Industrial and Engineering Chemistry Research, Volume 54, Issue 48, December 2015, pp. 12040–12053
DOI: 10.1021/acs.iecr.5b00844

Global approximation of self-optimizing controlled variables with average loss minimization

Lingjian Ye,*,† Yi Cao,‡ and Xiaofeng Yuan¶

†*Ningbo Institute of Technology, Zhejiang University, 315100, Ningbo, Zhejiang, China*

‡*School of Energy, Environment and Agrifood, Cranfield University, Cranfield, Bedford MK43 0AL, UK*

¶*Department of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang, P. R. China*

E-mail: Lingjian.ye@gmail.com

Abstract

Self-optimizing control (SOC) constitutes an important class of control strategies for real-time optimization (RTO) of chemical plants, by means of selecting appropriate controlled variables (CVs). Within the scope of SOC, this paper develops a CV selection methodology for a global solution which aims to minimise the average economic loss across the entire operation space. A major characteristic making the new scheme different from existing ones is that each uncertain scenario is independently considered in the new solution without relying on a linearised model, which was necessary in existing local SOC methods. Although global CV selection has been formulated as a nonlinear programming (NLP) problem, a tractable numerical algorithm for a rigorous solution is not available. In this work, a number of measures are introduced to ease the challenge. Firstly, we suggest to represent the economic loss as a quadratic function against the controlled variables through Taylor expansion, such that the average loss becomes an explicit function of the CV combination matrix, a direct optimizing

algorithm is proposed to approximately minimize the global average loss. Furthermore, an analytic solution is derived for a suboptimal but much more simplified problem by treating the Hessian of the cost function over the entire operating space as a constant. This approach is found very similar to one of existing local methods, except that a matrix involved in the new solution is constructed from global operating data instead of using a local linear model. The proposed methodologies are applied to two simulated examples, where the effectiveness of proposed algorithms are demonstrated.

Nomenclature

$\mathbf{B}$	arbitrary nonsingular matrix
$\mathbf{c}$	controlled variables
$\mathbf{d}$	disturbance variables
$\mathbf{e}_c,\ \mathbf{e}_u$	deviation of $\mathbf{c}$ and $\mathbf{u}$ from the optimum
$\mathbf{f}$	plant output model
$\mathbf{F}$	gain matrix between optimal outputs and $\mathbf{d}$
$\tilde{\mathbf{F}}$	intermediate matrix
$\mathbf{g}$	process constraints
$\mathbf{G}_y,\ \mathbf{G}_{yd}$	gain matrices of $\mathbf{y}$ for $\mathbf{u}$ and $\mathbf{d}$
$\mathbf{H},\ \mathbf{h}$	CV combination matrix/vector
J	cost function
$\mathbf{J}_{cc},\ \mathbf{J}_{uu},\ \mathbf{J}_{ud}$	second order sensitivity matrices
L, L_{lav}, L_{gav}	loss function, local and global average loss
$\bar{L}_{\mathrm{gav}}$, $\bar{\bar{L}}_{\mathrm{gav}}$	approximated global average loss in Algorithm 1 and 2
L^d, L^n	loss caused by $\mathbf{d}$ and $\mathbf{n}$
N	number of sampled disturbance scenarios
$\mathbf{M}_d,\ \mathbf{M}_n$	intermediate matrices
$\mathbf{n}$	measurement errors/noises
$\mathbf{u}$	manipulated variables
$\mathbf{W}_d,\ \mathbf{W}_n$	magnitude matrices for $\mathbf{d}$ and $\mathbf{n}$

$\mathbf{y}_m$, $\mathbf{y}$	measured and true measurement variables
$\mathbf{Y}$, $\tilde{\mathbf{Y}}$	optimal measurement matrix and extended measurement matrix
Convention	
$(\cdot)^{\mathrm{nom}}$	variables at the nominal point
$(\cdot)^{\mathrm{fb}}$	variables under closed-loop control
$(\cdot)_{(i)}$	variables for the ith disturbance scenario
Δ	deviation of a variable from the nominal point
$\overset{\circ}{(\cdot)}$	optimal variable values under different distrubance

Introduction

Chemical plants are often initially operated with satisfactory performances, which however, tend to decay with time. The main reason for such performance degradation is the occurrences of uncertainties/disturbances (e.g. equipment aging, raw material composition variation, heat exchanger fouling, etc), thus changing the process characteristics. When such a change occurs, the original operating policy is no longer appropriate and causes unnecessarily poor operation performance. To keep competitive in the global market, it is heavily desired for chemical factories to gain more profit by performing real-time optimization (RTO), which reoptimizes the process operation under changing operating conditions.

In the field of RTO, various approaches have been developed to improve process operation. According to Chachuat *et al.*,[1] these approaches can be classified into three categories, namely the model-parameter adaptation (two-step approach), modifier adaptation,[2] and direct input adaptation,[3,4] based on their own routines of realizing RTO. Among various approaches, self-optimizing control (SOC)[3] within the third class was demonstrated to be a promising one, which refers to a control strategy that when feedback controllers tracking constant set-points of some carefully selected controlled variables (CVs), the plant operation is automatically maintained as optimal

or near optimal without re-optimization in the present of various uncertainties. Furthermore, if the economic loss is acceptable, there is no need to activate an optimization layer, which plays an important role in a traditional hierarchical control structure. Several important characteristics of SOC are as follows:

- The motivation of SOC is using a simple control structure to achieve an acceptable economic performance, the operation is not guaranteed to eventually converge to the exact optimum, as compared to other RTO realizations.[1,3] In contrast, an acceptable loss is pursued.
- Parametric uncertain systems are considered. That is, the rigorous nonlinear process model with uncertain parameters (disturbances) are known, however, the true values of disturbances are unknown to operators during on-line operation. Structural plant-model mismatch and unexpected disturbances are not directly addressed in SOC.
- The optimizing speed of SOC is fast, which is attributed to the feature that control policy of SOC is designed off-line, whilst on-line, the plant operates under simple feedback loops. One does not need to extract necessary information by conducting on-line experiments, which are probably time consuming.

The key issue of SOC is the CV selection, which was revealed to be highly relevant to the economic performance of plant operation. In the past years since proposed, the methodology of CV selection has been developed in diverse directions to achieve better RTO performance, which is summarized as follows.

1. **CV selection methodology for static optimization.** This direction was the most intensively studied area of SOC. Halvorsen *et al.*[5] characterized CV selection as a non-linear programming (NLP) problem for economic operation of chemical plants. However, due to the complicacy of the formulated NLP, they alternatively derived an exact local method for CV selection . Furthermore, an approximate minimum singular value rule was also introduced

for quick screening CV candidates. Besides using single measurements as CVs, combinations of measurements were also advocated to give better performances later on. Alstad and Skogestad[6] proposed a null space method for selecting linear combinations of measurements, $\mathbf{c} = \mathbf{H}\mathbf{y}$, which achieves locally zero loss caused by disturbances. The null space method was later extended to reduce the economic loss caused by implementation error using extra measurements.[7] In parallel, Kariwala[8] and Kariwala *et al.*[9] developed eigenvalue decomposition approaches to minimize the local worst-case loss and the local average loss. A common feature of these methods is that they rely on a linearized model around the nominal operating point, thus leading to the self-optimizing performance only locally satisfactory. Recently, some global SOC approaches addressing process non-linearity were also considered. Ye *et al.*[10] proposed a necessary conditions of optimality (NCO) approximation method, where the CVs are selected as measurement functions to approximate the NCO. Since the information of the entire operation region is used, this approach is globally valid in contrast to those local SOC approaches. Jaschke and Skogestad[11] derived polynomial combinations of measurements as CVs, the unknown disturbances are eliminated using elimination theory. However, this method requires both the economic function and process models defined in a polynomial ring such that the cumbersome symbolic computations may become infeasible for more general non-linear processes, which restricts its practical usage to certain extent.

2. **Dynamic SOC.** Extensions to self-optimizing control dynamic systems have also been reported. Dahl-Olsen *et al.*[12] used maximum gain rule for CV selection of batch processes, where single measurement as CV was considered. Hu *et al.*[13] formulated dynamic SOC as an optimal control problem, a local perturbation control approach was proposed to determine the CV candidates. Jaschke and Fikar[14] and Ye *et al.*[15] controlled the Hamiltonian function to achieve SOC of batch processes, nonetheless, the Hamiltonian function requires to be analytically derived in their approaches.

3. **Measurement subset selection.** For practical usage, measurement subset selection for constructing CVs is of paramount importance. A general conclusion is that with more mea-

surements, better self-optimizing performance can be expected. However, the gained profit gradually decreases with using more measurements. Therefore, it is a reasonable way to use a few rather than all measurements to derive CVs while still achieves a similar performance. Due to the combinational nature of subset selection problem, the branch and bound (BAB) method was demonstrated to be the most suitable algorithm. Kariwala and Skogestad[16] provided a BAB algorithm for CV selection using minimum singular value rule. Cao and Kariwala[17–19] published a series of bidirectional BAB algorithms that were particularly efficient for screening CV candidates based on different criterions. In another work,[20] the BAB algorithm was further extended to deal with the regression based NCO approximation method, which is also applicable to general regression problem. Besides BAB, the Mixed Integer Quadratic Programming (MIQP) algorithm was also considered to select a measurement subset.[21]

4. **Other issues.** Although most SOC methods were developed on basis of invariant active constraints, in the work of Manum and Skogestad,[22] the nullspace method was extended to cover the situation with active set changes. In another work,[23] CV selection approach was proposed to minimize the local average loss while ensuring all the constraints satisfied over the allowable set of uncertainties. Ye *et al.*[24] proposed a data-driven SOC strategy, where the process model was identified from the historical operating data. SOC in the framework of a hierarchical control structure were also reported. For example, Jaschke and Skogestad[25] developed a control hierarchy where the lower layer implements SOC while the upper layer performs NCO tracking.[26] Ye *et al.*[27] devised a scheme to make the CVs adaptive to operating conditions in a new hierarchical control structure, hence improving the optimizing performance.

This paper is within the scope of SOC. Particularly, a new CV selection methodology for static optimization is proposed, which is devoted to deal with the SOC problem by minimizing the global average loss. A major characteristic that makes the new scheme different from other existing SOC approaches is that during the developments, each disturbance scenario within the entire operation

region is independently considered, rather than relying on a single linearized model, as was done in previous local SOC methods. Similarly to local SOC approaches, the numerical algorithm developed in this work uses a quadratic model to evaluate the loss. Nevertheless, the difference is that the quadratic model in the new algorithm is derived to fit data representing the whole operating region.

The reminder of this paper is organized as follows: Section 2 firstly presents some background knowledge of the SOC, particularly, a local SOC algorithm is reviewed.[7] Section 3 presents the SOC problem as a nonlinear programming (NLP) problem, which aims to minimize the global average economic loss in the entire uncertain space. Two algorithms are proposed to search optimal CVs. In Section 4, the proposed methodology is applied to two simulated examples. Finally, Section 5 concludes this paper.

Local self-optimizing control

Problem formulation

Consider a static optimization problem

$$\begin{aligned} &\min_{\mathbf{u}} J(\mathbf{u}, \mathbf{d}) \\ &\text{s.t. } \mathbf{g}(\mathbf{u}, \mathbf{d}) \leq 0 \end{aligned} \tag{1}$$

with measurements

$$\mathbf{y} = \mathbf{f}(\mathbf{u}, \mathbf{d}) \tag{2}$$

$$\mathbf{y}_m = \mathbf{y} + \mathbf{n} \tag{3}$$

where J is a scalar cost function to be minimized, $\mathbf{u} \in \mathbb{R}^{n_u}$, $\mathbf{d} \in \mathbb{R}^{n_d}$, $\mathbf{n}$, $\mathbf{y}_m$ and $\mathbf{y} \in \mathbb{R}^{n_y}$ are the manipulated variables, disturbances, measurement noise, measured and theoretical measure-

ments, respectively. $\mathbf{g} : \mathbb{R}^{n_u \times n_d} \rightarrow \mathbb{R}^{n_g}$ and $\mathbf{f} : \mathbb{R}^{n_u \times n_d} \rightarrow \mathbb{R}^{n_y}$ are the operational constraints and measurement models, respectively.

Assuming that at the optimal point, a subset of the operational constraints, say $\mathbf{g}_a$ (*i.e.* active constraints), should be satisfied. The active constraints are generally directly measured and controlled as the CVs, which consume the same number of the degrees of freedom. Furthermore, in this paper, the active set is assumed to keep unchanged in the whole operating region. We deal with the minimization problem of scalar cost function J with respect to the remaining degrees of freedom in the reduced unconstrained subspace. Hence, an unconstrained optimization problem can be equivalently stated as follows

$$\min_{\mathbf{u}} J(\mathbf{u}, \mathbf{d}) \tag{4}$$

Note, in the above equation, for simplicity, we use the same notation $\mathbf{u}$ as in the original problem (1) for the remaining unconstrained manipulated variables. Obviously, the necessary condition of optimality for this unconstrained problem is that the first order derivatives of J with respective to $\mathbf{u}$, $\mathbf{J}_u = 0$. However, $\mathbf{J}_u$ is generally hard to measure on-line because of the unknown disturbances. Moreover, direct evaluation of $\mathbf{J}_u$ may not be tractable due to the complexity of process model equations implicitly embedded in the cost function, J.

