Convex Optimization Approach to Design Sensor Networks using Information Theoretic Measures

Arjun M, and Nabil Magbool Jan*

Department of Chemical Engineering, Indian Institute of Technology Tirupati, Andhra Pradesh, 517619, India

Abstract

Accurate and precise estimation of process variables is key to effective process monitoring. The estimation accuracy depends on the choice of the sensor network. Therefore, this paper aims at developing convex optimization formulations for designing the optimal sensor network using information-theoretic measures in linear steady-state data reconciliation. To this end, the estimation errors are characterized by a multivariate Gaussian distribution, and thus the analytical form for entropy and Kullback-Leibler divergences (forward, reverse, and symmetric) of estimation errors can be obtained to formulate the optimal sensor network design. The proposed information theoretic-based optimal sensor selection problems are shown to be integer semidefinite programming problems where the relaxation of binary decision variables results in solving a convex optimization problem. Thus, we use a branch and bound method to obtain a globally optimal sensor network design. Demonstrative case studies are presented to illustrate the efficacy of the proposed optimal sensor selection formulations.

Keywords: data reconciliation, entropy, Kullback-Leibler divergence, sensor network design

*Corresponding author, Email: nabil@iittp.ac.in

1 Introduction

The selection of sensors is vital for efficiently performing process monitoring, fault detection and diagnosis, and process control. In recent years, sensor placement has played a significant role in the industrial internet of things. A typical process plant has many process variables; however, the number of sensors that can be placed is limited due to sensor cost or inability to place a sensor in the desired location, etc. Further, the installed sensors should enable one to infer the unmeasured variables using the process model. Therefore, the sensor selection problem involves selecting a subset of the available sensors to be used for monitoring the process variables. The goal is to determine the best combination of sensors to achieve the desired level of accuracy in the measurement of the process variables. This problem can be formulated as an optimal design problem where the goal is to optimize a utility function based on one or more of the following criteria: (a) minimizing the average estimation error that signifies the ability to obtain accurate estimates of process variables using data reconciliation,[1–5] (b) minimizing the operational cost or hardware cost of a plant,[6–10] (c) maximizing network reliability, or the sensor networks' ability to estimate process variables even in the situation where one or more sensors fail.[11–16] In addition, observability is often enforced such that the sensor networks can be used to estimate unmeasured variables using measurements and the process model. This work focuses on optimal sensor selection based on information-theoretic measures for the purpose of data reconciliation.

Vaclaveck and Loucka[17] developed a sensor network design approach for steady-state systems using graph theory to assure the observability of critical process variables. From a fault detection and diagnosis aspect, a diagnostic system's effectiveness is determined by a reliable set of sensors. Thus, the reliability of sensor networks is often used as a metric for the sensor selection problem. Ali and Narasimhan[11–13] presented a cutset-based graph theoretic approach to decide on measurement selection by maximizing reliability. Kotecha et al.[14] defined the reliability of steady-state linear flow processes as described by Ali and Narasimhan.[11] They also defined network variance, which measures uncertainty in estimates, as the highest estimation variance among all variables. They established that there is a dual relationship between maximizing reliability and minimizing variance. On the other hand, Bhushan et al.[18] defined unreliability as the probability of a fault occurring and remaining undetected and developed an optimization methodology to minimize the unreliability of detecting faults in the process.

From a data reconciliation perspective, an appropriate selection of sensors is crucial for improving the accuracy of the estimation of process variables. One of the earlier works on sensor placement for a steady-state, linear process using the data reconciliation aspect was proposed by Kretsovalis and Mah.[1] The trace of the covariance matrix of estimations of variables was minimized to improve the accuracy of process variable estimation, such that the obtained sensor network forms an observable set. Alternately, with growing global competition, maximizing the economic performance of process plants is of the utmost importance, leading to the adoption of capital cost as a vital metric for optimization. In this regard, Bagajewicz[6] proposed a sensor placement strategy to minimize the capital cost subject to precision bounds for reconciled values and robustness to gross errors. The optimization problem formulated by Bagajwicz was a Mixed Integer Non-Linear Programming (MINLP) problem. Bagajwicz and Sanchez[8] later considered an optimization problem that involved minimizing the total sensor network cost subject to the reliability of estimating a process variable. Following this, Chmielewski et al.[10] proposed a method to convert the non-linear constraints from the MINLP problem proposed by Bagajwicz into Linear Matrix Inequality (LMI) constraints. This resulted in a convex program that yield a globally optimal sensor network. Nabil and Narasimhan[2] developed a sensor network design formulation that incorporated the notion of process economics by defining a loss in operational profit, and it was shown to minimize the trace of the weighted error covariance matrix. Furthermore, Balaji et al.[3] applied the A-optimal sensor selection criterion in a reaction system with a linear measurement model.

Various information-theoretic measures, like entropy, correntropy, and mutual information entropy have been used for data reconciliation. Entropy is a measure of uncertainty or randomness in a system. Crowe[19] maximized entropy to determine the conditional distribution of data given true values with minimal incorporation of prior knowledge, and then the maximum likelihood of the resulting density function is used to obtain the reconciled estimates. Further, it was shown that, for the case when the bounds and sensor error variances are assumed to be known a priori, the developed entropy-based approach yields a well-known weighted least square solution. Correntropy is a measure of the similarity between two data sets and is defined as the expected value of the squared difference between the two data sets.[20] Further, Yang et al.[21] used mutual information for nonlinear data reconciliation to deal with correlations between nonlinearly related variables.

In the context of sensor selection, information-theoretic measures have also been used for sensor selection. The entropy of an estimated process variable quantifies the uncertainty of the estimate, incorporating the distribution of errors, which can be used to determine the most optimal sensors for the process. One illustration of the utilization of entropy in sensor selection can be observed in localization problems.[22] The algorithm proposed in this study selected the sensor network that resulted in the greatest reduction in entropy, based on the target location distribution and sensor locations. Another instance of the application of entropy in sensor selection can be found in the field of structural damage detection.[23] The authors employed entropy as a metric to reduce the number of required sensors while maintaining the accuracy of the damage detection model. These examples highlight the potential of entropy as a valuable metric in the selection of optimal sensors. The process of selecting the optimal set of sensors can be facilitated through the application of mutual information. One such application considers mutual information as a metric for evaluating sensor performance and is aimed at identifying the sensor network that provides the highest level of accuracy in tracking a specified number of objects.[24] By maximizing mutual information, the optimal sensor network is obtained, which is capable of effectively tracking the maximum number of objects while minimizing the use of sensors. Mutual information has also been used in power transmission applications.[25] By utilizing mutual information, a sensor network can be selected that provides the most comprehensive information about the target state, thereby optimizing data transfer and reducing waste. Very recently, Prakash and Bhushan[26] recently presented a reverse Kullback-Leibler (KL) divergence for designing sensor networks from a data reconciliation perspective for non-Gaussian noise distributions. Unlike the previous work on sensor selection using reverse KL divergence,[26] we assume that the estimation errors can be characterized using multivariate Gaussian distributions. Moreover, we use symmetric KL divergence to quantify the process variables' estimation accuracy in a linear steady-state data reconciliation problem. This is in contrast to asymmetric reverse KL divergence used by Prakash and Bhushan.[26] Further, they obtained the optimal sensor network through enumeration.

The main contributions of this work are as follows: (a) proposed information-theoretic measures such as entropy, forward, reverse, and symmetric KL divergences for optimal sensor selection for linear data reconciliation, (b) developed a computationally tractable integer semidefinite programming formulations for minimization of entropy of estimation errors, and minimization of reverse KL divergence, (c) proposed convex reformulations for the minimization of forward as well as symmetric KL divergences, and developed computationally tractable mixed-integer semidefinite programming formulations, (d) a globally optimal sensor network can be obtained using a branch and bound technique in which a semidefinite program is solved at each branching node of all the proposed formulations, and (e) demonstrated the efficacy the proposed approach using benchmark examples.

The rest of the paper is organized as follows. Firstly, we present the definition and properties of entropy and different variants of KL divergence. Secondly, we discuss the different formulations of data reconciliation. Thirdly, we propose an information-theoretic framework for optimal sensor selection using entropy, forward, reverse, and symmetric KL divergences, and developed computationally tractable optimization formulations. Next, three demonstrative case studies are presented. Finally, conclusions are presented.

2 Preliminaries

In this section, we briefly discuss the definitions and properties of information-theoretic measures such as entropy and KL divergences.

2.1 Entropy

In information theory, entropy is a measure of the uncertainty of a random variable. It measures the amount of information that is contained in a signal or in a set of data.[27] In general, the more uncertainty or randomness there is in a set of data, the higher its entropy will be. A set of data that is highly predictable will have a low entropy, while a set of data that is completely random will have a high entropy. There are several ways to measure entropy, but the most common is the Shannon entropy,[27] which is defined as:

$$\mathcal{H}(z) = -\int_S p(z) \log p(z) dz \tag{1}$$

where $\mathcal{H}(z)$ is the entropy of the random variable z, $S = \{z \,|\, p(z) > 0\}$ is the support set of z, and $p(z)$ is the probability density function. $\mathcal{H}(z)$ is always non-negative i.e., $\mathcal{H}(z) \geq 0$. Also to be noticed is the fact that the maximum value of Shannon entropy occurs when all events in the system are equally likely. Since the entropy is highest, it indicates maximum uncertainty or randomness. Another important property of Shannon's entropy is that Shannon's entropy is monotonically increasing with respect to the probabilities of events of a system. This property enables the comparison of the uncertainty of different systems. If Shannon's entropy is greater for one system than another, it can be concluded that the first system is more random or uncertain than the other. For a multivariate random variable $z \in \mathbb{R}^{n_z}$ following Gaussian distribution, $z \sim \mathcal{N}(\mu, \Sigma)$, the Shannon's entropy is defined as,[28]

$$\mathcal{H}(z) = \frac{n_z}{2}(1 + \log(2\pi)) + \frac{1}{2} \log |\Sigma| \tag{2}$$

where n_z is the dimensionality of z, and $|\cdot|$ denotes the determinant operation.

