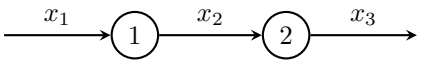


Fig. 1. An illustrative flow process

corresponding covariance matrix obtained in a manner similar to (15) as:

$$\mathbf{\Sigma_{\hat{x}_p}} = (\mathbf{A}^T {\mathbf{\Sigma_v}}^{-1} \mathbf{A})^{-1} \tag{16}$$

where ${\mathbf{\Sigma_v}}^{-1}$ is as given in (8d) for the designed sensor network. Thus, both the reference and the designed pdfs are Gaussian with the same mean. Using (2), the KLD objective in (13) can then be analytically obtained to be:

$$D_{KL}(f(\hat{\mathbf{x}}_\mathbf{p}^\mathbf{ref})||g(\hat{\mathbf{x}}_\mathbf{p})) = \frac{1}{2}\left[tr\left(\mathbf{\Sigma_f}{\mathbf{\Sigma_g}}^{-1}\right) + \ln\left(\frac{det(\mathbf{\Sigma_g})}{det(\mathbf{\Sigma_f})}\right) - (n-m)\right] \tag{17}$$

The expression for KLD for Gaussian measurement noise as given in (17) is derived for the uncorrelated measurement noise case. When the measurement noises are Gaussian but correlated, (17) continues to hold. The only modification is that the two-step data reconciliation procedure discussed in Section 2.2 needs to be used to obtain the covariance matrices $\mathbf{\Sigma_f}$ and $\mathbf{\Sigma_g}$.

(II) *Non-Gaussian measurement noise case*:
For this case, analytical expressions for $f(\hat{\mathbf{x}}_\mathbf{p}^\mathbf{ref})$ and $g(\hat{\mathbf{x}}_\mathbf{p})$ cannot be obtained in general. In this work, we have used a Monte Carlo based simulation approach to obtain these pdfs. In particular, a large number (N) of realizations of ML estimates are obtained corresponding to different measurement noise realizations. To obtain the pdfs from this sample, a Gaussian sum representation with the individual Gaussian pdfs centered at the N realizations of the ML estimates is obtained. KLD is then computed by numerically integrating the integral in (1), where reference and designed pdfs are represented as sum of Gaussian pdfs. Details are given in Appendix A.

3.2 Illustrative example for sensor network design

Consider a flow process (Fig. 1) with x_i corresponding to the mass flow rate of stream i, $i = 1, 2, 3$. For this process, the process model constraint matrix $\mathbf{C}$ in (4) can be written as: $\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$. With choice of x_1 as the primary variable (x_p), the $\mathbf{C_s}$, $\mathbf{C_p}$ matrices (in (5)), and the measurement matrix $\mathbf{A}$ (in (7)) are: $\mathbf{C_p} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{C_S} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix}$, $\mathbf{A} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Let the true value of $x_1 = 100$ kg/s. Also consider that the measurement noises have Gaussian mixture model (GMM) as:

$$v_1 \sim 0.6 \times \mathcal{N}(2, 0.2687^2) + 0.4 \times \mathcal{N}(-3, 0.4663^2)$$
$$v_2 \sim 0.7 \times \mathcal{N}(3, 0.4169^2) + 0.3 \times \mathcal{N}(-7, 0.4838^2)$$
$$v_3 \sim 0.8 \times \mathcal{N}(1, 0.3623^2) + 0.2 \times \mathcal{N}(-4, 0.1143^2)$$

The reference pdf $f(\hat{x}_1^{ref})$ of the estimate of x_1 corresponds to the case when all the variables are measured and is obtained using the procedure for non-Gaussian measurement noise case discussed in Section 3.1. To illustrate, different designed pdfs $g(\hat{x}_1)$ corresponding to various sensor networks are also obtained. Fig. 2 depicts $f(\hat{x}_1^{ref})$ and $g(\hat{x}_1)$ for various sensor networks. From this figure, it can be seen that the designed pdfs are sum of Gaussian pdfs. Further, Table 1 lists the corresponding KLD value (column 2). It can be observed that $\{y_3\}$ and $\{y_2, y_3\}$ are the optimal sensor networks when $N^* = 1$ and 2, respectively. Column 3 in Table 1 also lists the 95% confidence interval (CI) on error in estimate of x_1 obtained from $g(\hat{x}_1)$ for various sensor networks as depicted in Fig. 2. To compare, we also compute the CI on the error for the case when the measurement noise is assumed to be Gaussian. Towards this end, the GMM of the measurement noise is approximated as (Trailovic and Pao, 2002):