As an alternative to enforcing $\mathbf{J}_u = 0$, SOC achieves this goal by tracking constant set-points of some particularly chosen CVs, which are combinations of measurements, denoted as $\mathbf{c} = \mathbf{H}\mathbf{y}$. Here, $\mathbf{H}$ is a $(n_c \times n_y)$-dimensional combination matrix as

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_{n_c} \end{bmatrix} = \begin{bmatrix} h_{11} & \cdots & h_{1n_y} \\ \vdots & & \vdots \\ h_{n_c1} & \cdots & h_{n_cn_y} \end{bmatrix}$$

where $\mathbf{h}_i$ $(i = 1, \ldots, n_c)$ is a row vector of coefficients to be designed for an individual CV. Here, we generally have $n_c = n_u$ to form a square control problem. Besides, we require $n_c \leq n_y$, otherwise, the system is over-determined hence perfect control for $\mathbf{c}$ cannot be realized.

Local self-optimizing control methods

To make the self-optimizing control problem tractable, the first effort was to simplify the process model through linearisation around a reference point, where for a nominal disturbance, $\mathbf{d}^{\text{nom}}$, the input is nominally optimal, $\mathring{\mathbf{u}}^{\text{nom}} = \mathring{\mathbf{u}}(\mathbf{d}^{\text{nom}})$ and the corresponding measurement is, $\mathring{\mathbf{y}}^{\text{nom}}$. (In this paper, we use $\mathring{(\cdot)}$ to represent those variables under *optimal operation* and $(\cdot)^{\text{nom}}$ as those under the nominal condition.) Due to the linearisation involved, the controlled variables to be determined is only locally valid within a small region around the reference point, so that methods developed based on linearisation are referred to as local SOC methods such as.[5–9,21] The linearised process model is as follows.

$$\Delta\mathbf{y} = \mathbf{G}_y \Delta\mathbf{u} + \mathbf{G}_{yd} \Delta\mathbf{d} \tag{5}$$

where $\mathbf{G}_y$ and $\mathbf{G}_{yd}$ are the gain matrices of $\mathbf{y}$ associated with $\mathbf{u}$ and $\mathbf{d}$ respectively, the symbol "Δ" denotes the deviation of a variable from its nominal optimal value *i.e.* $\Delta\mathbf{y} = \mathbf{y} - \mathring{\mathbf{y}}^{\text{nom}}$, $\Delta\mathbf{u} = \mathbf{u} - \mathring{\mathbf{u}}^{\text{nom}}$ and $\Delta\mathbf{d} = \mathbf{d} - \mathbf{d}^{\text{nom}}$. In the following, we outline some basic results of SOC methodology, a complete summary of local SOC methods can be found in a recent book by Rangaiah and Kariwala, see ref.[28]

Evaluation of economic loss

As a criterion for CV selection, an economic loss is defined as

$$L = J(\mathbf{u}, \mathbf{d}) - \mathring{J}(\mathbf{d}) \tag{6}$$

In principle, L should be calculated from its original definition in (6). However, the calculation is often not straightforward, particularly, when the plant is operated in closed-loop. Alternatively, it is approximated through Taylor expansion of L with respect to some free variables, for example, either $\mathbf{u}$ or $\mathbf{c}$.

L in terms of u. Applying second-order Taylor expansion of L around the nominally optimal point $\mathring{\mathbf{u}}^{\text{nom}}$, L can be locally approximated as

$$L \approx \frac{1}{2}\mathbf{e}_u^{\mathrm{T}}\mathbf{J}_{uu}\mathbf{e}_u \tag{7}$$

where $\mathbf{e}_u \coloneqq \mathbf{u}^{fb} - \mathring{\mathbf{u}}^{\text{nom}}$ is input deviations from the nominally optimal point under closed-loop control, $\mathbf{u}^{fb}$ represents the location of $\mathbf{u}$ after controlling the selected CVs. $\mathbf{J}_{uu}$ is the Hessian matrix of J with respect to $\mathbf{u}$ evaluated at $\mathring{\mathbf{u}}^{\text{nom}}$.

Using the linear relationship of (5), $\mathbf{e}_u$ can be derived in terms of $\Delta\mathbf{d}$ and $\mathbf{n}$ as[5]

$$\mathbf{e}_u = (\mathbf{J}_{uu}^{-1}\mathbf{J}_{ud} - (\mathbf{H}\mathbf{G}_y)^{-1}\mathbf{H}\mathbf{G}_{yd})\Delta\mathbf{d} + (\mathbf{H}\mathbf{G}_y)^{-1}\mathbf{H}\mathbf{n} \tag{8}$$

where $\mathbf{J}_{ud}$ is a matrix representing the sensitivity of $\mathbf{J}_u$ with respect to $\mathbf{d}$.

L in terms of c. For a given $\mathbf{H}$, the nominally optimal CV value is $\mathring{\mathbf{c}}^{\text{nom}} = \mathbf{H}\mathring{\mathbf{y}}^{\text{nom}}$. Similarly, under $\mathbf{u}^{fb}$, the measurements after feedback control are $\mathbf{y}^{fb}$ and the corresponding CVs are $\mathbf{c}^{fb} = \mathbf{H}\mathbf{y}^{fb}$. With these notations, the loss can also be represented as a quadratic function in terms of $\mathbf{c}$.

$$L \approx \frac{1}{2}\mathbf{e}_c^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{e}_c \tag{9}$$

where $\mathbf{e}_c \coloneqq \mathbf{c}^{fb} - \mathring{\mathbf{c}}^{\text{nom}}$ is defined as the CV deviations around the nominally optimal point. $\mathbf{J}_{cc}$ is the Hessian matrix of J with respect to $\mathbf{c}$ evaluated at $\mathring{\mathbf{c}}^{\text{nom}}$

$$\mathbf{J}_{cc} = (\mathbf{H}\mathbf{G}_y)^{-\mathrm{T}}\mathbf{J}_{uu}(\mathbf{H}\mathbf{G}_y)^{-1} \tag{10}$$

and $\mathbf{e}_c$ can be further calculated from $\Delta\mathbf{d}$ and $\mathbf{n}$ based on the linear model (5) and the actual measurement equation (3).

$$\mathbf{e}_c = (-\mathbf{H}\mathbf{G}_y\mathbf{J}_{uu}^{-1}\mathbf{J}_{ud} + \mathbf{H}\mathbf{G}_{yd})\Delta\mathbf{d} + \mathbf{H}\mathbf{n} \tag{11}$$

Remark 1 *In above presentations, L is evaluated in terms of two different references,* $\mathbf{u}$ *and* $\mathbf{c}$*. In the local SOC methods,* $\mathbf{e}_c$ *and* $\mathbf{e}_u$ *can be mutually converted with a linear relationship, namely* $\mathbf{e}_c = (\mathbf{H}\mathbf{G}_y)\mathbf{e}_u$*. In such a situation, the choice of loss evaluation is not critical. In reported local SOC works, L in terms of* $\mathbf{u}$ *was typically used. However, for the global CV selection to be studied in this work, we do not rely on a linearized model hence such simple linear relationship no more holds. Then, it can be shown that the choice of reference for evaluating L is critical in the context of global SOC. This effect will be discussed again in the next section.*

Local average loss minimization

Let $\mathbf{d}$ and $\mathbf{n}$ be scaled as $\Delta\mathbf{d} = \mathbf{W}_d\mathbf{d}'$ and $\mathbf{n} = \mathbf{W}_n\mathbf{n}'$ where $\mathbf{W}_d$ and $\mathbf{W}_n$ are diagonal matrices containing the expected magnitudes of the disturbances and the measurement noises such that the allowable range for $\mathbf{d}'$ and $\mathbf{n}'$ are, for uniform distributions, $\left\| \begin{bmatrix} \mathbf{d}'^{\mathrm{T}} & \mathbf{n}'^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \right\|_\infty \leq 1$, or, for Gaussian distributions $\begin{bmatrix} \mathbf{d}'^{\mathrm{T}} & \mathbf{n}'^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \sim N(0,1)$. Throughout this paper, we assume that all elements in $\mathbf{d}$ and $\mathbf{n}$ are *independently and identically distributed*.

In previous works, both the worst case loss and average loss in the allowable range of uncertainties have been proposed and used as the criteria for CV selection. Kariwala *et al.*[9] derived expressions of local economic loss for both worst case and average loss under different distributions. Practically, the average loss is of more interest and more appropriate to be minimized. Actually, as demonstrated in,[9] minimizing the average loss simultaneously minimizes worst case loss. In this paper,

we aim to minimize the average loss. Follow the approach proposed in,[7] write

$$\Delta \mathring{\mathbf{y}} = \mathbf{F}\Delta \mathbf{d} \tag{12}$$

where $\mathbf{F}$ is defined as the sensitivity of optimal measurements with respect to the disturbances evaluated at the nominally optimal point, which can be numerically calculated through finite differences against $\mathbf{d}$ or through the following analytical expression

$$\mathbf{F} = -\mathbf{G}_y \mathbf{J}_{uu}^{-1} \mathbf{J}_{ud} + \mathbf{G}_{yd} \tag{13}$$

Based on the exact local method,[5] the local average loss L_{lav} was represented as[7]

$$L_{\text{lav}} = \frac{1}{2} \| \begin{bmatrix} \mathbf{M}_d & \mathbf{M}_n \end{bmatrix} \|_F^2 \tag{14}$$

$$\mathbf{M}_d = \mathbf{J}_{uu}^{1/2} (\mathbf{H}\mathbf{G}_y)^{-1} \mathbf{H}\mathbf{F}\mathbf{W}_d, \quad \mathbf{M}_n = \mathbf{J}_{uu}^{1/2} (\mathbf{H}\mathbf{G}_y)^{-1} \mathbf{H}\mathbf{W}_n \tag{15}$$

After introducing an augmented sensitivity matrix $\tilde{\mathbf{F}} \triangleq \begin{bmatrix} \mathbf{F}\mathbf{W}_d & \mathbf{W}_n \end{bmatrix}$ and a constraint of $\mathbf{H}\mathbf{G}_y = \mathbf{J}_{uu}^{1/2}$, the following optimization problem for CV selection is formulated to minimize the local average loss[7]

$$\begin{aligned} \min_{\mathbf{H}} \ & L_{\text{lav}} = \min_{\mathbf{H}} \ \frac{1}{2} \| \tilde{\mathbf{F}}^{\mathrm{T}} \mathbf{H}^{\mathrm{T}} \|_F^2 \\ \text{s.t.} \quad & \mathbf{H}\mathbf{G}_y = \mathbf{J}_{uu}^{1/2} \end{aligned} \tag{16}$$

Analytical solutions to problem (16) are referred to, e.g.,[7].[21]

Remark 2 *The trick to include a constraint utilizes the fact that the solution for optimal* $\mathbf{H}$ *is non-unique. One can verify this by checking the equivalence of using* $\mathbf{H}$ *and* $\mathbf{B}\mathbf{H}$ *as the combination matrix (*$\mathbf{B}$ *is an arbitrary* $n_c \times n_c$ *non-singular matrix). This property allows the designer to consider some additional requirements by imposing certain constraint for* $\mathbf{H}$*. For example, Alstad*

and Skogestad[6] proposed to enforce $\mathbf{HG}_y = \mathbf{I}$*, which achieves a decoupled steady-state response from* $\mathbf{u}$ *to* $\mathbf{c}$*. The same constraint was adopted in.[23] Alstad* et al.*,[7] also in,[21] further considered using* $\mathbf{HG}_y = \mathbf{J}_{uu}^{1/2}$*, which actually achieves* $\mathbf{J}_{cc} = \mathbf{I}$ *(see equation* (10)*), thus significantly simplifying the loss calculation. This constraint will be adopted in the present work due to its convenience.*

Global self-optimizing control

In this paper, we consider the self-optimizing control problem in the whole operating region, which is spanned by all expected uncertainties (disturbances $\mathbf{d}$ and measurements noises $\mathbf{n}$). Prior to introduction of global average loss, a feature of CV selection is reviewed as follows.