2.2 Kullback - Leibler Divergence

For two arbitrary probability density functions (PDF), $f(z)$ and $g(z)$, of a multivariate random variable $z \in \mathbb{R}^{n_z}$, Kullback-Leibler divergence is defined as,[28]

$$D_{KL}(f||g) = \int_{-\infty}^{\infty} f(z) \log\left(\frac{f(z)}{g(z)}\right) \, dz \tag{3}$$

It should be noted that $D_{KL}(f||g)$ is a non-negative (i.e., $D_{KL}(f||g) \geq 0$), asymmetric measure (i.e., $D_{KL}(f||g) \neq D_{KL}(g||f)$), and it does not satisfy the triangular inequality property[28]. KL divergence enables one to compare two distributions. For the case of two multivariate Gaussian distributions $f \sim \mathcal{N}(\mu_f, \Sigma_f)$ and $g \sim \mathcal{N}(\mu_g, \Sigma_g)$, the forward KL divergence, denoted by $\mathcal{D}_{KL}^f$, is defined as,

$$\begin{aligned} \mathcal{D}_{KL}^f := D_{KL}(f||g) =& \frac{1}{2}\Big[Tr(\Sigma_f {\Sigma_g}^{-1}) + \log\left(\frac{|\Sigma_g|}{|\Sigma_f|}\right) \\ &+ (\mu_g - \mu_f)^T {\Sigma_g}^{-1}(\mu_g - \mu_f) - n_z\Big] \end{aligned} \tag{4}$$

where $Tr(\cdot)$ and $|\cdot|$ denote the trace and determinant operations, respectively. Now, the reverse KL divergence, denoted by $\mathcal{D}_{KL}^r$, is defined as,

$$\begin{aligned} \mathcal{D}_{KL}^r := D_{KL}(g||f) =& \frac{1}{2}\Big[Tr(\Sigma_g {\Sigma_f}^{-1}) + \log\left(\frac{|\Sigma_f|}{|\Sigma_g|}\right) \\ &+ (\mu_f - \mu_g)^T {\Sigma_f}^{-1}(\mu_f - \mu_g) - n_z\Big] \end{aligned} \tag{5}$$

It should be noted that in the case of forward KL divergence, the logarithmic difference between $f(z)$ and $g(z)$ is weighted by $g(z)$, whereas in reverse KL divergence, it is weighted by $f(z)$. It is important to recall that $\mathcal{D}_{KL}^f \neq \mathcal{D}_{KL}^r$. Symmetric KL divergence ($\mathcal{D}_{KL}^s$) is often used to alleviate the asymmetric issue of forward and reverse KL divergences, and it is expressed as follows,

$$
\begin{aligned}
\mathcal{D}^s_{KL}(f,g) = \mathcal{D}^s_{KL}(g,f) &= \frac{1}{2}\big(\mathcal{D}^f_{KL} + \mathcal{D}^r_{KL}\big) \\
&= \frac{1}{4}\big[Tr(\Sigma_f {\Sigma_g}^{-1}) + Tr(\Sigma_g {\Sigma_f}^{-1}) \\
&\quad + (\mu_g - \mu_f)^T \big[{\Sigma_g}^{-1} + \Sigma_f^{-1}\big](\mu_g - \mu_f) - 2n_z\big]
\end{aligned} \tag{6}
$$

3 Data Reconciliation

Data reconciliation is a vital tool in modern process monitoring systems. It utilizes the process model and description of measurement errors to improve measured variables' accuracy and obtain estimates for unmeasured quantities.[29] Consider a measurement model of the form,

$$
y_m = x_m + \epsilon_m \tag{7}
$$

where $y_m \in \mathbb{R}^p$, $x_m \in \mathbb{R}^p$ and $\epsilon_m \in \mathbb{R}^p$ respectively signify the measurement vector, the true values of measured variables, and the vector of measurement noises. For steady-state linear processes, the model equations can be expressed as,

$$
Ax = 0 \tag{8}
$$

where $A \in \mathbb{R}^{r\times n}$ represents the process constraint matrix, r denotes the number of independent model equations, and $x \in \mathbb{R}^n$ represents the vector of all the process variables (both measured, x_m, and unmeasured variables, x_u), that is, $x = [x_m; x_u]^T$. In general, the formulation of a linear steady-state data reconciliation problem is as follows:

$$
\begin{aligned}
\min_{\hat{x}_m, \hat{x}_u} \quad & (y_m - \hat{x}_m)^T \Sigma_\epsilon^{-1} (y_m - \hat{x}_m) \\
s.t. \quad & A_m \hat{x}_m + A_u \hat{x}_u = 0
\end{aligned} \tag{9}
$$

where A_m and A_u are submatrices of A of appropriate dimensions, respectively. $\Sigma_\epsilon^{-1} = diag(1/\sigma_i^2)$ is a diagonal matrix whose elements are equal to the inverse of error variances of measurements. The solution to the above constrained optimization problem can be obtained using Lagrange's multipliers,

$$\hat{x}_m = y_m - \Sigma_\epsilon^{-1} A^T (A \Sigma_\epsilon^{-1} A^T)^{-1} A y_m \tag{10}$$

It should be noted that the performance of the data reconciliation depends critically on the set of measured variables, redundancy in measurements, and sensor precision.

Reformulated Data Reconciliation Problem

In this subsection, we recast the data reconciliation problem using the concept of primary variables following the works of Nabil and Narasimhan[2] and Chmielewski et al.[10] This allows one to conveniently express the data reconciliation in terms of potential measurements. The process variables of interest x can be categorized as primary variables, $x_p \in \mathbb{R}^{n-r}$, and secondary variables, $x_s \in \mathbb{R}^r$. This is in contrast to a typical data reconciliation problem, in which the process variables are categorized as measured, x_m, and unmeasured, x_u, variables. Any set of $n - r$ variables that guarantee network observability can be treated as primary variables, and the remaining variables correspond to the secondary variables. It is now possible to rewrite the process model constraints (8) in terms of primary and secondary variables as,

$$A_p x_p + A_s x_s = 0 \tag{11}$$

where A_p and A_s are submatrices of A of appropriate dimensions. Given that the primary variables form a minimum observable network, the matrix A_s will be invertible, and so the secondary variables are given by,

$$x_s = -{A_s}^{-1} A_p x_p \tag{12}$$

Defining $y \in \mathbb{R}^n$ as the measurement vector of all process variables, the measurement model can be written as follows,

$$y = Cx_p + \epsilon \tag{13}$$

where,

$$C = \begin{bmatrix} I \\ -{A_s}^{-1}A_p \end{bmatrix} \tag{14}$$

and I denotes identity matrix of size $n - r$. The data reconciliation problem is then posed as,

$$\min_{\hat{x}_p} \quad (y - C\hat{x}_p)^T Q(y - C\hat{x}_p) \tag{15}$$

where,

$$Q = \Sigma_\epsilon^{-1} = diag\{\frac{q_i}{{\sigma_i}^2}\} \tag{16}$$

Here q_i is a binary variable indicating if the particular variable i is measured or not, and ${\sigma_i}^2$ denotes the variance of measurement noise ϵ_i in the i^{th} measured variable. It should be noted that $1/{\sigma_i}^2$ denotes sensor precision, and it can be obtained from the vendor manual. Given the set of measured variables and their corresponding sensor precision values, the solution to the unconstrained optimization problem (15) above yields,

$$\hat{x}_p = (C^T QC)^{-1} C^T Qy \tag{17}$$

and the reconciled estimate of variables can be obtained as follows,

$$\hat{x} = C\hat{x}_p \tag{18}$$

It is essential to mention that the reconciled estimates obtained using equation (17) are the same as solving the optimization formulation (9). Equation (17) lets one quickly compute the reconciled estimates for different measurements by setting the appropriate $q_i = 1$. On the other hand, formulation (9) requires redefining the A_m and the A_u matrices for different measurements. Thus, equation (17) provides a convenient way to express reconciled estimates in terms of all potential measurements of interest. Furthermore, this helps one to obtain the analytical expression for the error covariance matrix Σ_e of the estimation error in $\hat{x}$. Defining the estimation error, $e = x - \hat{x}$, the error covariance matrix is given by,

$$\Sigma_e = C(C^T Q C)^{-1} C^T \tag{19}$$

Thus, the accuracy of reconciled estimates can be characterized using $e \sim \mathcal{N}(0, \Sigma_e)$, where zero mean indicates there are no gross errors. Notably, the selection of sensors significantly impacts the accuracy of reconciled estimates.

4 Optimal Sensor Selection using Information Theoretic Measures

In this section, we present an information-theoretic framework for sensor selection in data reconciliation.

4.1 Problem statement

In this work, we consider the problem of optimal sensor network design for linear, steady-state processes such that the most accurate reconciled estimates are obtained while performing data reconciliation if the determined optimal sensor network is employed. Therefore, the aim of this work is to determine the optimal set of N_s sensors among the N potential sensors such that the overall estimation error is minimized from an information-theoretic viewpoint.

It can be mathematically expressed as,

$$\begin{aligned} \min_{q_i \in \{0,1\}} \quad & \mathcal{I}(e) \\ s.t. \quad & \sum_i q_i = N_s \end{aligned} \tag{20}$$

where $\mathcal{I}(\cdot)$ denotes the chosen information-theoretic metric to quantify estimation errors of a sensor network for linear data reconciliation. More specifically, we propose the following information-theoretic measures for this purpose: entropy and variants of Kullback-Leibler (KL) divergence. These measures can be used to evaluate the quality of the reconciled estimate and also compare the accuracy of the estimate for different sensor network configurations. Therefore, we develop optimal sensor network design formulations based on these measures.