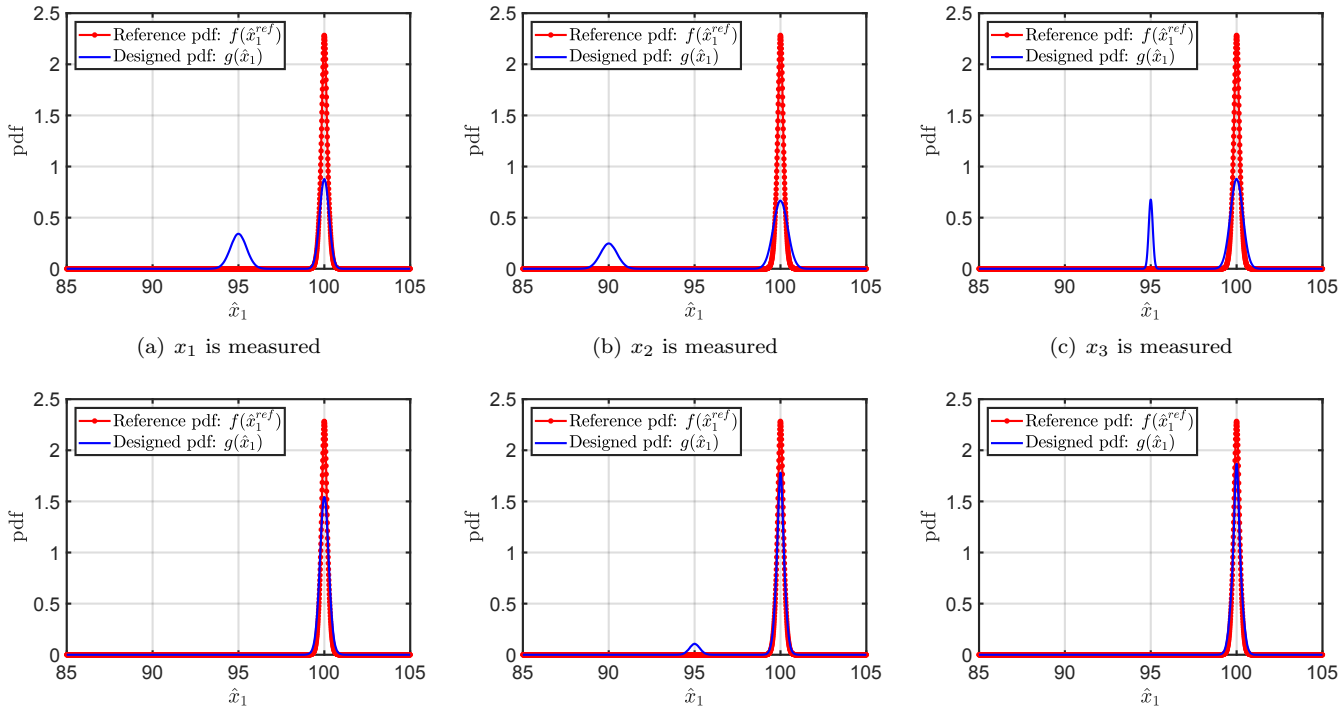
$$\text{Given, } v_i \sim \sum_l w_l \, \mathcal{N}(\mu_l, \sigma_l^2), \;\; \tilde{v}_i \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2) \tag{18}$$

$$\text{with } \tilde{\mu} = \sum_l w_l \mu_l, \;\; \tilde{\sigma}^2 = \sum_l w_l \left(\sigma_l^2 + \mu_l^2\right) - (\tilde{\mu})^2 \tag{19}$$

where $\tilde{v}_i$ is the Gaussian approximation of the GMM of v_i. With this Gaussian approximation, $f(\hat{x}_1^{ref})$ and $g(\hat{x}_1)$ would also be Gaussian (Section 3.1). The resulting 95% CI on errors in estimate of x_1 are also reported in Table 1 (column 4). Comparing columns 3 and 4, it can be seen that the confidence band on error in estimate of x_1 is tighter if noise is represented by its original pdf (GMM in this case), instead of approximating it with a Gaussian pdf. To further check the correctness of the confidence intervals obtained with a Gaussian approximation, the area (confidence level) under the sum of Gaussian pdfs $g(\hat{x}_1)$ with the limits obtained from the confidence intervals obtained with Gaussian approximation is also computed. This area is listed for various sensor networks in column 5 in Table 1. It can be seen that the obtained confidence level is either significantly lower or higher than the designed value of 95% thereby indicating the inadequacy of the Gaussian approximation. This example thus illustrates the need of explicitly incorporating the non-Gaussian nature of measurement noises in the sensor network design procedure.