Property 1 *Arbitrary CVs with nonzero setpoints (*$\mathbf{c}_s$*) can be generalized as CVs with zeros set-points as follows: define an augmented measurement vector* $\bar{\mathbf{y}} = \begin{bmatrix} y_0 & \mathbf{y}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$*, where* $y_0 = 1$ *is a constant. With* $\bar{\mathbf{y}}$ *as the measurements, the set-points of CVs can be unified to 0 by using an augmented combination matrix* $\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{h}_0 & \mathbf{H} \end{bmatrix}$*, where* $\mathbf{h}_0 = -\mathbf{c}_s$*.*

The correctness of Property 1 can be easily verified through rearrangement of the CV equation $\mathbf{Hy} = \mathbf{c}_s$. To simplify the notation, we use $\mathbf{y}$ for $\bar{\mathbf{y}}$ and, $\mathbf{H}$ for $\bar{\mathbf{H}}$ in the reminder of this work. Dimensions of $\mathbf{y}$ and $\mathbf{H}$ became $(n_y + 1)$ and $(n_c) \times n_y + 1$, respectively. Since we have $n_y \geq n_c$, $\mathbf{H}$ is a strict "fat" matrix in the present work. Using Property 1, arbitrary CVs can be expressed as $\mathbf{c} = \mathbf{Hy}$ with their setpoints at 0 without loss of generality.

Remark 3 *In local SOC methods,* $\mathbf{c}_s$ *is naturally determined by evaluating the CV values at the nominally optimal point to achieve locally zero loss. However, from the global operation point of view, the locally zero loss is not necessary if the overall aim is to achieve the global average loss corresponding to the entire operation range minimised. Therefore, in this paper the economic loss at the nominal point is not necessarily 0. Instead,* $\mathbf{c}_s$ *is a decision variable to minimize the global average loss, which is further incorporated in the combination matrix* $\bar{\mathbf{H}}$*, as indicated by Property 1.*

Global average loss

The average value of economic loss in the whole uncertain space can be expressed as an expectation form

$$L_{\text{gav}}(\mathbf{H}) = E[L(\mathbf{d}, \mathbf{n}, \mathbf{H})] = \int_{\mathbf{d}\in\mathcal{D},\mathbf{n}\in\mathcal{N}} \rho(\mathbf{d})\rho(\mathbf{n})L(\mathbf{d}, \mathbf{n}, \mathbf{H})\, \mathrm{d}\mathbf{n}\mathrm{d}\mathbf{d} \tag{17}$$

where $E[\cdot]$ and $\rho(\cdot)$ represent the expected value and the probability density of a random variable, respectively. $\mathcal{D}$ and $\mathcal{N}$ are the spaces spanned by $\mathbf{d}$ and $\mathbf{n}$, respectively. As seen from eq (6), L is originally defined as a function of $\mathbf{u}$ and $\mathbf{d}$ in open loop. However, since the closed loop performance is considered here, the feedback result of $\mathbf{u}^{fb}$ has to be identified, which depends on $\mathbf{d}$ and $\mathbf{n}$ for specific CVs, so that L is also a function of $\mathbf{d}$, $\mathbf{n}$ and $\mathbf{H}$ (implicitly defined). To calculate L, $\mathbf{u}^{fb}$ (or $\mathbf{y}^{fb}$) have to be determined over the region where the explicit function L lies.

Values of $\mathbf{u}^{fb}$ and $\mathbf{y}^{fb}$ are determined in a closed-loop operation, which is governed by the following equations including the rigorous non-linear process model and linear CVs functions

$$\mathbf{y}^{fb} = \mathbf{f}(\mathbf{u}^{fb}, \mathbf{d}) \tag{18}$$

$$\mathbf{c}_m^{fb} = \mathbf{H}\mathbf{y}_m^{fb} = \mathbf{H}(\mathbf{y}^{fb} + \mathbf{n}) = 0 \tag{19}$$

where $\mathbf{c}_m^{fb}$ is the "measured" CVs assuming to be perfectly maintained at zero through feedback control. The above analysis indicates that directly evaluating the loss requires to solve a set of non-linear equations for a given $\mathbf{H}$. The computation involved is cumbersome.

The formulation of minimizing the average loss for SOC design by taking (17) as the cost function and both (18) and (19) as constraints has been proposed by Halvorsen *et al.*.[5] However, the formulation was generally considered as intractable, hence had not been pursued any further in the literature due to cumbersome computations involved as explained above. Instead, two main simplifications were adopted in[5] and other later works to make the problem tractable. The first simplification was to approximate L by a second-order Taylor expansion and the second one is to

replace $\mathbf{f}$ by a linear model (5), so that L can be represented as an explicit function of $\mathbf{H}$.

Concerning the two simplifications mentioned above, the linear approximation of process model is more likely to result in a large loss when the operating condition drifts far away from the nominal point. In contrast, for a well-designed SOC structure, the average loss over the entire operating range should be reasonable small. Therefore, by taking the reference point to track the optimal operation conditions, a second-order approximation for L is generally acceptable even over the entire operating range. These observations fund the basis for the approaches to be developed in this work.

Loss evaluation using a nonlinear model

To simplify the global SOC problem with non-linear process model (18), it is important to explicitly link the loss with the combination matrix, $\mathbf{H}$. However, this cannot be achieved by taking $\mathbf{u}$ as the independent variable for the Taylor expansion because in this formation, L is implicitly dependent on $\mathbf{H}$ through $\mathbf{u}^{fb}$ determined in non-linear equations (18) and (19).

To address this issue, we choose to evaluate L in terms of $\mathbf{c}$ with $L \approx \frac{1}{2}\mathbf{e}_c^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{e}_c$, which makes L an explicit function of $\{\mathbf{d}, \mathbf{n}, \mathbf{H}\}$, as illustrated as follows. By definition, the CV deviations from the optimum with feedback control is $\mathbf{e}_c = \mathbf{c}^{fb} - \mathring{\mathbf{c}}$. Furthermore, since $\mathbf{c}^{fb} = \mathbf{H}\mathbf{y}^{fb} = \mathbf{H}(\mathbf{y}_m^{fb} - \mathbf{n})$ and the feedback result $\mathbf{H}\mathbf{y}_m^{fb} = 0$, we have $\mathbf{c}^{fb} = -\mathbf{H}\mathbf{n}$. Meanwhile, the optimal CV values can be simply obtained as $\mathring{\mathbf{c}} = \mathbf{H}\mathring{\mathbf{y}}$. Therefore, $\mathbf{e}_c$ can be calculated in a simple form, as

$$\mathbf{e}_c = -\mathbf{H}(\mathring{\mathbf{y}} + \mathbf{n}) \tag{20}$$

For $\mathbf{J}_{cc}$, it is given as

$$\mathbf{J}_{cc} = (\mathbf{H}\mathbf{G}_y)^{-\mathrm{T}}\mathbf{J}_{uu}(\mathbf{H}\mathbf{G}_y)^{-1} \tag{21}$$

Note that for a particular nominal point, $\mathbf{G}_y$ and $\mathbf{J}_{uu}$ in (21) are the same as those in local SOC methods. However, for the global CV design problem, we intend to evaluate it over the entire operation space, not only at a single operating point. In other words, above $\mathbf{G}_y$ and $\mathbf{J}_{uu}$, as well as $\mathring{\mathbf{y}}$, which depend solely on $\mathbf{d}$, can be obtained via off-line optimization for any given scenario of uncertainties. Based on above results, L is evaluated explicitly as

$$L \approx \frac{1}{2}(\mathring{\mathbf{y}} + \mathbf{n})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}(\mathring{\mathbf{y}} + \mathbf{n}) \tag{22}$$

where $\mathbf{J}_{cc}$ is computed from equation (21).

Remark 4 *The difference of choosing appropriate variables for evaluating L can be understood from the perspective of whether the system is operated in open loop or close loop. Using $\mathbf{u}$ is appropriate for the situation when the system is in open loop, because $\mathbf{u}$ are the "free" variables. However, since the loss here is evaluated for a closed-loop operation, $\mathbf{c}$ is preferable because the feedback results can be directly utilized.*

Proposition 1 *The average loss L_{gav} for a given $\mathbf{H}$ can be decomposed as*

$$L_{\mathrm{gav}}(\mathbf{H}) = E(L^d) + E(L^n) \tag{23}$$

where

$$L^d = \frac{1}{2}\mathring{\mathbf{y}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}\mathring{\mathbf{y}},\ L^n = \frac{1}{2}\mathrm{tr}(\mathbf{W}^2\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}) \tag{24}$$

$\mathrm{tr}(\cdot)$ *stands for the trace of a matrix, $\mathbf{W}^2 = E(\mathbf{n}\mathbf{n}^T)$ is diagonal provided that $\mathbf{n}$ are mutually independent.*

Proof: See Appendix A for a proof. ■

In above proposition, L^d can be seen as the loss induced by disturbances $\mathbf{d}$, whilst L^n is the effect of

measurement errors. Note that $\mathbf{n}$ has no influence on L^d, however, $\mathbf{d}$ has not been fully decoupled from L^n because the non-linear term $\mathbf{J}_{cc}$ is embedded in L^n. For $\mathbf{W}^2$, it is easy to show that for Gaussian distributions of $\mathbf{n}$, $\mathbf{W}^2 = \mathbf{W}_n^2$ whilst for uniform distributions, $\mathbf{W}^2 = \frac{1}{3}\mathbf{W}_n^2$, where $\mathbf{W}_n$ is the expected magnitudes of measurement errors as defined earlier. Also see[9] for a proof.

A direct optimizing algorithm

As can be seen from (57), the effect of $\mathbf{n}$ can be exactly estimated via a constant matrix $\mathbf{W}^2$. This happens because $\mathbf{n}$ is an additive terms imposed on $\mathbf{y}$, which further causes a linear deviation of $\mathbf{c}$. However, $\mathbf{y}$ and $\mathbf{J}_{cc}$ are nonlinear functions of $\mathbf{d}$, they do not obey particular distributions, rather, they heavily rely on the process characteristics. Consequently, for general nonlinear processes, both expectations of L^d and L^n cannot be analytically calculated, to the best knowledge of the authors.

Due to the complexity of nonlinear mapping functions, the global average loss can be evaluated through Monte Carlo simulation based on its original definition as indicated below,

$$\begin{aligned} L_{\text{gav}}(\mathbf{H}) &= E(L^d) + E(L^n) \\ &= \int \rho(\mathbf{d})(L^d + L^n)\mathrm{d}\mathbf{d} \\ &\approx \bar{L}_{\text{gav}}(\mathbf{H}) \triangleq \frac{1}{N}\sum_{i=1}^{N}[L^d_{(i)} + L^n_{(i)}] \end{aligned} \tag{25}$$

where N is the total number of sampled disturbance scenarios, the subscript $(\cdot)_{(i)}$ indicates those terms associated with the ith disturbance scenario $\mathbf{d}_{(i)}$, as sampled randomly from expected distributions. Comparing to (17) in the very beginning of definition of L_{gav}, here we just need to integrate the loss in the space of $\mathbf{d}$, because the impact of $\mathbf{n}$ has been analytically derived, which significantly simplifies the problem.

For those processes with discrete disturbance scenarios, which is also possible in real applications, $\bar{L}_{\text{gav}}$ is an exact expression of L_{gav}. However, the more general case of continuous disturbances

are considered in this work, in which case $\bar{L}_{\text{gav}}$ indicates an approximation of L_{gav}through Monte Carlo simulation.[29] Let $\mathbf{d}_{(i)}$ be sampled obeying its own distribution. Then, the approximation error approaches to 0 as $N \to \infty$, as supported by the *law of large numbers*.

Remark 5 *The continuous disturbance space is proposed to be sampled with N finite scenarios and each of them is solved as an NLP problem to minimize the cost. To ensure the global optimality of the CV derived, a reasonably large N is required, which may however make the problem computationally intense. Therefore, to ensure the approach proposed is tractable, it is assumed that an optimization model is readily available such that the global minimum can be solved off-line for all of these scenarios. Since the larger the N, the less the influence of N on the optimality, it is possible to determine an appropriate N off-line. For example, a common way is that we gradually increase N until the difference between successive cases is less than a small tolerance, whist the computation burden is maintained within an acceptable level.*

To sum up, $\bar{L}_{\text{gav}}$ is calculated as follows. Given the distributions of uncertainties ($\mathbf{d}$ and $\mathbf{n}$). Let $\{\mathbf{d}_{(i)}\},\ i = 1, ..., N$ be a sequence of disturbance scenarios which are randomly sampled through Monte Carlo simulation. For each $\mathbf{d}_{(i)}$, the optimization problem of minimizing the cost function J is performed, and $\mathring{\mathbf{y}}_{(i)}$, $\mathbf{J}_{uu,(i)}$ and $\mathbf{G}_{y,(i)}$ are respectively evaluated at the optimum. Their values are stored for calculating the $\bar{L}_{\text{gav}}$ using (58) and (25). Then, the following non-linear programming (NLP) problem is solved for CV selection.

$$\begin{aligned} \min_{\mathbf{H}}\ \bar{L}_{\text{gav}}(\mathbf{H}) &= \min_{\mathbf{H}}\ \frac{1}{N}\sum_{i=1}^{N}(L^d_{(i)} + L^n_{(i)}), \quad i = 1, \ldots, N \\ \text{s.t.} \qquad & \mathbf{H}\mathbf{G}_{y,(k)} = \mathbf{J}^{1/2}_{uu,(k)} \end{aligned} \tag{26}$$

Here, the kth disturbance scenario $\mathbf{d}_{(k)}$ is selected as a particular reference point to form the constraint $\mathbf{H}\mathbf{G}_{y,(k)} = \mathbf{J}^{1/2}_{uu,(k)}$. Typically, we can select the reference point as the nominal point to formulate this constraint. (Strictly speaking, there is no a particular nominal point in the context of global SOC. However, for the sake of comparison with local SOC methods, in this work, we

inherited the term "nominal operating point" and referred it to a particular operating point, which is treated as the nominal one in local SOC approaches.) Based on above developments, the NLP problem in (26) can, in theory, be solved via certain numerical algorithms directly, as proposed in the next.