4.2 Entropy-based optimal sensor selection

In this subsection, we develop an optimization formulation based on the entropy of estimation errors in reconciled estimates. As discussed previously, the choice of sensor network has a significant impact on the overall estimation error. Thus, to evaluate the performance of different sensor networks, entropy can be used as a measure of estimation error, with lower entropy values indicating a lower estimation error. Therefore, minimizing entropy can be employed as an objective function for sensor network design optimization problems. In other words, our goal is to find the optimal sensor network that results in the minimum entropy of estimation errors.

Recall that the accuracy of reconciled estimates is characterized by estimation errors, $e \sim \mathcal{N}(0, \Sigma_e)$, and hence it is given by,

$$\mathcal{N}(0, \Sigma_e) = \frac{1}{\sqrt{(2\pi)^{n_z}|\Sigma_e|}} exp\big(-\frac{1}{2} e^T \Sigma_e^{-1} e \big) \tag{21}$$

where n_z is the dimensionality of the distribution. Thus, the entropy of estimation error is given by,

$$
\begin{aligned}
\mathcal{H}(e) &= -\,\mathbb{E}\left[\log \mathcal{N}(0, \Sigma_e)\right] \\
&= -\,\mathbb{E}\left[\log[(2\pi)^{-n_z/2}|\Sigma_e|^{-1/2} exp(-\frac{1}{2}e^T\Sigma_e^{-1}e)]\right] \\
&= \frac{n_z}{2}(1+\log(2\pi)) + \frac{1}{2}\log|\Sigma_e|
\end{aligned}
\tag{22}
$$

Assuming that gross errors do not exist and that the estimation errors depend only on the sensor network and its precision values. The estimation error of the sensor network can be modeled as a Gaussian distribution with mean zero and covariance matrix Σ_e, where $\Sigma_e = C(C^TQC)^{-1}C^T$ with $q_i = 1$ if $i \in SN$. Now, the entropy-based optimal sensor selection problem can be expressed as follows,

$$
\begin{aligned}
\min_{q_i \in \{0,1\}} \quad & \frac{n_z}{2}(1+\log(2\pi)) + \frac{1}{2}\log\det(C(C^TQC)^{-1}C^T) \\
s.t. \quad & \sum_i q_i = N_s
\end{aligned}
\tag{23}
$$

It is known that the error covariance matrix Σ_e may not be invertible due to degeneracy issues, as reported by Prakash and Bhushan.[26] This poses a challenge in computing the entropy of the estimation error distribution. To circumvent this degeneracy issue, we recast the optimization problem as minimizing the entropy of estimation errors in primary variables. Thus, the accuracy of reconciled estimates is characterized using the estimation errors in primary variables, $e_p \sim \mathcal{N}(0, \Sigma_{e,p})$ where $\Sigma_{e,p} = (C^TQC)^{-1}$. Now, the entropy-based optimal sensor selection problem is recast as follows,

$$
\begin{aligned}
\min_{q_i \in 0,1} \quad & \frac{n_p}{2}(1+\log(2\pi)) - \frac{1}{2}\log\det(C^TQC) \\
s.t. \quad & \sum_i q_i = N_s
\end{aligned}
\tag{24}
$$

where $Q = diag\{q_i/\sigma_i^2\}$, where q_i is a binary variable indicating if sensor i is selected or not, and n_p is the number of primary variables. The resulting optimization problem (24) is an integer nonlinear programming problem as the second term involves the computation of determinants involving binary decision variables. However, the integer relaxation of the binary decision variable yields $-log\ det(.)$ which is known to be a convex function in terms of continuous variables. Thus, the relaxed problem is convex, therefore, it can be solved for global optimality. Hence, the integer programming problem can be solved to global optimality using the branch and bound algorithm by solving the relaxed problem at every branching step.

4.3 Kullback-Leibler divergence based optimal sensor selection

In this subsection, we propose the use of the Kullback-Leibler (KL) divergence as a metric for optimal sensor selection. As it is widely known that KL divergence can be used as a similarity (or dissimilarity) measure between two distributions, say f and g. Therefore, it can be used to compute the similarity between the estimation error distributions obtained by two different sensor networks. It is well-known that the most accurate values of process variables can be obtained through data reconciliation when all variables are measured and no gross errors exist in the measurements. Let g represent the distribution of estimation errors obtained when all sensors are chosen, and it is assumed to follow a normal distribution with mean 0 and covariance $\Sigma_g = C(C^T\bar{Q}C)^{-1}C^T$, where $\bar{Q}$ represents the diagonal Q matrix for the case where all sensors are chosen. Similarly, let f represent the distribution of estimation errors when a sensor network (denoted by SN) is chosen. The distribution f follows a normal distribution with mean 0 and covariance $\Sigma_f = C(C^TQC)^{-1}C^T$ with $q_i = 1$ if $i \in SN$.

4.3.1 Sensor Network Design using Forward Kullback-Leibler divergence

As mentioned in the preliminaries section, there exist two variants of KL divergence viz. forward and reverse KL divergences. The optimal sensor selection problem based on forward

KL divergence can be formulated as follows,

$$\begin{aligned} \min \quad & \mathcal{D}_{KL}^{f}(f, g) \\ s.t. \quad & \sum_i q_i = N_s \end{aligned} \tag{25}$$

In light of the degeneracy issues highlighted by Prakash and Bhushan,[26] we rewrite the formulation in terms of the estimation errors in primary variables. The variance-covariance matrix of estimation errors in primary variables is given by:

$$\Sigma_{x_p} = (C^T Q C)^{-1} \tag{26}$$

Let $g_p \sim \mathcal{N}(0, \Sigma_{gp})$ represent the distribution of estimation errors in primary variables when all variables are measured, and $f_p \sim \mathcal{N}(0, \Sigma_{fp})$ represent the distribution of estimation errors in primary variables for a given sensor network SN. $\Sigma_{gp} = (C^T \bar{Q} C)^{-1}$ when all sensors are selected, and $\Sigma_{fp} = (C^T Q C)^{-1}$ when $q_i = 1$ if $i \in SN$. The formulation thus becomes,

$$\begin{aligned} \min \quad & \mathcal{D}_{KL}^{f}(f_p, g_p) \\ s.t. \quad & \sum_i q_i = N_s \end{aligned} \tag{27}$$

Upon substitution of Σ_{fp} and Σ_{gp} in the expressions for forward KL (4), we get

$$\begin{aligned} \min \quad & \frac{1}{2}\big[Tr\{(C^T Q C)^{-1}(C^T \bar{Q} C)\} + \log\det(C^T \bar{Q} C)^{-1} \\ & - \log\det(C^T Q C)^{-1} - n_p\big] \\ s.t. \quad & \sum_i q_i = N_s \end{aligned} \tag{28}$$

The resulting formulation is an integer nonlinear programming problem that is hard to solve in general. Therefore, in the following subsection, we present a convex reformulation to obtain a computationally tractable optimization problem.

Convex Reformulation of Forward Kullback-Leibler divergence

In this subsection, we reformulate the proposed optimization problem (28) using convex optimization. To this end, we provide some basic definitions enabling the reformulation. More details on these concepts can be found in the original reference.[30]

Definition 1: An epigraph of a function $f : \mathbb{R}^t \mapsto \mathbb{R}$ is the subset of $\mathbb{R}^{t+1}$ given by,

$$epi(f) = \{(u, v)|\ x \in domain(f), f(u) \leq v\}$$

Definition 2: Let G be positive definite i.e. $G \succ 0$, then the matrix $T = E - FG^{-1}F^T$ is the Schur complement of G in the matrix

$$X = \begin{bmatrix} E & F \\ F^T & G \end{bmatrix}$$

Then the condition for positive definiteness of block matrix X is if $G \succ 0$, then $X \succ 0$ only if $T \succ 0$.

Using the above definitions, and letting $W_3 = (C^T\bar{Q}C)^{-1}$ and $W_4 = (C^T\bar{Q}C)$, we can reformulate the original problem (32) as,

$$\begin{aligned} \min \quad & \frac{1}{2}\left[Tr\{YW_4\} + \log\det(W_3) - \log\det(Y) - n_p\right] \\ s.t. \quad & Y \succ (C^TQC)^{-1} \\ & \sum_i q_i = N_s \end{aligned} \tag{29}$$

This can be written as,

$$
\begin{aligned}
\min \quad & \frac{1}{2}\big[Tr\{YW_4\} + \log\det(W_3) - \log\det(Y) - n_p\big] \\
s.t. \quad & \begin{bmatrix} Y & I \\ I & C^TQC \end{bmatrix} \succ 0 \\
& \sum_i q_i = N_s
\end{aligned} \tag{30}
$$

Notice that the second and fourth terms in the equation (30) are known constants, and thus do not affect the optimal solution. The resulting formulation is a Mixed Integer SemiDefinite Programming (MISDP) problem, which when relaxed becomes a semidefinite programming problem, which is known to be convex. A branch and bound algorithm can be used to find the globally optimal solution to this problem.