4. CASE STUDY

We now present an Ammonia process case study (Ali and Narasimhan, 1993; Nabil and Narasimhan, 2012; Kotecha et al., 2008). The process graph (Fig. 3) consists of 6 nodes (including 1 environmental node) and 8 streams (variables x). Independent mass balances can be written across any 5 nodes. Then the process model constraint matrix $\mathbf{C}$ in (4) is written in (20). For this process, a minimum of three measurements are required to satisfy the observability criteria. Thus, considering primary variables as: $\mathbf{x_p} = \{x_1, x_4, x_6\}$, the measurement matrix $\mathbf{A}$ in (7) is given in (20). The measurement noises are considered to be Gaussian with randomly selected variances as: {0.2687, 0.4663, 0.4169, 0.4838, 0.3623, 0.1143, 0.4397, 0.4736}. Optimal SND (solution of Formulation II) is now obtained for N^* ranging from 3-8 and listed in Table 2. For the sake of comparison, we also list the optimal sensor network design obtained with A-, D-, T-, and E- optimal design approaches (Formulation I). All the optimization problems (Formulations I and II) were solved by enumeration.



(a) x_1 is measured (b) x_2 is measured (c) x_3 is measured

(d) x_1 and x_2 are measured (e) x_1 and x_3 are measured (f) x_2 and x_3 are measured

Fig. 2. Probability density function for various sensor networks

$$\mathbf{C} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & -1 \end{bmatrix}, \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \quad (20)$$

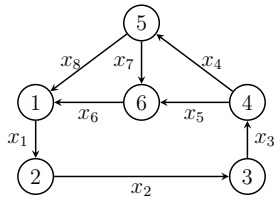


Fig. 3. Simplified Ammonia flowsheet

From the results in Table 2, the following can be observed:
1) The various existing design approaches lead to different optimal designs in general. Further, the optimal designs obtained by these approaches are in general different from the optimal designs obtained by the proposed KLD minimization criteria.
2) As N^* increases, the objective functions in all the design approaches decrease. However, only for the proposed KLD objective function, does the objective value decreases to 0 for the case when all the variables in the process are measured. This happens since for this case, the designed pdf exactly matches the reference pdf. For other values of N^*, the KLD value indicates the distance of the designed pdf from the reference pdf. For the existing design approaches, the objective function does not go to 0 even for the case when all variables are measured. Hence, it may be difficult for the end-user to interpret the objective function values obtained by the existing design approaches vis-a-vis her requirements.

5. CONCLUSIONS

In this work, we have considered the sensor network design problem for data reconciliation in a steady state linear process. In particular, we presented a novel sensor network design formulation based on the information theoretic concept of Kullback-Leibler divergence. The advantages of this formulation are that it does not assume Gaussian

Table 1. KLD value and 95% CI on error in estimate of x_1 for various sensor networks

Sensor network	D_{KL}	GMM: [CI interval] (CI width)	Gaussian: [CI interval] (CI width)	Area under GMM assuming CI interval from Gaussian case
$\{y_1\}$	0.6053	$[-5.72, 0.472]$ (6.1920)	$[-4.8528, 4.8528]$ (9.7057)	0.7514
$\{y_2\}$	0.7328	$[-10.68, 0.755]$ (11.4350)	$[-9.0228, 9.0228]$ (18.0456)	0.7068
$\{y_3\}$	0.4921	$[-5.135, 0.68]$ (5.8150)	$[-3.9724, 3.9724]$ (7.9448)	0.8003
$\{y_1, y_2\}$	0.0772	$[-0.53, 0.535]$ (1.0650)	$[-4.2739, 4.2739]$ (8.5478)	1
$\{y_1, y_3\}$	0.1060	$[-5.14, 0.46]$ (5.60)	$[-3.0739, 3.0739]$ (6.1477)	0.9197
$\{y_2, y_3\}$	0.0464	$[-0.52, 0.52]$ (1.040)	$[-3.6356, 3.6356]$ (7.2713)	1
$\{y_1, y_2, y_3\}$	0	$[-0.395, 0.395]$ (0.79)	$[-2.9097, 2.9097]$ (5.8193)	1