Algorithm 1: direct numerical optimization. The optimal $\mathbf{H}$ minimizing $\bar{L}_{\text{gav}}$ is directly solved using numerical optimization algorithms, such as the sequential quadratic programming (SQP), interior-point, etc. The detailed procedure is summarized in Table 2.

Table 2: Procedure of implementing Algorithm 1

Step 1: Preparations
1.1 Sampling N finite disturbance scenarios through Monte Carlo simulation
1.2 For each $\mathbf{d}_{(i)}$, calculate and store $\mathring{\mathbf{y}}_{(i)}$, $\mathbf{J}_{uu,(i)}$ and $\mathbf{G}_{y,(i)}$ with off-line optimization (minimizing J)
1.3 $\mathbf{W}^2 = \mathbf{W}_n^2$ or $\frac{1}{3}\mathbf{W}_n^2$ for Gaussian and uniform $\mathbf{n}$, respectively.
Step 2: Performing optimizing scheme
2.1 Define the objective function $\bar{L}_{\text{gav}}$ from (58) and (25)
2.2 Define the constraint: $\mathbf{H}\mathbf{G}_{y,(k)} = \mathbf{J}_{uu,(k)}^{1/2}$, where $\mathbf{G}_{y,(k)}$ and $\mathbf{J}_{uu,(k)}^{1/2}$ are evaluated at the nominal point
2.3 Select initial values for $\mathbf{H}$, which can be selected from random guesses, or CV candidates from other SOC meth
2.4 Apply a numerical optimization solver: SQP, interior-point, etc.

A major drawback for Algorithm 1 is the non-convexity of the NLP problem, which may probably lead to a local optimum. The non-convexity mainly comes from the dependence of $\mathbf{J}_{cc}$ on $\mathbf{H}$ as shown in (21). To find the true optimal solution of the NLP, we have to use global optimization tools, such as the Global Optimization Toolbox in MATLAB, LINGO, etc. However, these tool are either not efficient, for example the Global Optimization Toolbox, which is based on the genetic algorithm and pattern search,[30] or bespoke based on the divide and conquer principles.[31] Fortunately, SOC does not pursue true optimal, but an acceptable performance. Hence, a solution is acceptable as long as it is good enough, particularly if it is better than a local SOC solution although it cannot been proven whether this solution is globally optimal. To achieve this, a local SOC solution can be used as an initial guess for a NLP solver. This will ensure any solution obtained from the NLP solver will be better than the initial local SOC solution.

Remark 6 *It turns out that the constraint* $\mathbf{H}\mathbf{G}_{y,(k)} = \mathbf{J}_{uu,(k)}^{1/2}$ *not only guarantees the uniqueness of*

optimal CVs, but also facilitate solving the optimization problem. In the case studies, the algorithm can be successfully solved with random initial values for searching optimal $\mathbf{H}$*, even though we are not able to prove the convexity of the optimization problem. In contrast, when other constraints are used, the algorithm encounters local optima. The underlying reason can be explained by the fact that this constraint attempts to restrict* $\mathbf{J}_{cc}$ *as a constant, which eliminates, at least in certain extent, the major non-linear source in the loss function.*

A simplified algorithm

In this subsection, we demonstrate that with a further approximation, a sub-optimal solution to problem (26) can be analytical derived in the reformulated framework. Briefly speaking, the algorithm developed here uses a single quadratic model for the loss function as was done in local SOC methods. The quadratic model, however, is fitted using those data sampled from the whole operating region instead of using a linearised model.

Recall that in local SOC, $\mathbf{e}_{c,(i)}$ is linear in terms of $\mathbf{d}$ and $\mathbf{J}_{cc,(i)}$ is a constant for all disturbances (given a specific $\mathbf{H}$). This paper mainly addresses the non-linearity where both $\mathbf{e}_{c,(i)}$ and $\mathbf{J}_{cc,(i)}$ are non-linear mapping functions of $\mathbf{d}_{(i)}$. To further ease the CV selection problem, we will relax one restriction in later developments, by considering $\mathbf{J}_{cc,(i)}$ constant over all disturbance scenarios, similar to local SOC methods. However, $\mathbf{e}_{c,(i)}$ is still calculated using a non-linear model. This way makes sense because $L \approx \frac{1}{2}\mathbf{e}_c^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{e}_c$ is an Euclidean norm of a vector $\mathbf{e}_c$ with a weighting factor $\mathbf{J}_{cc}$. Generally, the Euclidean distance itself is more important than the weighting factor. Note, relaxation of $\mathbf{J}_{cc,(i)}$ as a constant matrix means that the loss is approximated by a quadratic function in terms of the CVs, just as in local SOC. However, the difference is that the approach here linearizes the system differently in the whole operating region and the summation of norm of $\mathbf{e}_{c,(i)}$ is minimized.

As described below, $\mathbf{J}_{cc} = \mathbf{I}$ is adopted to simplify the loss function. However, it is worth to point out that this condition, equivalent to $\mathbf{HG} = \mathbf{J}_{uu}^{1/2}$, is an assumption added to the loss function

$\bar{L}_{\text{gav}}$, not explicitly enforced in the optimization problem. Therefore, to ensure the uniqueness of solution, as well as to enforce all $\mathbf{J}_{cc}$ close to $\mathbf{I}$, an extra constraint, $\mathbf{HG}_{y,(k)} = \mathbf{J}_{uu,(k)}^{1/2}$ at a selected reference point is still required. The nominal point is selected as the reference point in the reminder of this paper.

By approximating $\mathbf{J}_{cc} = \mathbf{I}$, let us firstly consider the expectation of L^n

$$E[L^n] = \frac{1}{2}E[\text{tr}(\mathbf{W}^2\mathbf{H}^{\text{T}}\mathbf{H})] = \frac{1}{2}E[\text{tr}(\mathbf{H}\mathbf{W}^2\mathbf{H}^{\text{T}})] = \frac{1}{2}\|\mathbf{W}\mathbf{H}^{\text{T}}\|_{\text{F}}^2 \tag{27}$$

which is an explicit function of $\mathbf{H}$ and irreverent to $\mathbf{d}$.

For $E(L^d)$ with $\mathbf{J}_{cc} = \mathbf{I}$,

$$E(L^d) = \frac{1}{2}E[\mathring{\mathbf{y}}^{\text{T}}\mathbf{H}^{\text{T}}\mathbf{H}\mathring{\mathbf{y}}] \approx \frac{1}{2N}\sum_{i=1}^{N}\mathring{\mathbf{y}}_{(i)}^{\text{T}}\mathbf{H}^{\text{T}}\mathbf{H}\mathring{\mathbf{y}}_{(i)} \tag{28}$$

Furthermore, in a matrix form, define the following measurement set

$$\mathbf{Y} = \begin{bmatrix} \mathring{\mathbf{y}}_{(1)}^{\text{T}} \\ \mathring{\mathbf{y}}_{(2)}^{\text{T}} \\ \vdots \\ \mathring{\mathbf{y}}_{(N)}^{\text{T}} \end{bmatrix} = \begin{bmatrix} 1 & \mathring{y}_{1,(1)} & \cdots & \mathring{y}_{n_y,(1)} \\ 1 & \mathring{y}_{1,(2)} & \cdots & \mathring{y}_{n_y,(2)} \\ \vdots & \vdots & & \vdots \\ 1 & \mathring{y}_{1,(N)} & \cdots & \mathring{y}_{n_y,(N)} \end{bmatrix} \tag{29}$$

whose ith row vector corresponds to the noise-free measurements sampled at the *optimal status* under $\mathbf{d}_{(i)}$. It can be verified that the $E(L^d)$ in (28) equals to

$$E(L^d) = \frac{1}{2N}\text{tr}(\mathbf{H}\mathbf{Y}^{\text{T}}\mathbf{Y}\mathbf{H}^{\text{T}}) = \frac{1}{2N}\|\mathbf{Y}\mathbf{H}^{\text{T}}\|_{\text{F}}^2 \tag{30}$$

Let $\bar{\bar{L}}_{\text{gav}}$ denote the further approximated average loss under all uncertainties with an enforced

condition $\mathbf{J}_{cc} = \mathbf{I}$. It is derived as follows

$$\begin{aligned}\bar{\bar{L}}_{\text{gav}}(\mathbf{H}) &= E(L^d) + E(L^n) = \frac{1}{2N}\|\mathbf{Y}\mathbf{H}^{\mathrm{T}}\|_{\mathrm{F}}^2 + \frac{1}{2}\|\mathbf{W}_n\mathbf{H}^{\mathrm{T}}\|_{\mathrm{F}}^2 \\ &= \frac{1}{2}\|\tilde{\mathbf{Y}}\mathbf{H}^{\mathrm{T}}\|_{\mathrm{F}}^2 \end{aligned} \tag{31}$$

where the extended matrix $\tilde{\mathbf{Y}}$ is defined as

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \frac{1}{\sqrt{N}}\mathbf{Y} \\ \mathbf{W}_n \end{bmatrix} \tag{32}$$

To sum up, the following optimization problem is formulated with constraint

$$\begin{aligned} \min_{\mathbf{H}} \ & \bar{\bar{L}}_{\text{gav}}(\mathbf{H}) = \frac{1}{2}\|\tilde{\mathbf{Y}}\mathbf{H}^{\mathrm{T}}\|_{\mathrm{F}}^2 \\ \text{s.t.} \quad & \mathbf{H}\mathbf{G}_{y,(k)} = \mathbf{J}_{uu,(k)}^{1/2} \end{aligned} \tag{33}$$

where $\mathbf{G}_{y,(k)}$ and $\mathbf{J}_{uu,(k)}^{1/2}$ can be typically selected as those values evaluated at the nominal point.

This optimization problem is convex with an available analytical solution. Precisely, by making an analogy of above optimization problem to the local SOC approach as presented in equation (16) in Section 2, the only difference is that in global approach, we construct an extended matrix $\tilde{\mathbf{Y}}$ to replace $\tilde{\mathbf{F}}^{\mathrm{T}}$ as in the local SOC approach.[7] Therefore, the next two propositions solving (33) readily follow.

Proposition 2 *The analytical solution solving* (33) *is given as*[7]

$$\mathbf{H}^{\mathrm{T}} = (\tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{Y}})^{-1}\mathbf{G}_{y,(k)}(\mathbf{G}_{y,(k)}^{\mathrm{T}}(\tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{Y}})^{-1}\mathbf{G}_{y,(k)})^{-1}\mathbf{J}_{uu,(k)}^{1/2} \tag{34}$$

Remark 7 *In the work of ,*[7] *above result is not applicable to noise-free cases when* $n_y > n_d$*. This is because when* $\mathbf{W}_n = 0$*,* $\tilde{\mathbf{F}}\tilde{\mathbf{F}}^{\mathrm{T}}$ *loses rank hence the inverse of* $\tilde{\mathbf{F}}\tilde{\mathbf{F}}^{\mathrm{T}}$ *does not exist. However, in the new global approach,* $\tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{Y}}$ *still remains full rank for noise-free cases as long as* $N > (n_y + 1)$

for non-linear systems, which can be easily achieved by increasing the sampling number. Therefore, the proposed approach is also applicable to noise-free case hence more general.

Proposition 3 *A more concise yet equivalent expression for optimal CVs is given as*[21]

$$\mathbf{H}^{\mathrm{T}} = (\tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{Y}})^{-1}\mathbf{G}_{y,(k)} \tag{35}$$

This follows by simply choosing a transformation matrix $\mathbf{B} = (\mathbf{G}_{y,(k)}^{\mathrm{T}}(\tilde{\mathbf{Y}}^{\mathrm{T}}\tilde{\mathbf{Y}})^{-1}\mathbf{G}_{y,(k)})^{-1}\mathbf{J}_{uu,(k)}^{1/2}$, which implies that analytical CVs can be obtained without knowing $\mathbf{J}_{uu,(k)}$. However, if we want to obtain the numerical value of $\bar{\bar{L}}_{\mathrm{gav}}$, $\mathbf{J}_{uu,(k)}$ is still required to calculate $\mathbf{H}$ matching the constraint.[21]

The following algorithm is summarized for minimizing $\bar{\bar{L}}_{\mathrm{gav}}$.