4.3.2 Sensor Network Design using Reverse Kullback-Leibler divergence

The optimal sensor selection problem based on reverse KL divergence can be formulated as follows,

$$
\begin{aligned}
\min \quad & \mathcal{D}^r_{KL}(f, g) \\
s.t. \quad & \sum_i q_i = N_s
\end{aligned} \tag{31}
$$

As discussed above, owing to the degeneracy issues, the optimization formulation resembles the formulation for reverse KL divergence and is given by,

$$
\begin{aligned}
\min \quad & \mathcal{D}^r_{KL}(f_p, g_p) \\
s.t. \quad & \sum_i q_i = N_s
\end{aligned} \tag{32}
$$

Upon substitution of Σ_{fp} and Σ_{gp} in the expressions for reverse KL (5), we get

$$\begin{aligned} \min \quad & \frac{1}{2}\big[Tr(\{C^T\bar{Q}C)^{-1}(C^TQC)\} - \log\det(C^T\bar{Q}C)^{-1} \\ & + \log\det(C^TQC)^{-1} - n_p\big] \\ s.t. \quad & \sum_i q_i = N_s \end{aligned} \tag{33}$$

The resulting optimization formulation for reverse KL divergence is an Integer SemiDefinite Programming (ISDP) problem, which upon relaxation of binary variables results in a convex function. Thus the problem can be solved to global optimality using a branch and bound algorithm.

4.3.3 Sensor Network Design using Symmetric Kullback-Leibler divergence

In this subsection, we formulate an optimization problem for optimal sensor selection using the symmetric Kullback-Leibler divergence. The optimal sensor selection problem in terms of estimation errors in primary variables is formulated as follows:

$$\begin{aligned} \min_{q_i \in \{0,1\}} \quad & \mathcal{D}^s_{KL}(f_p, g_p) \\ s.t. \quad & \sum_i q_i = N_s \end{aligned} \tag{34}$$

The symmetric KL divergence in the above formulation (34) can be easily computed, alleviating the degeneracy issues in (31). Assuming no gross errors, the $\mathcal{D}^s_{KL}(f_p, g_p)$ can be obtained using (6) as,

$$\begin{aligned} \mathcal{D}^s_{KL}(f_p, g_p) = \frac{1}{4}\big[& Tr(\Sigma_{fp}{\Sigma_{gp}}^{-1}) + Tr(\Sigma_{gp}{\Sigma_{fp}}^{-1}) \\ & - 2n_p\big] \end{aligned} \tag{35}$$

It is worth noting that Σ_{gp} is an invertible and known matrix. Substituting for $\Sigma_{fp} = (C^T QC)^{-1}$ and letting $W_1 = \Sigma_{gp}^{-1}$ and $W_2 = \Sigma_{gp}$, we get,

$$\begin{aligned} \mathcal{D}^s_{KL}(f_p, g_p) = \frac{1}{4}\big[&Tr((C^T QC)^{-1}W_1) + Tr(W_2(C^T QC)) \\ &- 2n_p\big] \end{aligned} \tag{36}$$

Recall that $Q = diag\{q_i/\sigma_i^2\}$ where q_i is a binary variable indicating whether the i-th sensor is selected or not. The first term is nonlinear owing to inverse operation, whereas the second term is linear in terms of decision variable q_i, and the last term is a constant. Hence, the resulting optimization problem is a nonlinear integer programming problem, which is hard to solve in general. Therefore, we provide a convex reformulation of the symmetric KL divergence-based optimal sensor selection in the next subsection.

Convex Reformulation of Symmetric Kullback-Leibler divergence

By factorizing, $W_1 = BB^T$, where B is the square root of W_1, and using the property that cyclic permutations have no effect on the trace operator, we get

$$\begin{aligned} \mathcal{D}^s_{KL}(f_p, g_p) = \frac{1}{4}\big[&Tr(B^T(C^T QC)^{-1}B) \\ &+ Tr(W_2(C^T QC)) - 2n_p\big] \end{aligned} \tag{37}$$

Using definition 1 in (37), the optimization problem (34) can be rewritten as,

$$\begin{aligned} \min_{q_i \in \{0,1\}, \alpha} \quad & \frac{1}{2}\big[\alpha + Tr(W_2(C^T QC)) - 2n_p\big] \\ s.t. \quad & Tr(B^T(C^T QC)^{-1}B) \leq \alpha \\ & \sum_i q_i = N_s \end{aligned} \tag{38}$$

Introducing the internal matrix variable Y and using definition 2, the overall optimization problem can be recast as follows:

$$
\begin{aligned}
\min_{q_i \in \{0,1\}, \alpha, Y} \quad & \frac{1}{4}\left[\alpha + Tr(W_2(C^T Q C)) - 2n_p\right] \\
s.t. \quad & Tr(Y) \leq \alpha \\
& Y \succ 0 \\
& \begin{bmatrix} Y & B \\ B^T & C^T Q C \end{bmatrix} \succ 0 \\
& \sum_i q_i = N_s
\end{aligned}
\tag{39}
$$

The resulting formulation (39) is a Mixed Integer Semidefinite Programming (MISDP) problem. It is important to notice that the integer relaxation of the binary variable q_i requires one to solve an optimization problem with a linear cost function subject to linear matrix inequalities and linear equality. Thus, the relaxed problem is a semidefinite programming problem known to be convex in general. Therefore, we use branch and bound to solve the integer problem in which a semidefinite programming solver is used at each branching node. Hence, the proposed sensor selection problem employing the symmetric KL divergence can be solved using a branch and bound algorithm for global optimality. The optimization problem was solved using the SDPT3 toolbox in MATLAB R2021b.[31]

5 Case studies

5.1 Two splitter system

Consider a two-splitter system shown in Figure 1. There are a total of five flow streams and two process units. The steady-state balance equations governing the process are given by,

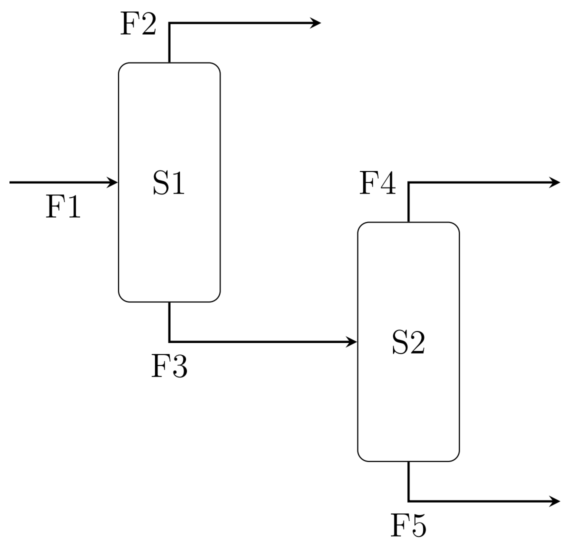

Figure 1: A simple splitter system

$$
\begin{aligned}
F_1 &= F_2 + F_3 \\
F_3 &= F_4 + F_5
\end{aligned}
\tag{40}
$$

Therefore, the degree of freedom of this system is three, i.e., any three variables that form an observable sensor network can be chosen as primary variables. Let us choose the primary variables as $x_p = \{F_1, F_2, F_5\}$ and the corresponding measurement matrix C can be obtained using (14) as,

$$
C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \tag{41}
$$

The number of observable sensor networks in this case is 8. In this work, the error variance values are considered to be unequal, and are given by, $\Sigma_\epsilon = diag\{0.37, 0.29, 0.45, 0.61, 0.49\}$. Table 1 presents the forward, reverse and symmetric KL divergence values of all the observable sensor networks. It can also be inferred from Table 1 that the optimal sensor networks obtained by minimizing the trace of estimation errors[29] and by minimizing the maximum eigenvalue[30] are different. The optimal sensor network obtained using entropy and the symmetric KL was found to be $\{F_1, F_2, F_5\}$ with the optimal $\mathcal{H}$ value of 2.30 and an optimal $\mathcal{D}^s_{KL}$ value of 0.57. Also to be noted is the fact that the optimal sensor network obtained by reverse KL divergence is the same as that of entropy and symmetric KL divergence, while the optimal network obtained using forward KL divergence was found to be $\{F_2, F_3, F_5\}$ with an optimal $\mathcal{D}^f_{KL}$ value of 0.81. This example illustrates that minimizing the forward KL is not the same as minimizing the reverse KL divergence as these measures are asymmetric. Therefore, we propose to use symmetric KL as a measure for sensor network design as it is an average of forward and reverse KL. Thus, determining the optimal sensor network based on symmetric KL signifies the minimal average estimation error obtained by accounting for the complete distribution information of estimation errors.

Table 1: Comparison of sensor network design obtained using different measures for the case of minimum observable networks

Sensor	$\mathrm{Tr}(\Sigma_e)$[29]	$\log\det\Sigma_{ep}$[30]	$\lambda_{max}(\Sigma_e)$[30]	$\mathcal{H}$	$\mathcal{D}^f_{KL}$	$\mathcal{D}^r_{KL}$	$\mathcal{D}^s_{KL}$
F_1,F_2,F_4	3.20	-2.73	2.11	2.89	1.06	0.40	0.73
F_1,F_2,F_5	2.96	-2.95	1.98	2.78[a]	0.82	0.34[c]	0.58[d]
F_1,F_3,F_4	3.31	-2.29	1.90	3.11	1.31	0.51	0.91
F_1,F_3,F_5	3.07	-2.51	1.78	3.00	1.11	0.45	0.78
F_1,F_4,F_5	4.04	-2.20	2.94	3.16	2.53	0.64	1.58
F_2,F_3,F_4	3.15	-2.53	1.87	2.99	1.02	0.43	0.72
F_2,F_3,F_5	2.91	-2.75	1.75	2.88	0.81[b]	0.36	0.59
F_2,F_4,F_5	3.88	-2.45	2.90	3.03	2.00	0.55	1.28

[a,b,c,d] are optimal sensor networks obtained using (24), (30), (33), and (39) respectively.