Table 2. Comparative SND results for Ammonia process

N^*	D_{KL}	A_{op}	D_{op}	T_{op}	E_{op}
3	$\{y_1, y_5, y_6\}$ (0.45881)	$\{y_1, y_4, y_6\}$ (0.86680)	$\{y_1, y_5, y_6\}$ (0.01113)	$\{y_6, y_7, y_8\}$ (−19.79471)	$\{y_1, y_4, y_6\}^a$ (−2.06697)
4	$\{y_1, y_4, y_6, y_7\}$ (0.23291)	$\{y_1, y_4, y_5, y_6\}$ (0.59208)	$\{y_1, y_4, y_5, y_6\}$ (0.00483)	$\{y_5, y_6, y_7, y_8\}$ (−25.31500)	$\{y_1, y_4, y_5, y_6\}$ (−2.77297)
5	$\{y_1, y_4, y_5, y_6, y_8\}$ (0.11774)	$\{y_1, y_3, y_4, y_5, y_6\}$ (0.50318)	$\{y_1, y_4, y_5, y_6, y_7\}$ (0.00266)	$\{y_1, y_5, y_6, y_7, y_8\}$ (−29.03662)	$\{y_1, y_3, y_4, y_5, y_6\}$ (−3.42948)
6	$\{y_1, y_3, y_4, y_5, y_6, y_7\}$ (0.05452)	$\{y_1, y_3, y_4, y_5, y_6, y_7\}$ (0.44385)	$\{y_1, y_4, y_5, y_6, y_7, y_8\}$ (0.00170)	$\{y_1, y_3, y_5, y_6, y_7, y_8\}$ (−31.43527)	$\{y_1, y_2, y_3, y_4, y_5, y_6\}$ (−3.77613)
7	$\{y_1, y_3, y_4, y_5, y_6, y_7, y_8\}$ (0.01049)	$\{y_1, y_3, y_4, y_5, y_6, y_7, y_8\}$ (0.39011)	$\{y_1, y_3, y_4, y_5, y_6, y_7, y_8\}$ (0.00125)	$\{y_1, y_2, y_3, y_5, y_6, y_7, y_8\}$ (−33.57982)	$\{y_1, y_3, y_4, y_5, y_6, y_7, y_8\}$ (−4.27694)
8	$\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8\}$ (0)	$\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8\}$ (0.35945)	$\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8\}$ (0.00101)	$\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8\}$ (−35.64679)	$\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8\}$ (−4.72331)

[a] The E-optimal design approach for $N^* = 3$ yields three different optimal sensor networks.
The other two sensor networks are $\{y_2, y_4, y_6\}$ and $\{y_3, y_4, y_6\}$.

measurement noises and can directly incorporate non-Gaussian nature of the measurement noises. Additionally, it enables the end-user to set a target for the sensor network design performance by specifying a reference probability density function of the estimates. The significance of the proposed KLD based sensor network design approach was illustrated using a small example and a literature case study. Development of efficient optimization methods to solve larger problems with hundreds of variables is currently under investigation.

REFERENCES

Ali, Y. and Narasimhan, S. (1993). Sensor network design for maximizing reliability of linear processes. *AIChE Journal*, 39(5), 820–828.

Balaji, B.S., Bhatt, N., and Narasimhan, S. (2018). Optimal selection of reference components and measurements in reaction systems. *Industrial & Engineering Chemistry Research*, 57(44), 15096–15104.

Chmielewski, D.J., Palmer, T., and Manousiouthakis, V. (2002). On the theory of optimal sensor placement. *AIChE Journal*, 48(5), 1001–1012.

Hershey, J.R. and Olsen, P.A. (2007). Approximating the Kullback Leibler divergence between Gaussian mixture models. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 4, IV–317. IEEE.

Joshi, S. and Boyd, S. (2008). Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, 57(2), 451–462.