Algorithm 2: Let $\tilde{\mathbf{Y}}$ be constructed as in (32) and (29), the optimal $\mathbf{H}$ minimizing $\bar{\bar{L}}_{\mathrm{gav}}$ are given as (34) or (35). The detailed procedure of Algorithm 2 is summarized in Table 3.

Table 3: Procedure of implementing Algorithm 2

Step 1: Preparations
1.1 Sampling N finite disturbance scenarios through Monte Carlo simulation
1.2 For each $\mathbf{d}_{(i)}$, calculate $\mathring{\mathbf{y}}_{(i)}$ with off-line optimization (minimizing J)
1.3 Evaluate $\mathbf{G}_{y,(k)}$, $\mathbf{J}_{uu,(k)}$ (optional) at the nominal point
1.4 $\mathbf{W}^2 = \mathbf{W}_n^2$ or $\frac{1}{3}\mathbf{W}_n^2$ for Gaussian and uniform $\mathbf{n}$, respectively.
1.5 Constructing $\tilde{\mathbf{Y}}$ from (32) and (29)
Step 2: Performing optimizing scheme
2.1 $\mathbf{H}$ is given as (34) or (35)

Summary and discussions

Approximations in proposed methodology

Within the scope of SOC, a CV selection methodology is developed to minimize the global average loss for non-linear processes. However, due to the problem complexity, several approximations are introduced in the proposed methodology to make the problem tractable:

- The loss function is evaluated with a second-order Taylor expansion with respect to $\mathbf{c}$.
- For those processes with continuous disturbances, the average loss is approximated based on N finite samplings from the disturbance space through Monte Carlo simulation. However, since CVs are designed off-line, we are able to increase N to reduce the approximation error, as long the algorithm can be completed in an affordable time.
- In the simplified Algorithm 2, $\mathbf{J}_{cc}$ is considered to be constant for all sampled disturbance scenarios, which is reasonable, however, introduces some extra errors.

Differences and links to local SOC

At this stage, it is necessary to make a comparison to local SOC methods, based on which some of the results presented in this paper are built. The major difference between the methodology developed in this paper and those local SOC methods is that, we do not rely on a linearized model $\Delta\mathbf{y} = \mathbf{G}_y\Delta\mathbf{u} + \mathbf{G}_{yd}\Delta\mathbf{d}$ but the rigorous non-linear process model $\mathbf{y} = \mathbf{f}(\mathbf{u}, \mathbf{d})$. This treatment is more accurate and helps improve self-optimizing performances for practical plants. Another merit of the proposed approach worth mentioning is that it is more likely to guarantee feasibility when controlling the obtained CVs in the whole operating region. This is because the process is more likely to be operated toward optimum for all occurred disturbances, where feasibility is guaranteed upon optimality. Nevertheless, some challenges to seek a global solution for SOC have to be highlighted as follows:

1. The linear relationship between $\mathbf{e}_u$ and $\mathbf{e}_c$ no more holds. Therefore, choosing the right form of loss evaluation is critical, which was not the case in local SOC methods.
2. The analytical expectation of loss induced by disturbances L^d cannot be easily derived, owing to the uncertain distributions of $\mathring{\mathbf{y}}$ and $\mathbf{J}_{cc}$, which are nonlinear mappings of $\mathbf{d}$. Therefore, the average loss is calculated from its original definition with finite sampling in the whole disturbance space. On the other hand, the effect of measurement/implementation errors are

handled in the same fashion as local SOC methods, because their statistic influences can be exactly estimated.

3. The non-convexity of optimization problem (26) may cause failure (not always but just a risk) if an initial value or numerical optimizing algorithm is not appropriately chosen. On the other hand, the simplified Algorithm 2 is convex where an analytical optimal solution exists, at the price of introducing some extra error. Nevertheless, Algorithm 2 is still stronger than local SOC in the sense that the non-linear term $\mathbf{e}_c$ is preserved, which is generally more important than $\mathbf{J}_{cc}$ when evaluating a loss.

In the next, it is of interest to illustrate that, when the methodology is applied to a linear plant, it will reduce to a local SOC method. To show this, it is firstly noted that for a linear plant, $\mathbf{J}_{cc}$ remains constant for a specific $\mathbf{H}$, hence minimizing $\bar{\bar{L}}_{\text{gav}}$ (Algorithm 2) is a special case of minimizing $\bar{L}_{\text{gav}}$ (Algorithm 1) proposed in this paper. Meanwhile, observing (16), (33) and the definitions of $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{F}}$, it is found that the effect of measurement errors is the same. Thus, we just need to check the equivalence of $\frac{1}{2}\|\frac{1}{\sqrt{N}}\mathbf{Y}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\|_F^2$ (or $E[L^d]$) and $\frac{1}{2}\|(\mathbf{F}\mathbf{W}_d)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\|_F^2$.

For a linear system, we have

$$\mathring{\mathbf{y}} = \mathbf{F}\Delta\mathbf{d} + \mathbf{y}^{\text{nom}} \tag{36}$$

therefore

$$\begin{aligned} E(L^d) &= \frac{1}{2}E(\mathring{\mathbf{y}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\mathring{\mathbf{y}}) = \frac{1}{2}E[\mathrm{tr}(\mathring{\mathbf{y}}\mathring{\mathbf{y}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H})] \\ &= \frac{1}{2}E\{\mathrm{tr}[(\mathbf{F}\Delta\mathbf{d}\Delta\mathbf{d}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}} + \mathbf{F}\Delta\mathbf{d}(\mathbf{y}^{\text{nom}})^{\mathrm{T}} + \mathbf{y}^{\text{nom}}(\mathbf{F}\Delta\mathbf{d})^{\mathrm{T}} + \mathbf{y}^{\text{nom}}(\mathbf{y}^{\text{nom}})^{\mathrm{T}})\mathbf{H}^{\mathrm{T}}\mathbf{H}]\} \end{aligned} \tag{37}$$

where since the expectation of $\Delta\mathbf{d}$ is 0, the middle two terms vanish. Furthermore,

$$E[\mathrm{tr}(\mathbf{y}^{\text{nom}}(\mathbf{y}^{\text{nom}})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H})] = (\mathbf{y}^{\text{nom}})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\mathbf{y}^{\text{nom}} = 0 \tag{38}$$

The equality holds because in the linear case, we select $\mathbf{H}$ such that at the nominal point the loss is 0. Note, we complement the set of measurements with a constant 1, the set-points are also decision variables in $\mathbf{H}$. Consequently,

$$E(L^d) = \frac{1}{2}E[\mathrm{tr}(\mathbf{F}\Delta\mathbf{d}\Delta\mathbf{d}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H})] = \frac{1}{2}\|(\mathbf{F}\mathbf{W}_d)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\|_F^2 \tag{39}$$

which proves the equivalence.

The only restriction is that in the proposed algorithms, $E(L^d)$ is estimated via finite sampling. However, this approximation error is small when N is large. On the other hand, it is very rare for a practical process to be exactly linear. Hence, the profit gained by non-linear treatment is generally larger than the extra economic loss induced by finite sampling.

Case studies

A toy example

Consider the following toy example

$$J = \frac{1}{2}(u-d)^2 \tag{40}$$

where both u and d are scalars. Two measurements are available as follows

$$y_1 = u \tag{41}$$

$$y_2 = \frac{1}{4}u^2 + d \tag{42}$$

The nominal disturbance is $d^{\mathrm{nom}} = 0$. Correspondingly, the optimal point is then defined by $u^{\mathrm{nom}} = 0$, where $J^{\mathrm{nom}} = 0$, $y_1^{\mathrm{nom}} = 0$ and $y_2^{\mathrm{nom}} = 0$. The possible variation of d is uniformly distributed between -1 and 1, *i.e.* $d \in \begin{bmatrix} -1 & 1 \end{bmatrix}$. Measurement noises are not considered in this

toy example. For self-optimizing control of this toy example, we consider

$$c = \mathbf{hy} = h_0 + h_1 y_1 + h_2 y_2 \tag{43}$$

as the controlled variable.

To more clearly illustrate the proposed algorithms, let us firstly consider a naive scenario by selecting $N = 3$. Furthermore, the disturbance range is evenly discreted as $d_{(i)} \in \{-1, 0, 1\}$. Since the optimum of this toy example is given as $\mathring{u}_{(i)} = d_{(i)}$. The optimal measurements $\mathring{\mathbf{y}}$ and their first order derivatives $\mathbf{G_y}$ for each scenario are easily obtained. For all scenarios, $J_{uu} = 1$.

$$\mathbf{Y} = \begin{bmatrix} \mathring{\mathbf{y}}^{\mathrm{T}}_{(1)} \\ \mathring{\mathbf{y}}^{\mathrm{T}}_{(2)} \\ \mathring{\mathbf{y}}^{\mathrm{T}}_{(3)} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -0.75 \\ 1 & 0 & 0 \\ 1 & 1 & 1.25 \end{bmatrix}, \qquad \begin{bmatrix} \mathbf{G}^{\mathrm{T}}_{y,(1)} \\ \mathbf{G}^{\mathrm{T}}_{y,(2)} \\ \mathbf{G}^{\mathrm{T}}_{y,(3)} \end{bmatrix} = \begin{bmatrix} 0 & 1 & -0.5 \\ 0 & 1 & 0 \\ 0 & 1 & 0.5 \end{bmatrix} \tag{44}$$

Note an augmented measurement $y_0 = 1$ is added.

(1) Algorithm 1. The objective function is given as $\bar{L}_{\mathrm{gav}} = \frac{1}{6}\sum_{i=1}^{3}(\mathring{\mathbf{y}}^{\mathrm{T}}_{(i)}\mathbf{h}^{\mathrm{T}}\mathbf{J}_{cc,(i)}\mathbf{h}\mathring{\mathbf{y}}_{(i)})$, where $J_{cc,(i)} = (\mathbf{hG}_{y,(i)})^{-\mathrm{T}}\mathbf{J}_{uu}(\mathbf{hG}_{y,(i)})^{-1}$. Using the values in (44), $\bar{L}_{\mathrm{gav}}$ is rearranged as

$$\bar{L}_{\mathrm{gav}} = \frac{1}{6}\left\{\frac{(h_0 - h_1 - 0.75h_2)^2}{(h_1 - 0.5h_2)^2} + \frac{h_0^2}{h_1^2} + \frac{(h_0 + h_1 + 1.25h_2)^2}{(h_1 + 0.5h_2)^2}\right\} \tag{45}$$

The constraint is selected as $\mathbf{hG}_{y,(2)} = J_{uu}$, i.e. $h_1 = 1$. The SQP method is used for minimizing $\bar{L}_{\mathrm{gav}}$, the solution gives a CV as

$$c_1 = 0.1348 + y_1 - 0.9201y_2 \tag{46}$$

with the minimal $\bar{L}_{\mathrm{gav}} = 0.005561$.

(2) Algorithm 2. Since no measurement errors involved, the extended measurement set $\tilde{\mathbf{Y}}$ is a

constant factor $\frac{1}{\sqrt{3}}$ of $\mathbf{Y}$. Using Proposition 2, the optimal CV minimizing $\bar{\bar{L}}_{\text{gav}}$ is

$$c_2 = 0.1633 + y_1 - 0.9796 y_2 \tag{47}$$

with the minimal $\bar{\bar{L}}_{\text{gav}} = 0.006803$.

Comment: For Algorithm 2, different solutions will be obtained when other reference points are used to enforce the constraint. However, such differences are generally small. Here, in the case of using the other two points ($i = 1, 3$), the calculated self-optimizing performances are almost identical to using the nominal point (relatively less than 0.8%).

(3) Local SOC. Since there are no measurement noises, the local null space method is applied,[6] the resultant CV is

$$c_{\text{loc}} = y_1 - y_2 \tag{48}$$

(4) Analytical optimal CV. For this toy example, the analytical optimal CV can actually be derived using Matlab Symbolic Toolbox[32]

$$c_{\text{ana}} = 0.0705 + y_1 - 0.9231 y_2 \tag{49}$$

Above results are summarized in Table 4, from which the following results can be observed: (1) As expected, among 4 CVs, Algorithm 1 and Algorithm 2 minimizes $\bar{L}_{\text{gav}}$ (0.005561) and $\bar{\bar{L}}_{\text{gav}}$ (0.006803), respectively. (2) Both Algorithm 1 and Algorithm 2 are effective as compared to the local SOC method, the true average loss (analytically calculated for this toy example) for them are 0.005431 and 0.006344, respectively. The performance of Algorithm 1 is better than Algorithm 2. On the other hand, the local SOC gives a big average loss as 0.01831. (3) There are still some loss between the true average loss and the proposed algorithms, which can be improved by increasing N.