5.2 Ammonia system

Let us consider the representative process network of an ammonia process as presented in Figure 2.[2] It has five process units and eight flow streams. The steady-state mass balance equations of the process are given by,

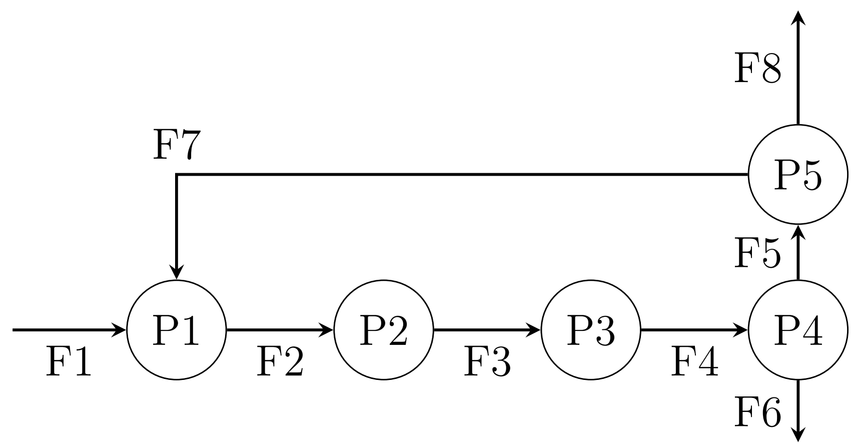

Figure 2: Simplified flow network of an ammonia process

$$
\begin{aligned}
F_2 - F_1 - F_7 &= 0 \\
F_3 - F_2 &= 0 \\
F_4 - F_3 &= 0 \\
F_5 + F_6 - F_4 &= 0 \\
F_7 + F_8 - F_5 &= 0
\end{aligned}
$$

The degree of freedom for the system is three. Therefore, any three variables that form an observable sensor network can be chosen as primary variables. In this work, we choose the primary variables as $x_p = \{F_1, F_2, F_5\}$, and the corresponding measurement matrix C is given by,

$$
C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 1 \end{bmatrix}^T \tag{42}
$$

Notice that the choice of primary variables does not affect the optimal solution, and the number of observable sensor networks for this process is 32. Table 2 compares average error, forward KL, reverse KL, and symmetric KL divergences of all the minimal observable

Table 2: Comparison of information-theoretic based sensor network design formulations

#	Sensor Networks	$\text{Tr}(\Sigma_e)^2$		$\mathcal{H}$		$\mathcal{D}^f_{KL}$		$\mathcal{D}^r_{KL}$		$\mathcal{D}^s_{KL}$	
		case 1	case 2	case 1	case 2	case 1	case 2	case 1	case 2	case 1	case 2
1	F_1,F_2,F_5	12	5.04	4.26	2.94	2.77	2.51	0.80	0.75	1.78	1.63
2	F_1,F_2,F_6	11	4.93	4.26	3.05	2.27	2.36	0.80	0.81	1.53	1.59
3	F_1,F_2,F_8	12	5.05	4.26	2.96	2.77	2.54	0.83	0.79	1.80	1.67
4	F_1,F_3,F_5	12	4.98	4.26	2.93	2.77	2.45	0.80	0.74	1.78	1.60
5	F_1,F_3,F_6	11	4.88	4.26	3.04	2.27	2.31	0.80	0.81	1.53	1.56
6	F_1,F_3,F_8	12	5.00	4.26	2.95	2.77	2.49	0.83	0.78	1.80	1.64
7	F_1,F_4,F_5	12	4.39	4.26	2.79^a	2.77	1.89	0.80	0.64^c	1.78	1.27^d
8	F_1,F_4,F_6	11	4.38	4.26	2.90	2.27	1.84^b	0.80	0.71	1.53	1.28
9	F_1,F_4,F_8	12	4.50	4.26	2.82	2.77	2.02	0.83	0.68	1.80	1.35
10	F_1,F_5,F_6	14	6.61	4.26	3.06	3.77	4.48	0.89	0.94	2.33	2.71
11	F_1,F_5,F_7	14	6.16	4.26	3.01	3.77	3.90	0.89	0.89	2.33	2.39
12	F_1,F_5,F_8	16	6.86	4.26	2.97	4.77	5.01	0.92	0.91	2.84	2.96
13	F_1,F_6,F_7	14	6.37	4.26	3.11	3.77	4.19	0.89	0.96	2.33	2.58
14	F_1,F_7,F_8	13	5.73	4.26	3.03	3.27	3.51	0.92	0.94	2.09	2.22
15	F_2,F_5,F_7	11	4.94	4.26	3.07	2.27	2.42	0.80	0.83	1.53	1.62
16	F_2,F_5,F_8	12	5.29	4.26	3.04	2.77	2.90	0.83	0.85	1.80	1.87
17	F_2,F_6,F_7	12	5.69	4.26	3.17	2.77	3.20	0.80	0.89	1.78	2.05
18	F_2,F_6,F_8	12	5.71	4.26	3.14	2.77	3.22	0.83	0.92	1.80	2.07
19	F_2,F_7,F_8	12	5.49	4.26	3.09	2.77	2.93	0.83	0.87	1.80	1.90
20	F_3,F_7,F_8	11	4.89	4.26	3.06	2.27	2.37	0.80	0.82	1.53	1.59
21	F_3,F_6,F_8	12	5.24	4.26	3.02	2.77	2.85	0.83	0.84	1.80	1.84
22	F_3,F_6,F_7	12	5.63	4.26	3.16	2.77	3.14	0.80	0.89	1.78	2.01
23	F_3,F_5,F_8	12	5.66	4.26	3.13	2.77	3.17	0.83	0.91	1.80	2.04
24	F_3,F_5,F_7	12	5.44	4.26	3.08	2.77	2.88	0.83	0.86	1.80	1.87
25	F_4,F_7,F_8	11	4.39	4.26	2.92	2.27	1.89	0.80	0.72	1.53	1.30
26	F_4,F_6,F_8	12	4.74	4.26	2.89	2.77	2.37	0.83	0.74	1.80	1.55
27	F_4,F_6,F_7	12	5.03	4.26	3.03	2.77	2.53	0.80	0.79	1.78	1.66
28	F_4,F_5,F_8	12	5.16	4.26	3.00	2.77	2.70	0.83	0.81	1.80	1.75
29	F_4,F_5,F_7	12	4.94	4.26	2.94	2.77	2.40	0.83	0.76	1.80	1.58
30	F_5,F_6,F_7	14	6.84	4.26	3.19	3.77	4.91	0.89	1.02	2.33	2.96
31	F_5,F_6,F_8	13	6.31	4.26	3.15	3.27	4.23	0.92	1.04	2.09	2.63
32	F_6,F_7,F_8	16	7.94	4.26	3.21	4.77	6.33	0.92	1.06	2.84	3.70

^{a,b,c,d} are optimal sensor networks obtained using (24), (30), (33), and (39) respectively

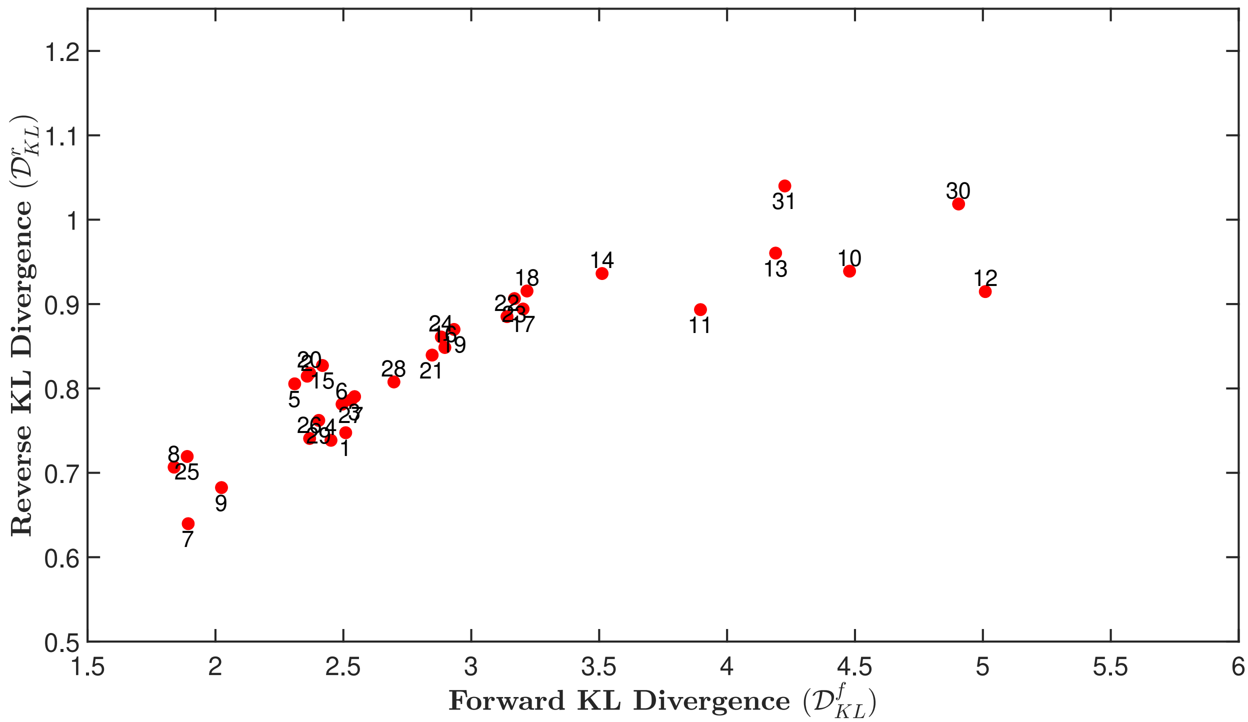

Figure 3: $\mathcal{D}_{KL}^f$ vs $\mathcal{D}_{KL}^r$ values of minimum observable sensor networks of ammonia process

sensor networks. Further, we consider two cases. In case 1, we assume equal, and unit error variances for all potential measured variables, that is, $\Sigma_\epsilon = diag\{1, 1, 1, 1, 1, 1, 1, 1\}$. All the optimal sensor networks obtained using the average error formulation are also optimal in terms of forward and symmetric KL divergence. The optimal sensor networks obtained are as follows: $\{F_1, F_2, F_6\}$, $\{F_1, F_3, F_6\}$, $\{F_1, F_4, F_6\}$, $\{F_2, F_5, F_7\}$, $\{F_3, F_5, F_7\}$, and $\{F_4, F_5, F_7\}$. These sensor networks have a symmetric KL divergence value of 1.53. In addition, reverse KL divergence have a few more optimal sensor networks that are not optimal with respect to symmetric KL divergence. They are as follows: $\{F_1, F_2, F_5\}$, $\{F_1, F_3, F_5\}$, $\{F_1, F_4, F_5\}$, $\{F_2, F_6, F_7\}$, $\{F_4, F_6, F_7\}$.