Jusoh, R.M. and Ampountolas, K. (2019). Optimal selection of traffic sensors: An information-theoretic framework. In *2019 American Control Conference (ACC)*, 3297–3302. IEEE.

Kotecha, P.R., Bhushan, M., and Gudi, R.D. (2008). Design of robust, reliable sensor networks using constraint programming. *Computers & Chemical Engineering*, 32(9), 2030–2049.

Krause, A., Singh, A., and Guestrin, C. (2008). Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(Feb), 235–284.

Kretsovalis, A. and Mah, R.S. (1987). Effect of redundancy on estimation accuracy in process data reconciliation. *Chemical Engineering Science*, 42(9), 2115–2121.

Nabil, M. and Narasimhan, S. (2012). Sensor network design for optimal process operation based on data reconciliation. *Industrial & Engineering Chemistry Research*, 51(19), 6789–6797.

Narasimhan, S. and Jordache, C. (1999). *Data reconciliation and gross error detection: An intelligent use of process data*. Elsevier.

Trailovic, L. and Pao, L.Y. (2002). Variance estimation and ranking of Gaussian mixture distributions in target tracking applications. In *Proceedings of the 41st IEEE Conference on Decision and Control, 2002.*, volume 2, 2195–2201. IEEE.

Appendix A. KLD COMPUTATION FOR NON-GAUSSIAN MEASUREMENT NOISE

When the measurement noises $v_i, i = 1, 2, \ldots, n$ are independent random variable, the likelihood function of $\mathbf{x}_p$ can be written as: $\mathcal{L}(\mathbf{x_p}|\boldsymbol{\rho}) = \prod_{i \in M} f_{y_i}(\rho_i|\mathbf{x_p})$, where ρ_i is the realized value of the i^{th} measurement y_i, and set $M = \{j : q_j = 1\}$ contains the indices of measured variables. Using (7), we get: $\mathcal{L}(\mathbf{x_p}|\boldsymbol{\rho}) = \prod_{i \in M} f_{v_i}(\rho_i - [\mathbf{A}^{(i)}]^T \mathbf{x_p})$, where $[\mathbf{A}^{(i)}]^T$ is the i^{th} row of matrix $\mathbf{A}$. The maximum likelihood estimate $\hat{\mathbf{x}}_\mathbf{p}$ is obtained by maximizing the likelihood function as:

$$\hat{\mathbf{x}}_\mathbf{p} = \arg\max_{\mathbf{x_p}} \mathcal{L}(\mathbf{x_p}|\boldsymbol{\rho}) \tag{A.1}$$

For non-Gaussian measurement noises $\mathbf{v}$, (A.1) does not have an analytical solution in general, and numerical procedures have to be used. In the current work, the pdf of $\hat{\mathbf{x}}_\mathbf{p}$ is obtained by performing Monte Carlo simulations. In this approach, for a given $\mathbf{x_p}$ (the true value of variables in the process), large number (N) of independent and identically distributed samples $\rho_i^{(k)}$ from $f_{y_i}(\rho_i|\mathbf{x_p}), \forall i \in M$ are obtained, where $k = 1, 2, \ldots, N$ indicates the sample number. For each of these realizations, (A.1) is solved to obtain the corresponding realization of the maximum likelihood estimator $\hat{\mathbf{x}}_\mathbf{p}^{(k)}$. The pdf for the maximum likelihood estimator can then be obtained by using a kernel density estimation procedure with a Gaussian kernel. This overall procedure is used to obtain the reference pdf ($f(\hat{\mathbf{x}}_\mathbf{p}^\mathbf{ref})$) as well. Once the pdfs $f(\hat{\mathbf{x}}_\mathbf{p}^\mathbf{ref})$ and $g(\hat{\mathbf{x}}_\mathbf{p})$ are obtained, KLD in (1) is computed by numerical integration.

For the example presented in Section 3.2, Gaussian mixture models for measurement noises were considered with $x_p = 100$. $N = 2 \times 10^6$ realizations were considered in the Monte Carlo based approach. For a given realization, grid search was used to solve (A.1) to obtain the corresponding realization of the maximum likelihood estimator. The pdf of the estimator was obtained in MATLAB version 2018a using its inbuilt *ksdensity* function with the bandwidth parameter chosen to be the optimal for Gaussian densities. KLD in (1) was computed by numerically integrating the integral in the range [80,120] using trapezoidal rule with 40,000 equal width intervals in this range.