The performances of increasing N from 3 and 30 are shown in Figure 1. For Algorithm 1, it

Table 4: Performance comparisons for toy example

Algorithms	CV	$\bar{L}_{\text{gav}}$	$\bar{\bar{L}}_{\text{gav}}$	L_{gav}
Algorithm 1	c_1	**0.005561**	0.008179	0.005431
Algorithm 2	c_2	0.007624	**0.006803**	0.006344
Local SOC	c_{loc}	0.04630	0.02083	0.01831
Analytical optimal	c_{ana}	0.009213	0.0136	**0.002778**

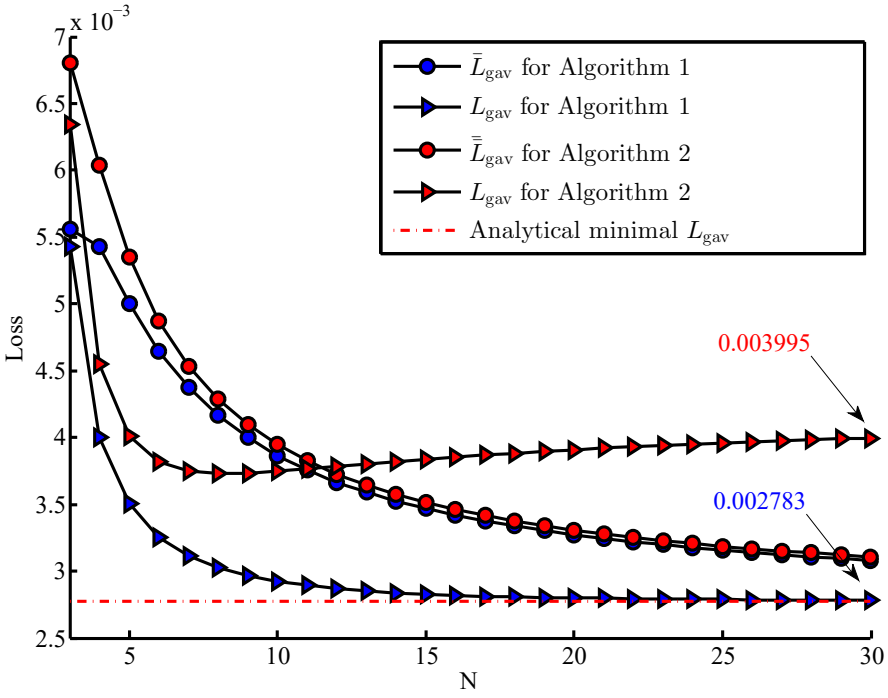

Figure 1: Performance validation of increasing N

does not have multiple local optima in a wide range of the solution space. For various starting points, the algorithm converges to the same optimal solution. Both the losses $\bar{L}_{\text{gav}}$ and true average loss L_{gav} gradually decrease as we increase N. At $N = 30$, $\bar{L}_{\text{gav}}$ (0.002783) is almost identical to the true optimal 0.002778, where the slight difference stems from the approximation error of Taylor expansion. For Algorithm 2, it is noted that $\bar{\bar{L}}_{\text{gav}}$ gradually decreases, however, L_{gav} decreases at first but slightly increase when N continues to grow. This can be explained by different effects imposed on the estimated average loss by N and treating J_{cc} as a constant. Increasing N is generally beneficial, hence all the losses decrease except L_{gav} for Algorithm 2, which is due to the enforced condition $\mathbf{J}_{cc} = \mathbf{I}$, the true global loss increases in spite of the decreasing simplified loss, $\bar{\bar{L}}_{\text{gav}}$. However, such a behaviour is not general, but case dependent. For this toy example, interpretation can be provided as follows: according to (17) and the calculation of $\bar{L}_{\text{gav}}$, Algorithm 1 is equivalent to implementing Algorithm 2 but with samples obeying a distribution of $\tilde{\rho}(d) = \frac{\rho(d)J_{cc}(d)}{\int \rho(d)J_{cc}(d)\mathrm{d}d}$. Simulation result shows that $J_{cc}(d)$ monotonically grows with d and its mean around $d = 0.7$. Hence, using data generated under original uniform distribution, the true global loss may increase when N increase. This is because more samples under a wrong distribution cause more errors. Nevertheless, the final loss (0.003995) is still much smaller than the local SOC method (0.01831).

Evaporator

A forced-circulation evaporator is investigated, which was modified from its original version,[33] as shown in Figure 2. This example has also been investigated in several papers,[9,21,23] however, some results provided here are new. The process involves 3 state variables and 20 process variables, which are explained in Table 5 with their nominal values. For more process descriptions and model equations, we also refer readers to other publications.[9,33]

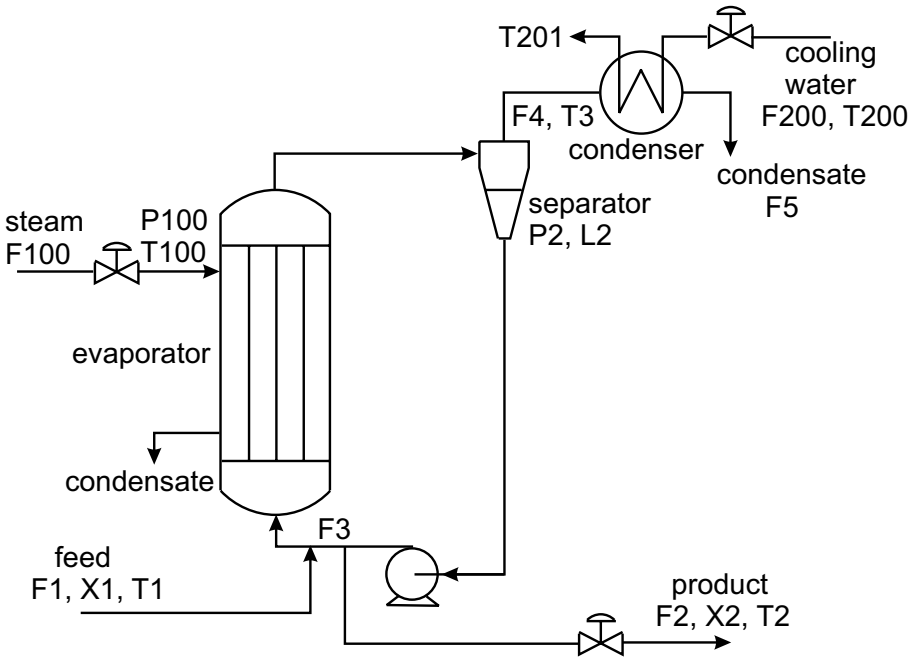

Figure 2: Evaporator

Table 5: Process variables and nominal values for evaporator process

Variable	Physical description	Nominal value	Unit
F_1	Feed flowrate	9.469	kg·min^{-1}
F_2	Product flowrate	1.334	kg·min^{-1}
F_3	Circulating flowrate	24.721	kg·min^{-1}
F_4	Vapor flowrate	8.135	kg·min^{-1}
F_5	Condensate flowrate	8.135	kg·min^{-1}
X_1	Feed composition	5	%
X_2	Product composition	35.5	%
T_1	Feed temperature	40	°C
T_2	Product temperature	88.400	°C
T_3	Vapor temperature	81.066	°C
L_2	Separator level	1	meter
P_2	Operating pressure	51.412	kPa
F_{100}	Steam flowrate	9.434	kg·min^{-1}
T_{100}	Steam temperature	151.52	°C
P_{100}	Steam pressure	400	kPa
Q_{100}	Heat duty	345.292	kW
F_{200}	Cooling water flowrate	217.738	kg·min^{-1}
T_{200}	Inlet cooling water temperature	25	°C
T_{201}	Outlet cooling water temperature	45.550	°C
Q_{200}	Condenser duty	313.210	kW

The objective function is described as

$$J = 600F_{100} + 0.6F_{200} + 1.009(F_2 + F_3) + 0.2F_1 - 4800F_2 \tag{50}$$

with manipulated variables and disturbances as

$$\mathbf{u} = \begin{bmatrix} F_{200} & F_1 \end{bmatrix}^T \tag{51}$$

$$\mathbf{d} = \begin{bmatrix} X_1 & T_1 & T_{200} \end{bmatrix}^T \tag{52}$$

where the variation ranges for disturbances are defined as $\pm 5\%$ for X_1 and $\pm 20\%$ for T_1 and T_{200} (uniform) of their nominal values, respectively. Here, the considered $\mathbf{u}$ is a subset of original manipulated variables, other degrees of freedom are used to satisfy the active process constraints. In the present work, we consider the reduced unconstrained optimizing problem.

The following measurements are available

$$\mathbf{y} = \begin{bmatrix} P_2 & T_2 & T_3 & F_2 & F_{100} & T_{201} & F_3 & F_5 & F_{200} & F_1 \end{bmatrix}^T \tag{53}$$

with expected noise magnitudes of 2%, 2.5% and 1°C (uniform) for flowrates, pressures and temperatures, respectively. At the nominal point, the following matrices are obtained

$$\begin{aligned} \mathbf{J}_{uu} &= \begin{bmatrix} 0.0056 & -0.1334 \\ -0.1334 & 16.7366 \end{bmatrix}, \\ \mathbf{G}_y &= \begin{bmatrix} -0.093 & -0.052 & -0.047 & 0 & -0.001 & -0.094 & -0.032 & 0 & 1 & 0 \\ 11.678 & 6.559 & 5.921 & 0.141 & 1.115 & 2.170 & 6.594 & 0.859 & 0 & 1 \end{bmatrix}^{\mathrm{T}} \end{aligned} \tag{54}$$

which are incorporated into the constraint condition for the combination matrix of CVs.

In this evaporator case study, we investigate the measurement subset selection problem via applying Algorithm 2, where $\bar{\bar{L}}_{\mathrm{gav}}$ is used as the criterion for subset screening. This is because the analytical

minimal $\bar{\bar{L}}_{\text{gav}}$ can be computed in one step, once the matrix $\tilde{\mathbf{Y}}$ has been constructed, through which means one can quickly find promising measurement subsets. For this scale of problem, exhaustive search for optimal subset is quickly completed, note that, faster MIQP and BAB algorithms are also available.[19,22]

Table 6: Best measurement subset selection for self-optimizing control of evaporator

n_y	Best subset (minimal $\bar{\bar{L}}_{\text{gav}}$)	$\bar{\bar{L}}_{\text{gav}}$	Best subset (minimal $\frac{1}{2}\|\tilde{\mathbf{F}}^{\text{T}}\mathbf{H}^{\text{T}}\|_F^2$)	$\bar{\bar{L}}_{\text{gav}}$
2	$[F_{100}\ F_{200}]$	19.150	$[F_3\ F_{200}]$	20.694
3	$[F_2\ F_{100}\ F_{200}]$	4.0214	-[a]	-
4	$[F_2\ F_{100}\ F_5\ F_{200}]$	3.5303	$[F_2\ T_{201}\ F_3\ F_{200}]$	13.706
5	$[F_2\ F_{100}\ F_3\ F_5\ F_{200}]$	3.1190	$[F_2\ F_{100}\ T_{201}\ F_3\ F_{200}]$	3.6269
6	$[F_2\ F_{100}\ F_3\ F_5\ F_{200}\ F_1]$	3.0742	$[F_2\ F_{100}\ T_{201}\ F_3\ F_5\ F_{200}]$	3.1005
7	$[P_2\ F_2\ F_{100}\ T_{201}\ F_3\ F_5\ F_{200}]$	3.0389	-	-
8	$[P_2\ F_2\ F_{100}\ T_{201}\ F_3\ F_5\ F_{200}\ F_1]$	3.0153	$[P_2\ T_2\ F_2\ F_{100}\ T_{201}\ F_3\ F_5\ F_{200}]$	3.0328
9	$[P_2\ T_2\ F_2\ F_{100}\ T_{201}\ F_3\ F_5\ F_{200}\ F_1]$	3.0106	-	-
10	$[P_2\ T_2\ T_3\ F_2\ F_{100}\ T_{201}\ F_3\ F_5\ F_{200}\ F_1]$	3.0083	-	-

[a] The two approaches give the same optimal measurement subset

The disturbance space is randomly sampled by $N = 1000$ uniformly distributed scenarios, which is tested to be enough for this example. Using the number of measurements ranging from 2–10, the best subsets with minimal $\bar{\bar{L}}_{\text{gav}}$ are obtained. The best measurement subsets and their corresponding $\bar{\bar{L}}_{\text{gav}}$ values are given in Table 6. Detailed CVs are provided in Appendix B (Table 8).