In case 2, we consider unequal error variances for all potential measured variables, that is, $\Sigma_\epsilon = diag\{0.380,\ 0.430, 0.420, 0.320, 0.440, 0.545, 0.490, 0.460\}$. The optimal sensor network obtained by minimizing the average error is $\{F_1, F_4, F_6\}$. In contrast, the optimal sensor network obtained by entropy and the proposed symmetric KL divergence was found to be $\{F_1, F_4, F_5\}$ with the optimal $\mathcal{H}$ value of 2.79 and $\mathcal{D}_{KL}^s$ value of 1.26. Coincidentally,

$\{F_1, F_4, F_5\}$ is also the optimal sensor network based on reverse KL divergence; however, it is not an optimal sensor network in terms of forward KL divergence. The optimal sensor network obtained using forward KL divergence is $\{F_1, F_4, F_6\}$, with a $\mathcal{D}^f_{KL}$ value of 1.84. This demonstrates that the sensor network that is optimal in terms of forward KL divergence may not be optimal in terms of reverse KL divergence, as these measures are asymmetric. This can also be clearly inferred using Figure 3.

Figure 4 shows the distribution of estimation errors of primary variables for different observable sensor networks. The reference distribution (best achievable) indicates the case where all sensors are chosen. It is provided to compare the relative accuracy of primary variables that can be achieved when the optimal sensors are placed. It can be seen from the plots that sensor network 7 is closest to the reference distribution in terms of similarity, while sensor network 32 is the furthest. Thus symmetric KL divergence provides a quantifying metric for designing sensor networks by accounting for the complete distribution of estimation errors in reconciled estimates.

Table 4 lists the optimal sensor networks obtained for different values of N_s. Notice that as the number of sensors increases, the estimation errors of process variables approach the least achievable estimation errors when all sensors are selected. This can also be seen from the symmetric KL divergence values. The diagonal plots in Figure 5 present the distribution of estimation errors in primary variables of the ammonia process. The reference distribution is again the same, i.e., it indicates the case where all sensors are chosen. The off-diagonal plots show the contours of estimation errors between two primary variables with a 95% confidence interval. For example, the contours in subplot(2,1) represent the estimation errors between F_1 and F_2. Thus, minimizing average error, $Tr(\Sigma_e)$, does not account for estimation errors between two different variables, although it will be affected by the choice of the sensor network. Therefore, the proposed symmetric KL divergence measure can be considered more meaningful since it accounts for the complete distribution of estimation errors in the sensor selection.

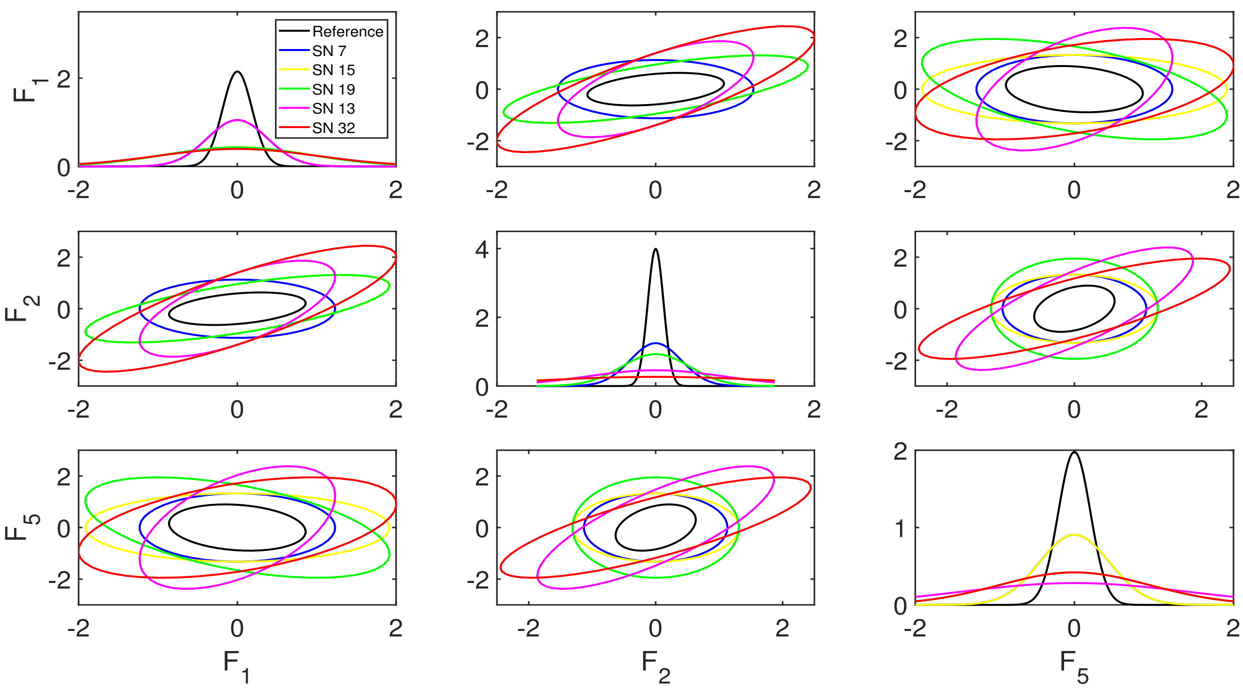

Figure 4: Plots showing estimation errors of various sensor networks

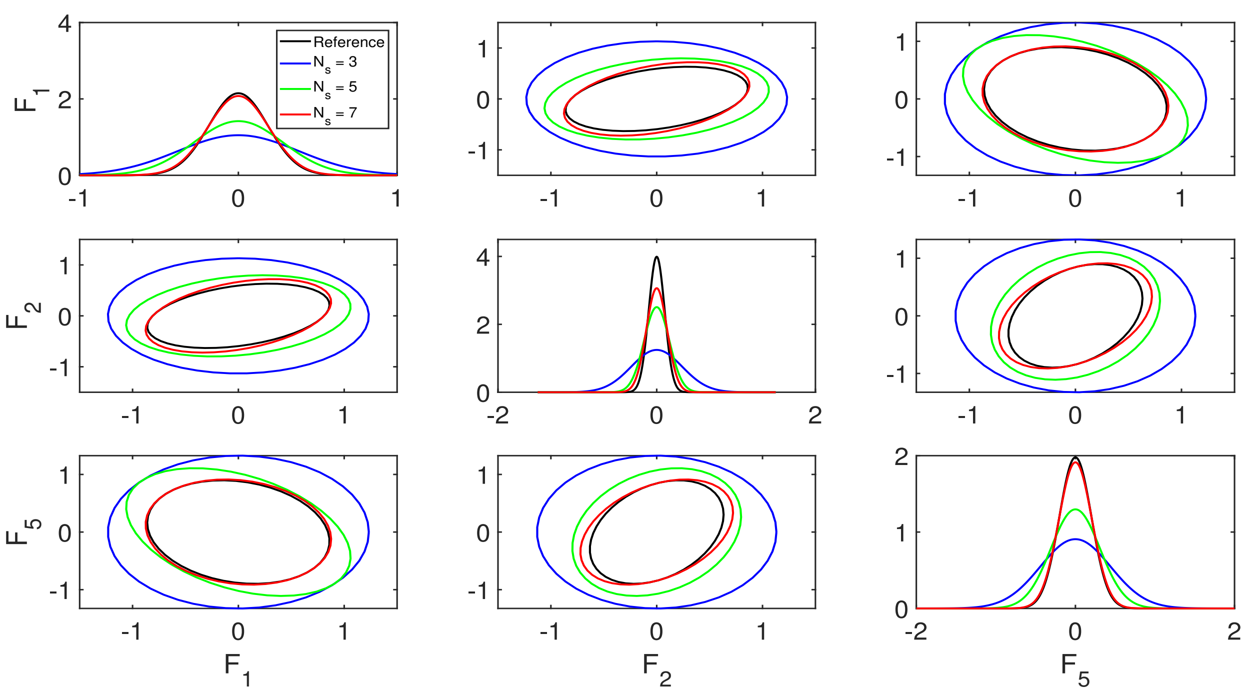

Figure 5: Plots showing estimation errors of primary variables when optimal sensor networks are chosen based on $\mathcal{D}^s_{KL}$ with $N_s = 3$, 5 and 7

Table 3: Optimal redundant sensor networks based on $\mathcal{H}$ for a given N_s

N_s	Optimal Sensor Networks	$\mathcal{H}$
3	F_1, F_4, F_5	2.79
4	F_1, F_4, F_5, F_8	2.16
5	F_1, F_4, F_5, F_7, F_8	1.87
6	$F_1, F_4, F_5, F_6, F_7, F_8$	1.60
7	$F_1, F_3, F_4, F_5, F_6, F_7, F_8$	1.41
8	$F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8$	1.28

Table 4: Optimal redundant sensor networks based on $\mathcal{D}_{KL}^s$ for a given N_s

N_s	Optimal Sensor Networks	$\mathcal{D}_{KL}^s$
3	F_1, F_4, F_5	1.26
4	F_1, F_4, F_5, F_8	0.37
5	F_1, F_3, F_4, F_5, F_8	0.18
6	$F_1, F_3, F_4, F_5, F_7, F_8$	0.09
7	$F_1, F_3, F_4, F_5, F_6, F_7, F_8$	0.01
8	$F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8$	0.00

5.3 Steam Metering Network

Consider the steam metering system as shown in Figure 6. The steam metering network has twenty eight flow streams and eleven process systems. Hence the degree of freedom for the system is 17, i.e., any seventeen variables that form an observable sensor network can be chosen as primary variables. In this work, we choose the primary variables as $x_p = \{F_1, F_2, F_5, F_6, F_7, F_8, F_{12}, F_{14}, F_{15}, F_{16}, F_{17}, F_{19}, F_{20}, F_{21}, F_{23}, F_{24}, F_{27}\}$. For details on the model, the reader is referred to the Supplementary Material.