Basically, using $\bar{\bar{L}}_{\text{gav}}$ as the criterion for subset selection leads to some similar conclusions as revealed earlier with a local SOC method.[9] That is, the loss index gradually decreases as the number of measurements n_y increases from 2 to 10, as shown in Figure 3. Using 2 measurements is not good enough for self-optimizing controlling this evaporator, which gives an average loss of 19.150. This can be explained by the fact that there are 3 disturbances affecting the process, hence using 2 measurements is not enough to achieve good self-optimizing performance. On the other hand, when $n_y \geq 3$, the loss can be substantially reduced as expected. Among all, $n_y = 10$ gives the minimal average loss of 3.0083. Nonetheless, it is obvious that using all measurements is not necessary, in contrast, $n_y = 3$ or 4 gives a relatively good trade-off between the economic performance and CV complexity.

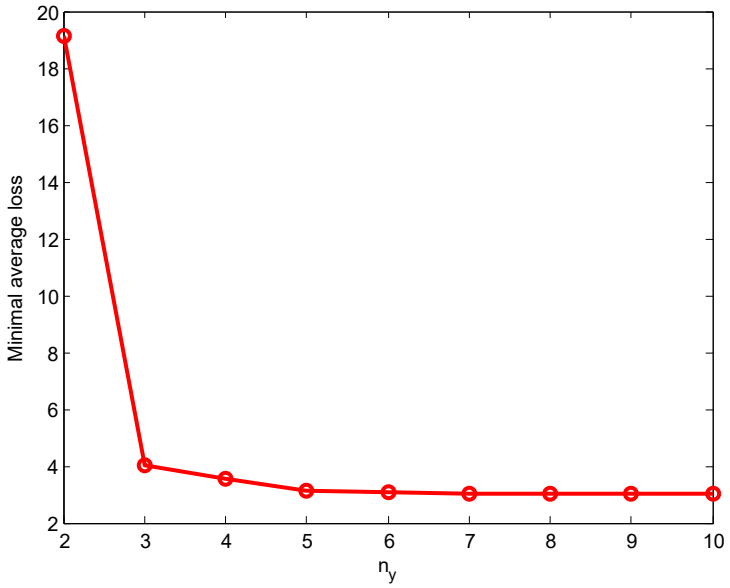

Figure 3: Minimal average losses obtained from Algorithm 2 with 2 to 10 measurements

For comparisons, the best measurement subsets are also identified by using the local SOC method, where $\frac{1}{2}\|\tilde{\mathbf{F}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\|_F^2$ is used as an index for subset selection. The results are compared in Table 6. Firstly, the obtained subsets with a local SOC analysis are the same as in,[21] where a Mixed Integer Quadratic Programming (MIQP) algorithm was used for subset selection. However, when compared to the results obtained by the algorithm proposed in this paper, some differences can be observed. When $n_y = 2, 4, 5, 6, 8$, the proposed algorithm gives different optimal measurement subsets, as shown in Table 6. For other cases when $n_y = 3, 7, 9, 10$, they gave the same optimal subsets.

In all scenarios, the proposed method suggests include F_{100} and F_{200}, instead of F_3 and F_{200} as encouraged by a local SOC analysis. With non-linear model verification of 100 random uncertainties, we find that the average loss can be reduced from 20.032 to 15.857 by using $[F_{100}, F_{200}]$ instead of $[F_3, F_{200}]$. In the work of,[9] $[F_{100}, F_{200}]$ is the third best subset when 2 individual measurements are controlled. However, they did not identify it as the best choice and implement it for non-linear model verification.

The largest discrepancy occurs when $n_y = 4$, the proposed approach recommends using $[F_2\ F_{100}\ F_5\ F_{200}]$ with a minimal $\bar{\bar{L}}_{\mathrm{gav}} = 3.5303$. In contrast, the local SOC method gives a subset of $[F_2\ T_{201}\ F_3\ F_{200}]$,

whose $\bar{\bar{L}}_{\text{gav}}$ index is calculated to be 13.706, indicating that it is a bad choice for 4-measurement subset. Actually, these two subsets have been investigated in,[9] they verified that $[F_2\ T_{201}\ F_3\ F_{200}]$ results in a poor performance although it has a locally minimal loss. Alternatively, they found that $[F_2\ F_{100}\ F_5\ F_{200}]$ is a very promising candidate via an ad hoc search, where they reported a non-linear model evaluated average loss of 2.808. In this paper, we further confirm it as the best subset using the $\bar{\bar{L}}_{\text{gav}}$ index. For other cases where the two algorithms give different subsets ($n_y = 5, 6, 8$), the $\bar{\bar{L}}_{\text{gav}}$ values differ not much.

In the following, we further explore the the cases of $n_y = 2$ and 4 using best measurement subsets identified in this paper, by applying both Algorithm 1 and 2, as well as the local SOC method. Their non-linear model evaluated average losses with 100 groups of random disturbances and measurement errors are tabulated in Table 7. When $n_y = 2$, these algorithms give almost identical results, with negligible differences. When $n_y = 4$, the proposed algorithms behave relatively better with smaller average losses, as shown in Table 7. Above results are actually not surprising, because the model between identified measurements (F_{100}, F_{200}) and system inputs can be verified to be very linear. In this situation, the differences among various algorithms are not significant. Besides, for the case of $n_y = 4$, the overall average loss $\bar{L}_{\text{gav}}$ for different CVs (e.g. obtained by Algorithm 1) is decomposed as $L^d = 0.809$ and $L^n = 3.531$, which indicate that the measurement errors contribute most of the economic loss. Since the Algorithm1 and 2 are mainly proposed to reduce the loss caused by the disturbances, the overall economic performance is only slightly improved. Nevertheless, it is the new algorithm proposed in this work that identifies these good subsets successfully.

Table 7: Average loss evaluations with 100 groups of random disturbances and measurement errors for $n_y = 2$ and 4

n_y	Measurement subset	Algorithm 1	Algorithm 2	Local SOC
2	$[F_{100}\ F_{200}]$	15.859	15.857	15.866
4	$[F_2\ F_{100}\ F_5\ F_{200}]$	3.760	3.802	3.921

With further explorations, we demonstrate that even with comparable average losses, Algorithm 1 and 2 still give more reliable estimate of the economic loss. For the simulated 100 groups of random uncertainties, we calculate the single loss using different formulas (with the same **H**

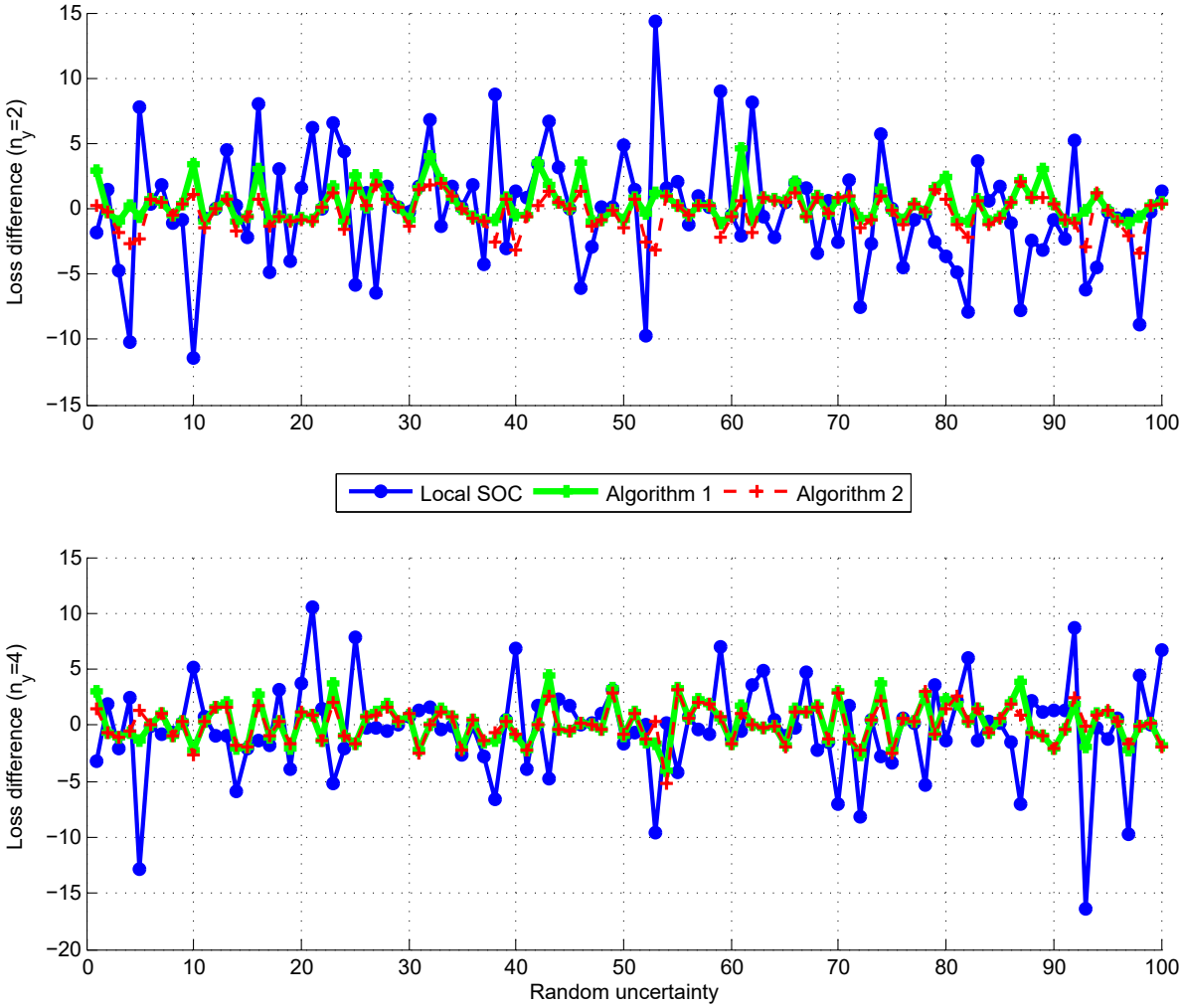

Figure 4: Differences between loss estimates from various algorithms and the true evaluated losses

results from Algorithm 2 for illustration), namely $L_1 = \frac{1}{2}(\mathring{\mathbf{y}} + \mathbf{n})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}(\mathring{\mathbf{y}} + \mathbf{n})$ (Algorithm 1), $L_2 = \frac{1}{2}(\mathring{\mathbf{y}} + \mathbf{n})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}(\mathring{\mathbf{y}} + \mathbf{n})$ (Algorithm 2) and, $L_{loc} = \frac{1}{2}\mathbf{e}_u^{\mathrm{T}}\mathbf{J}_{uu}\mathbf{e}_u$, where $\mathbf{e}_u$ is calculate from (8). Their differences between the true non-linear model evaluated loss are shown in Figure 4, where in the case of $n_y = 2$, the standard deviations of the loss difference are 1.266, 1.267 and 4.200 for Algorithm 1, 2 and the local SOC, respectively. In the other case of $n_y = 4$, the standard deviations are 1.356, 1.361 and 4.857. These results show that both L_1 and L_2 provide much more accurate estimation of the true economic loss, in other words, the relative good performance of local SOC method stems partially from the fact that, the effects of those not well estimated losses are counteracting with each other, because this evaporator behaves somewhat "symmetrically" around the nominal point. However, this feature may not be generally expected for other plants.

Finally, as a summary for this evaporator case study, we demonstrate that the proposed methodology is more suitable for measurement subset selection. Besides, it gives more reliable estimate of economic loss by using the formula as presented in this paper.

Conclusions

This paper proposed a global solution of the self-optimizing control problem, which aims to minimize the average loss under uncertain operating conditions. A major characteristic of the methodology developed in this paper is that we do not rely on a linearized process model, rather, we use the original non-linear model. During the developments, the economic loss was proposed to be evaluated in terms of $\mathbf{c}$, such that the loss can be expressed as an explicit function of combination matrix. Two algorithms for solving the global SOC problem were developed, namely a direct optimizing scheme and an analytical approach by treating $\mathbf{J}_{cc}$ under all scenarios as the same. Differences and links of the proposed methodology to local SOC approaches were also discussed.

Two case studies were provided with applications of the presented methodology. The results showed that both the algorithms proposed were effective in the context of global SOC problem. In the

simulated chemical process, Algorithm 2 exhibited no worse (even better) performances than Algorithm 1, even though the former is a further simplified formulation. Since the analytical optimal solution can be readily derived, Algorithm 2 is also particularly suitable for applying SOC strategy, namely, CV selection to fast screening of measurement subset candidates, as was done in the evaporator example.