In this case, the minimum number of observable sensor networks are 12,43,845. We again consider the case of unequal variances and is given by, $\Sigma_\epsilon = diag\{0.021, 0.025, 2.795, 2.748, 1.331, 2.806, 0.058, 4.101, 0.021, 1.310, 0.371, 1.681, 2.781, 2.296, 1.50, 0.591, 0.818, 0.405, 0.198, 0.262, 2.181, 0.136, 0.064, 1.166, 2.136, 2.033, 1.769, 1.805\}$ taken from the work by Narasimhan.[29] In this case, the optimal sensor network obtained by minimizing the average error is $\{F_1, F_2, F_4, F_7, F_9, F_{10}, F_{11}, F_{13}, F_{16}, F_{18}, F_{19}, F_{20}, F_{21}, F_{22}, F_{23}, F_{24}, F_{27}\}$, whereas the one obtained using entropy is $\{F_1, F_2, F_7, F_9, F_{10}, F_{11}, F_{16}, F_{17}, F_{18}, F_{19}, F_{20}, F_{21}, F_{22}, F_{23}, F_{24}, F_{27}F_{28}\}$ with an en-

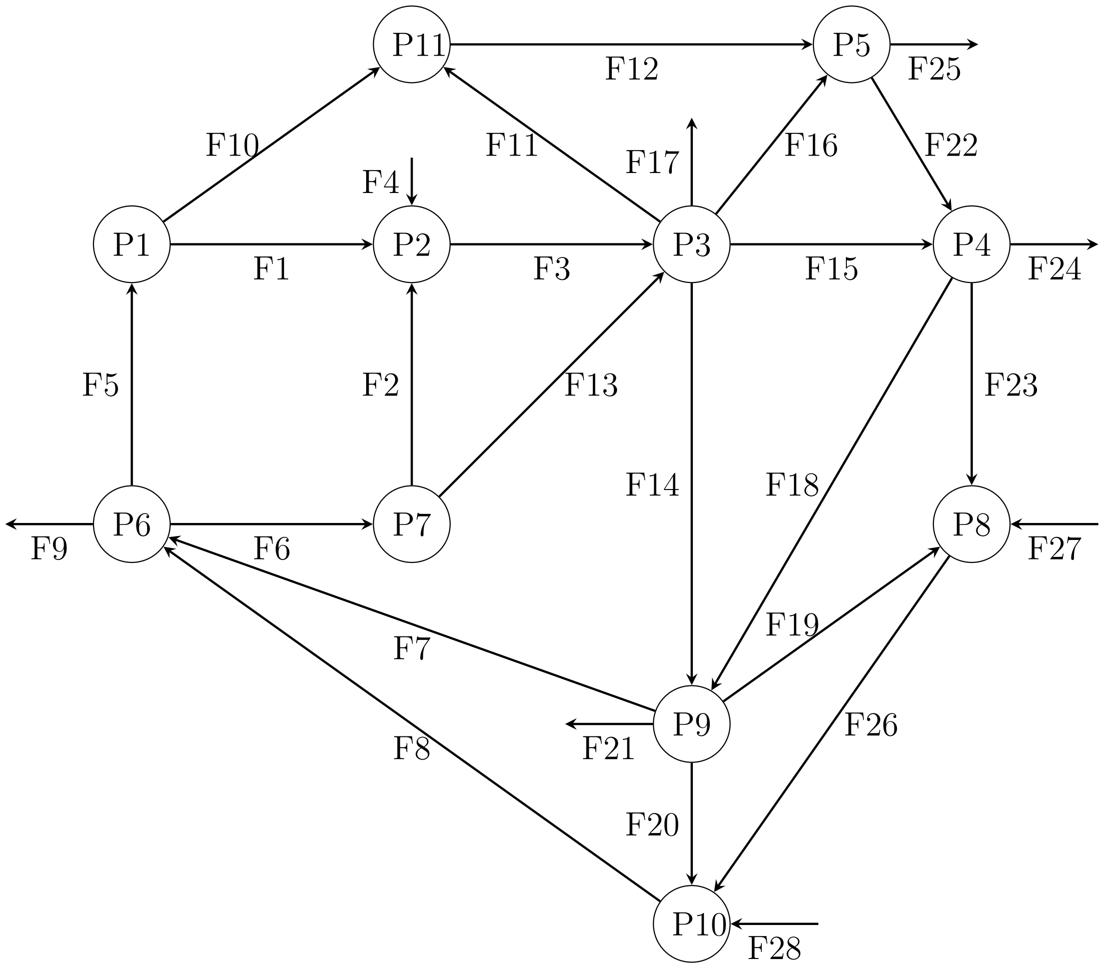

Figure 6: Steam metering system

tropy value of 13.04 and that using the proposed approach is $\{F_1, F_2, F_7, F_9, F_{10}, F_{11}, F_{13}, F_{16}, F_{17}, F_{18}, F_{19}, F_{20}, F_{21}, F_{22}, F_{23}, F_{24}, F_{27}\}$ with the symmetric KL divergence value of 3.62. Also to be noted is the fact that using forward KL divergence also gives the same optimal sensor network as using symmetric KL divergence, whereas reverse KL divergence gives the same optimal network as using entropy. Table 5 presents average error, entropy, forward KL divergence values, reverse KL divergence, and symmetric KL divergence values of a few of the minimum observable sensor networks, respectively. Due to solver issues, forward KL did not converge in this case. Figure 7 shows the convergence plots of the trace based formulation with the proposed entropy and symmetric KL formulation in a steam metering system. It is worth mentioning that the entropy formulation converged in 293 seconds. The

Table 5: Representative sensor network designs in steam metering system

Sensor Network	$\mathrm{Tr}(\Sigma_e)^2$	$\mathcal{H}$	$\mathcal{D}^f_{KL}$	$\mathcal{D}^r_{KL}$	$\mathcal{D}^s_{KL}$
$F_1,F_2,F_3,F_5,F_6,F_7,F_8,F_{11},F_{14},F_{15},F_{16},F_{18},F_{19},F_{20},F_{24},F_{27},F_{28}$	91.20	19.40	300.96	6.70	153.83
$F_1,F_2,F_7,F_9,F_{10},F_{11},F_{16},F_{17},F_{18},F_{19},F_{20},F_{21},F_{22},F_{23},F_{24},F_{27},F_{28}$	59.08	13.04^a	6.04	1.56^c	3.80
$F_1,F_2,F_3,F_6,F_7,F_8,F_{10},F_{11},F_{14},F_{15},F_{16},F_{22},F_{23},F_{24},F_{26},F_{27},F_{28}$	80.13	19.31	223.52	6.43	114.97
$F_1,F_2,F_7,F_9,F_{10},F_{11},F_{13},F_{16},F_{17},F_{18},F_{19},F_{20},F_{21},F_{22},F_{23},F_{24},F_{27}$	55.36	13.26	5.60^b	1.64	3.62^d
$F_1,F_3,F_6,F_7,F_8,F_9,F_{11},F_{13},F_{14},F_{17},F_{19},F_{20},F_{22},F_{23},F_{24},F_{25},F_{28}$	133.68	17.83	157.48	5.40	81.44
$F_2,F_4,F_5,F_6,F_7,F_8,F_{10},F_{12},F_{14},F_{15},F_{18},F_{19},F_{20},F_{22},F_{23},F_{25},F_{26}$	76.20	20.17	264.66	7.01	135.83
$F_1,F_2,F_4,F_7,F_9,F_{10},F_{11},F_{13},F_{16},F_{18},F_{19},F_{20},F_{21},F_{22},F_{23},F_{24},F_{27}$	53.38	13.86	8.86	2.01	5.43
$F_1,F_2,F_4,F_5,F_6,F_7,F_8,F_{12},F_{14},F_{15},F_{16},F_{18},F_{19},F_{20},F_{22},F_{23},F_{26}$	67.96	17.48	203.24	4.87	104.05
$F_3,F_4,F_5,F_6,F_7,F_9,F_{10},F_{11},F_{14},F_{15},F_{16},F_{19},F_{22},F_{23},F_{24},F_{27},F_{28}$	89.84	19.93	248.62	7.07	127.8
$F_3,F_4,F_5,F_6,F_8,F_{12},F_{13},F_{14},F_{15},F_{17},F_{18},F_{21},F_{24},F_{25},F_{26},F_{27},F_{28}$	177.16	29.12	1074.69	14.77	544.73
$F_1,F_2,F_7,F_9,F_{10},F_{11},F_{14},F_{16},F_{17},F_{18},F_{19},F_{21},F_{23},F_{24},F_{25},F_{27},F_{28}$	84.08	15.50	31.74	3.51	17.63

^{a,c,d} are optimal sensor networks obtained using (24), (33), and (39) respectively. ^{b} is obtained by enumeration.

proposed symmetric KL divergence formulation converged in 3308 seconds, however, the average error (trace of estimation error) formulation did not converge even after 24 hours. This demonstrates the computational efficacy of the proposed information theoretic-based sensor network design formulations.

6 Data Availability and Reproducibility Statement

The details of the linear process model, corresponding C matrix, and the sensor error variances of all three case studies are provided in Supplementary Material. Additionally, the estimation error covariances of primary variables for different sensor networks that are used for plotting Figures 4 and 5 have also been provided. The .ZIP file contains all the numerical data for plotting the distribution contours (Figures 4 and 5), and the convergence plot (Figure 7). Data used for plotting Figure 3 are presented in Table 2 of this manuscript.

7 Conclusions

This work presents a convex optimization approach to the design of sensor networks for data reconciliation in a linear steady-state process. We have proposed different information theoretic measures that account for the complete distribution of estimation errors in the

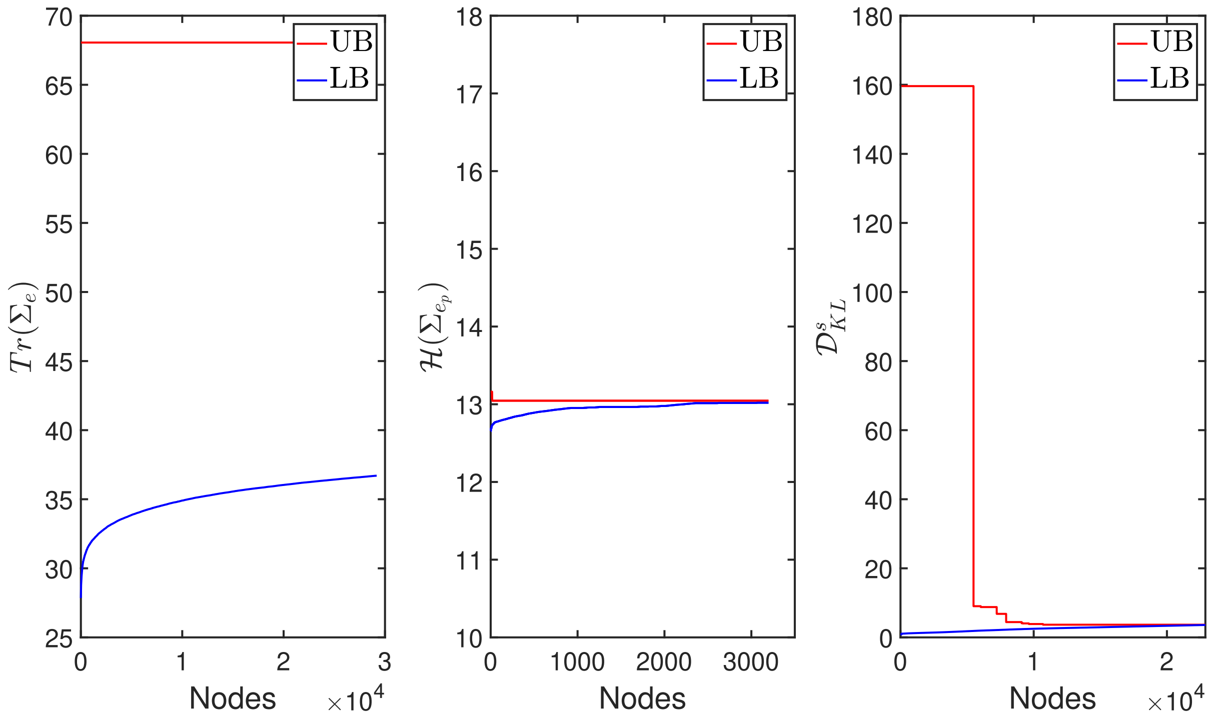

Figure 7: Convergence plots of the $Tr(\Sigma_e)$, $\mathcal{H}$, and $\mathcal{D}^s_{KL}$ formulations

design of sensor networks. Firstly, it was shown that the optimal sensor selection problem formulated using the entropy of estimation errors and reverse KL are integer semidefinite programming problems. Further, it was shown that the minimization of entropy of estimation errors is the same as the maximization of the logarithm of the determinant of estimation errors. Secondly, it was shown that the design of an optimal sensor network using forward, and symmetric KL, are formulated as mixed integer semidefinite programming problems. All the proposed sensor network design formulations using information-theoretic measures are convex when binary variables are relaxed. Thus we obtain the globally optimal sensor networks using all the proposed formulations. Furthermore, we highlighted that the forward and reverse KL divergences yield different optimal sensor networks even when the reference distribution is the same due to the asymmetric property. On the other hand, a symmetric KL divergence-based sensor network design problem determines the sensor network that minimizes the average between the forward and the reverse KL divergences as it alleviates the asymmetric issue. Therefore, we propose to use symmetric KL as a measure to design sensor networks as it is symmetric, and accounts for the complete distribution information of estimation errors. Case studies demonstrated that the proposed formulations yield globally optimal sensor networks, as well as, their computational effectiveness. Demonstration of the proposed symmetric KL divergence-based sensor selection on a large-scale process network is currently under investigation.

Acknowledgement

We appreciate the funding support from Science and Engineering Research Board through Start-Up Research Grant number SRG/2021/001653. We express our gratitude to Prof. Johan Löfberg from Linköping University, Sweden, for providing us with invaluable assistance pertaining to the SDPT3 solver.

References

1. Kretsovalis A, Mah RS. Effect of redundancy on estimation accuracy in process data reconciliation. *Chemical Engineering Science.* 1987;42(9):2115–2121.

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

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

4. Wei J, Zhang Q, Yuan Z. A unified approach to multiscenario sensor network design for data reconciliation. *AIChE Journal.* 2022;68(1):e17404.

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

6. Bagajewicz MJ. Design and retrofit of sensor networks in process plants. *AIChE Journal.* 1997;43(9):2300–2306.

7. Bagajewicz MJ, Sanchez MC. Design and upgrade of nonredundant and redundant linear sensor networks. *AIChE Journal.* 1999;45(9):1927–1938.

8. Bagajewicz M, Sanchez M. Cost-optimal design of reliable sensor networks. *Computers & Chemical Engineering.* 2000;23(11-12):1757–1762.

9. Chmielewski DJ. Convex methods in sensor placement. *IFAC Proceedings Volumes.* 2001;34(27):267–272.

10. Chmielewski DJ, Palmer T, Manousiouthakis V. On the theory of optimal sensor placement. *AIChE Journal.* 2002;48(5):1001–1012.

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

12. Ali Y, Narasimhan S. Redundant sensor network design for linear processes. *AIChE Journal.* 1995;41(10):2237–2249.

13. Ali Y, Narasimhan S. Sensor network design for maximizing reliability of bilinear processes. *AIChE Journal.* 1996;42(9):2563–2575.

14. Kotecha PR, Bhushan M, Gudi R, Keshari M. A duality based framework for integrating reliability and precision for sensor network design. *Journal of Process Control.* 2008;18(2):189–201.

15. Prakash O, Bhushan M, Narasimhan S, Rengaswamy R. Sensor network design based on system-wide reliability criteria. Part II: Formulations and applications. *Journal of Process Control.* 2020;93:14–27.

16. Prakash O, Bhushan M, Narasimhan S, Rengaswamy R. Sensor network design based on system-wide reliability criteria. Part I: Objectives. *Journal of Process Control.* 2020;93:66–82.

17. Vaclavek V, Loučka M. Selection of measurements necessary to achieve multicomponent mass balances in chemical plant. *Chemical Engineering Science.* 1976;31(12):1199–1205.

18. Bhushan M, Narasimhan S, Rengaswamy R. Robust sensor network design for fault diagnosis. *Computers & Chemical Engineering.* 2008;32(4-5):1067–1084.

19. Crowe CM. Formulation of linear data reconciliation using information theory. *Chemical Engineering Science.* 1996;51(12):3359–3366.

20. Liu W, Pokharel PP, Principe JC. Correntropy: A localized similarity measure. in *2006 IEEE International Joint Conference on Neural Network Proceedings*:4919–4924.

21. Yang C, Xie S, Yuan X, Wang X, Xie Y. A new data reconciliation strategy based on mutual information for industrial processes. *Industrial & Engineering Chemistry Research.* 2018;57(38):12861–12870.

22. Wang H, Yao K, Pottie G, Estrin D. Entropy-based sensor selection heuristic for target localization. in *Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks*:36–45 2004.

23. Tjen J, Smarra F, D'Innocenzo A. An entropy-based sensor selection algorithm for structural damage detection. in *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*:1566–1571.

24. Adurthi N, Singla P, Majji M. Mutual information based sensor tasking with applications to space situational awareness. *Journal of Guidance, Control, and Dynamics.* 2020;43(4):767–789.

25. Chithra B, Muniraj N. A mutual information based sensor selection and information controlled transmission power adjustment. in *2013 International Conference on Optical Imaging Sensor and Security (ICOSS)*:1–10.

26. Prakash O, Bhushan M. A Novel Information Theoretic Measure Based Sensor Network Design Approach for Steady State Linear Data Reconciliation. *IFAC-PapersOnLine.* 2020;53(2):3583–3588.

27. Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal.* 1948;27(3):379–423.

28. Thomas M, Joy AT. *Elements of information theory.* Wiley-Interscience, 2006.

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

30. Boyd S, Vandenberghe L. *Convex optimization.* Cambridge university press, 2004.

31. Löfberg J. YALMIP : A Toolbox for Modeling and Optimization in MATLAB. in *In Proceedings of the CACSD Conference,*(Taipei, Taiwan) 2004.