Acknowledgement

The first and third authors gratefully acknowledge the financial support from the National Natural Science Foundation of China (NSFC) (61304081), Zhejiang Provincial Natural Science Foundation of China (LQ13F030007), National Project 973 (2012CB720500) and Ningbo Innovation Team (2012B82002,2013B82005).

References

(1) Chachuat, B.; Srinivasan, B.; Bonvin, D. Adaptation strategies for real-time optimization. *Comput. Chem. Eng.* **2009**, *33*, 1557–1567.

(2) Marchetti, A.; Chachuat, B.; Bonvin, D. Modifier-Adaptation Methodology for Real-Time Optimization. *Ind. Eng. Chem. Res.* **2009**, *48*, 6022–6033.

(3) Skogestad, S. Plantwide control: The search for the self-optimizing control structure. *J. Process Control* **2000**, *10*, 487–507.

(4) Srinivasan, B.; Bonvin, D.; Visser, E.; Palanki, S. Dynamic optimization of batch processes - II. Role of measurements in handling uncertainty. *Comput. Chem. Eng.* **2003**, *27*, 27–44.

(5) Halvorsen, I. J.; Skogestad, S.; Morud, J. C.; Alstad, V. Optimal selection of controlled variables. *Ind. Eng. Chem. Res.* **2003**, *42*, 3273–3284.

(6) Alstad, V.; Skogestad, S. Null space method for selecting optimal measurement combinations as controlled variables. *Ind. Eng. Chem. Res.* **2007**, *46*, 846–853.

(7) Alstad, V.; Skogestad, S.; Hori, E. S. Optimal measurement combinations as controlled variables. *J. Process Control* **2009**, *19*, 138–148.

(8) Kariwala, V. Optimal Measurement Combination for Local Self-Optimizing Control. *Ind. Eng. Chem. Res.* **2007**, *46*, 3629–3634.

(9) Kariwala, V.; Cao, Y.; Janardhanan, S. Local Self-Optimizing Control with Average Loss Minimization. *Ind. Eng. Chem. Res.* **2008**, *47*, 1150–1158.

(10) Ye, L.; Cao, Y.; Li, Y.; Song, Z. Approximating Necessary Conditions of Optimality as Controlled Variables. *Ind. Eng. Chem. Res.* **2013**, *52*, 798–808.

(11) Jaschke, J.; Skogestad, S. Optimal controlled variables for polynomial systems. *J. Process Control* **2012**, *22*, 167 – 179.

(12) Dahl-Olsen, H.; Narasimhan, S.; Skogestad, S. Optimal output selection for control of batch processes. Proc. of American Control Conference. 2008.

(13) Hu, W.; Mao, J.; Xiao, G.; Kariwala, V. Selection of Self-Optimizing Controlled Variables for Dynamic Processes. 8th Intl. Symposium on ADCHEM. 2012; pp 774–779.

(14) Jaschke, J.; Fikar, S., M.and Skogestad Self-optimizing invariants in dynamic optimization. CDC-ECE. 2011; pp 7753–7758.

(15) Ye, L.; Kariwala, V.; Cao, Y. Dynamic optimization for batch processes with uncertainties via approximating invariant. the 8th IEEE Conference on Industrial Electronics and Applications (ICIEA). 2013; pp 1786–1791.

(16) Kariwala, V.; Skogestad, S. Branch and Bound Methods for Control Structure Design. Proc. 16th ESCAPE and 9th Intl. Symposium on PSE. Garmisch-Partenkirchen, Germany, 2006.

(17) Cao, Y.; Kariwala, V. Bidirectional branch and bound for controlled variable selection: Part I. Principles and minimum singular value criterion. *Comput. Chem. Engng.* **2008**, *32*, 2306–2319.

(18) Kariwala, V.; Cao, Y. Bidirectional branch and bound for controlled variable selection: Part II. Exact Local Method for Self-optimizing Control. *Comput. Chem. Eng.* **2009**, *33*, 1402–1412.

(19) Kariwala, V.; Cao, Y. Bidirectional branch and bound for controlled variable selection: Part III. Local Average Loss Minimization. *IEEE Trans. Ind. Informat.* **2010**, *6*, 54–61.

(20) Kariwala, V.; Ye, L.; Cao, Y. Branch and bound method for regression-based controlled variable selection. *Comput. Chem. Eng.* **2013**, *54*, 1–7.

(21) Yelchuru, R.; Skogestad, S. Convex formulations for optimal selection of controlled variables and measurements using Mixed Integer Quadratic Programming. *J. Proc. Control* **2012**, *22*, 995–1007.

(22) Manum, H.; Skogestad, S. Self-optimizing control with active set changes. *J. Process Control* **2012**, *22*, 873–883.

(23) Hu, W.; Umar, L. M.; Xiao, G.; Kariwala, V. Local self-optimizing control of constrained processes. *J. Process Control* **2012**, *22*, 488–493.

(24) Ye, L.; Cao, Y.; Li, Y.; Song, Z. A Data-driven Approach for Selecting Controlled Variables. 8th Intl. Symposium on ADCHEM. Singapore, 2012; pp 904–909.

(25) Jaschke, J.; Skogestad, S. NCO tracking and self-optimizing control in the context of real-time optimization. *J. Process Control* **2011**, *21*, 1407–1416.

(26) Francois, G.; Srinivasan, B.; Bonvin, D. Use of measurements for enforcing the necessary conditions of optimality in the presence of constraints and uncertainty. *J. Process Control* **2005**, *15*, 701–712.

(27) Ye, L.; Cao, Y.; Ma, X.; Song, Z. A Novel Hierarchical Control Structure with Controlled Variable Adaptation. *Ind. Eng. Chem. Res.* **2014**, *53*, 14695–14711.

(28) Rangaiah, G. P.; Kariwala, V. *Plantwide control: Recent developments and applications*; John Wiley & Sons, 2012.

(29) Sakalauskas, L. L. Nonlinear stochastic programming by Monte-Carlo estimators. *European Journal of Operational Research* **2002**, *137*, 558–573.

(30) *The Global Optimization Toolbox User's Guide*; The MathWorks, 2015.

(31) Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. *Introduction to algorithms*, 3rd ed.; MIT press, 2009.

(32) Ye, L.; Cao, Y. A formulation for globally optimal controlled variable selection. UKACC International Conference on Control. 2012; pp 136–141.

(33) Newell, R. B.; Lee, P. *Applied process control : a case study*; Prentice-Hall, 1989.

Proof for Proposition 1

Based on (22), the average loss L_{gav} is expected as

$$\begin{aligned} L_{\text{gav}}(\mathbf{H}) &= E[L(\mathbf{d},\mathbf{n},\mathbf{H})] \\ &= \frac{1}{2}E\left\{(\mathring{\mathbf{y}}+\mathbf{n})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}(\mathring{\mathbf{y}}+\mathbf{n})\right\} \\ &= \frac{1}{2}E\left\{\mathrm{tr}[(\mathring{\mathbf{y}}+\mathbf{n})(\mathring{\mathbf{y}}+\mathbf{n})^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}]\right\} \\ &= \frac{1}{2}E[\mathring{\mathbf{y}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}\mathring{\mathbf{y}}] + \frac{1}{2}E[\mathrm{tr}(\mathring{\mathbf{y}}\mathbf{n}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H})] + \frac{1}{2}E[\mathrm{tr}(\mathbf{n}\mathring{\mathbf{y}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H})] + \frac{1}{2}E[\mathrm{tr}(\mathbf{n}\mathbf{n}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H})] \end{aligned} \tag{55}$$

For the second term,

$$E[\mathrm{tr}(\mathring{\mathbf{y}}\mathbf{n}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H})] = E[\mathbf{n}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}\mathring{\mathbf{y}}] = 0 \tag{56}$$

This is because both $\mathbf{J}_{cc}$ and $\mathring{\mathbf{y}}$ are functions of $\mathbf{d}$, but independent from $\mathbf{n}$ (note that $\mathbf{n}$ does not affect the location of optimum). Therefore, $\mathbf{n}^{\mathrm{T}}$ and $\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}\mathring{\mathbf{y}}$ are independent and their inner product vanishes (expectations for $\mathbf{n}$ are 0). Similarly, for the third term, the trace of covariance matrix between $\mathbf{n}$ and $\mathring{\mathbf{y}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}$ also vanishes. Lastly, for the fourth term, since $\mathbf{n}\mathbf{n}^{\mathrm{T}}$ is independent from $\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}$,

$$E[\mathrm{tr}(\mathbf{n}\mathbf{n}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H})] = \mathrm{tr}[E(\mathbf{n}\mathbf{n}^{\mathrm{T}})E(\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H})] = \mathrm{tr}[\mathbf{W}^2E(\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H})] = E[\mathrm{tr}(\mathbf{W}^2\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H})] \quad (57)$$

Combining above results, we have

$$L_{\mathrm{gav}}(\mathbf{H}) = \frac{1}{2}E[\mathring{\mathbf{y}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H}\mathring{\mathbf{y}}] + \frac{1}{2}E[\mathrm{tr}(\mathbf{W}^2\mathbf{H}^{\mathrm{T}}\mathbf{J}_{cc}\mathbf{H})] \quad (58)$$

based on which Proposition 1 follows.

Numerical CVs for the evaporator

Table 8: Best measurement subset and numerical CVs for the evaporator using Algorithm 2

$n_{\mathring{y}}$	Best subset (minimal $\bar{L}_{\mathrm{gav}}$)	$\bar{L}_{\mathrm{gav}}$	$\mathbf{H}$
2	$[F_{100}\ F_{200}]$	19.150	$\begin{bmatrix} -14.291 & -0.0288 & 0.0673 \\ -28.160 & 3.669 & -0.0286 \end{bmatrix}$
3	$[F_2\ F_{100}\ F_{200}]$	4.0214	$\begin{bmatrix} -8.434 & -49.129 & 6.177 & 0.0732 \\ -16.547 & -97.402 & 15.971 & -0.017 \end{bmatrix}$
4	$[F_2\ F_{100}\ F_5\ F_{200}]$	3.5303	$\begin{bmatrix} -9.659 & -50.810 & 4.788 & 2.078 & 0.072 \\ -20.436 & -102.737 & 11.563 & 6.596 & -0.0211 \end{bmatrix}$
5	$[F_2\ F_{100}\ F_3\ F_5\ F_{200}]$	3.1190	$\begin{bmatrix} -6.219 & -52.050 & 3.8896 & 0.1459 & 2.3275 & 0.0757 \\ -11.951 & -105.797 & 9.348 & 0.360 & 7.211 & -0.0118 \end{bmatrix}$
6	$[F_2\ F_{100}\ F_3\ F_5\ F_{200}\ F_1]$	3.0742	$\begin{bmatrix} -6.314 & -52.650 & 3.744 & 0.147 & 2.055 & 0.0756 & 0.4727 \\ -12.400 & -108.62 & 8.665 & 0.365 & 5.931 & -0.0123 & 2.224 \end{bmatrix}$
7	$[P_2\ F_2\ F_{100}\ T_{201}\ F_3\ F_5\ F_{200}]$	3.0389	$\begin{bmatrix} 6.738 & 0.0919 & -52.513 & 3.646 & -0.199 & 0.164 & 1.833 & 0.0658 \\ 19.631 & 0.2254 & -106.901 & 8.751 & -0.485 & 0.402 & 6.004 & -0.0358 \end{bmatrix}$
8	$[P_2\ F_2\ F_{100}\ T_{201}\ F_3\ F_5\ F_{200}\ F_1]$	3.0153	$\begin{bmatrix} 6.124 & 0.0877 & -52.781 & 3.588 & -0.190 & 0.164 & 1.724 & 0.0662 & 0.256 \\ 15.063 & 0.195 & -108.895 & 8.317 & -0.420 & 0.402 & 5.200 & -0.0330 & 1.678 \end{bmatrix}$
9	$[P_2\ T_2\ F_2\ F_{100}\ T_{201}\ F_3\ F_5\ F_{200}\ F_1]$	3.0106	$\begin{bmatrix} 3.637 & 0.0637 & 0.0591 & -52.762 & 3.569 & -0.206 & 0.163 & 1.696 & 0.0654 & 0.2054 \\ 9.548 & 0.1412 & 0.1310 & -108.854 & 8.276 & -0.456 & 0.398 & 5.138 & -0.0346 & 1.633 \end{bmatrix}$
10	$[P_2\ T_2\ T_3\ F_2\ F_{100}\ T_{201}\ F_3\ F_5\ F_{200}\ F_1]$	3.0083	$\begin{bmatrix} 2.378 & 0.0521 & 0.0483 & 0.0436 & -52.753 & 3.560 & -0.214 & 0.162 & 1.682 & 0.0651 & 0.1955 \\ 6.758 & 0.1155 & 0.1071 & 0.0967 & -108.834 & 8.256 & -0.473 & 0.397 & 5.108 & -0.0354 & 1.611 \end{bmatrix}$