RESEARCH ARTICLE

Lagrange-Based Global Self-Optimizing Control for Constraint Activeness Varying Processes

HONGXIN SU[1,2], YI CAO[1,2], (Senior Member, IEEE),
SHUANG-HUA YANG[1,2,3], (Senior Member, IEEE), AND LINGJIAN YE[4], (Member, IEEE)

[1]College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, Zhejiang 310058, China
[2]Institute of Zhejiang University-Quzhou, Quzhou, Zhejiang 324000, China
[3]Department of Computer Science, University of Reading, RG6 6DH Reading, U.K.
[4]School of Engineering, Huzhou University, Huzhou, Zhejiang 313000, China

Corresponding author: Shuang-Hua Yang (yangsh@zju.edu.cn)

ABSTRACT Self-optimizing control (SOC) aiming to select most appropriate controlled variables (CVs), is a promising control strategy in the field of real-time optimization. An approach, global self-optimizing control (gSOC) was proposed to find globally optimal CVs by minimizing the global average of the economic loss over the whole operation space. However, as the gSOC was developed from the local SOC, it inherited the same theoretical basis by assuming invariant constraint activeness. Nevertheless, this will significantly restrict the applicable range of SOC as in many real systems activeness varying constraints are common. The difficulty for the gSOC to consider activeness varying constraints is the degrees of freedom inconsistency in CV selection. To tackle the problem, this paper rebuilds the gSOC approach based on the Lagrange function and the well known Karush-Kuhn-Tucker conditions to incorporate active and inactive constraints uniformly in a Lagrange-based global average loss expression. As the gSOC approach is based on optimal measurement data, optimal values of the newly introduced Lagrange multiplies can also be obtained from the same optimization results as well. With this novel Lagrange-based loss function, an optimization problem for CV selection is formulated although non-convex. Thus, the short-cut algorithm of the original gSOC approach is amended for the Lagrange-based gSOC (LgSOC) problem to derive a closed-form solution. Furthermore, the existing cascade SOC structure for a single constraint is generalized to guarantee all constraints satisfied in the whole space. The proposed LgSOC method was proved effective to solve constraint activeness varying gSOC problems through an evaporator case study.

INDEX TERMS Activeness varying constraints, global self-optimizing control, Lagrange-based method.

I. INTRODUCTION

Self-optimizing control (SOC) [1], [2] is a control strategy for real-time optimization (RTO) of chemical processes. SOC focuses on the selection of appropriate controlled variables (CVs) under disturbances so that the near-optimal operation (with acceptable economic loss) can be achieved by just keeping CVs at constant setpoints through simple feedback control.

Traditional RTO approaches usually work in a two layer structure [3], which needs a RTO layer (which is open-loop) to update the optimal steady-state setpoints for the

The associate editor coordinating the review of this manuscript and approving it for publication was Taous Meriem Laleg-Kirati.

regulatory control loops in the presence of disturbances and uncertainties. Whereas, SOC works in a feedback control structure, which runs faster than an upper optimizer and no longer needs waiting time for reaching another steady state in traditional RTO. Besides, the RTO layer can be omitted when the selected self-optimizing CVs are able to achieve an acceptable economic loss as shown in Fig. 1, hence there is no need for SOC to update setpoints of CVs frequently to maintain optimal operation, just keep CVs at constant setpoints. Therefore, the core of SOC is to search for optimal CVs.

A. MOTIVATIONS

Most SOC methods have assumed that disturbances do not alter active constraints. However, in general, disturbances
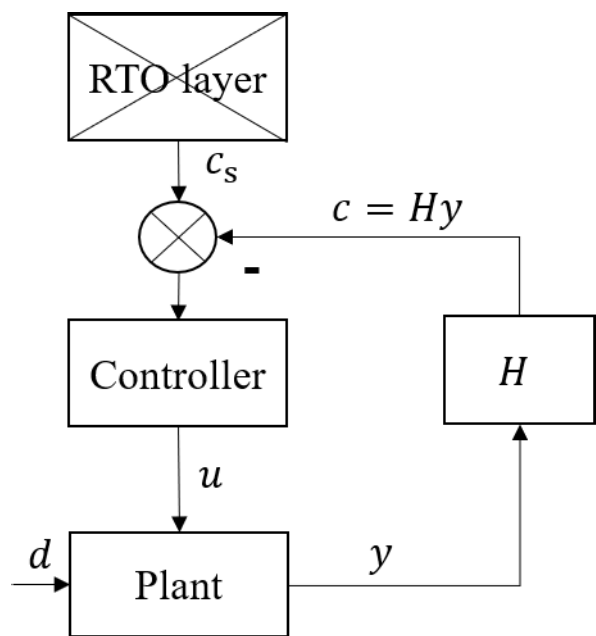
FIGURE 1. Structure of self-optimizing control (c and c_s: self-optimizing controlled variables and setpoints; H: linear combination matrix; y: measurements; u: manipulated variables; d: disturbances).

can cause changes to the active set and may require different operation of the plant in different regions defined by active constraints, which means that different CVs should be selected in different regions. Nevertheless, it results in relatively complicated implementation structure and we are motivated to propose a simple yet efficacious method to address the issue.

B. RELATED WORKS ABOUT LOCAL AND GLOBAL SOC METHODS

Under the assumption that the constraint activeness does not change, a great number of approaches have been proposed to select optimal CVs by minimizing the economic loss in the case of disturbances and uncertainties. The exact local method [4] was proposed based on linearization around the nominal optimal point and the optimal linear combinations of measurements were selected as CVs through minimization of worst case local economic loss, which was evaluated by the second-order Taylor expansion of the cost function at the nominal point. To derive an analytical solution of linear combination matrix H, the null space method [5] and its extended version [6] were proposed without and with considering measurement noise. The null space method chooses optimal H as the left null space of optimal sensitivity matrix with respect to disturbances, hence just minimizing the loss caused by disturbances. The extended method uses extra degrees of freedom to deal with the effect of measurement noise after minimizing the effect of disturbances. Since the above methods are all based on a linearized model, their optimality is locally valid and the overall performance, in most cases, is not satisfactory.

To address the local deficiency, the global optimal CVs were proposed to approximate the necessary conditions of optimality (NCO) through regression over the entire operation space with data produced by Monte Carlo simulation [7], [8], [9], [10], [11]. It overcomes the limitation of local methods and a near-optimal operation can be achieved globally when the CVs are kept at zero. The achievable loss is greatly dependent on the regression error, which is associated with the sampling points. To reduce the regression error, the sampling points should be close to the optimum, where the final system is expected to work around. However, the gradient is zero at the optimum, which makes the regression problem singular. To overcome this, another global SOC (gSOC) method [12] was proposed using the optimal data for CV selection to minimize the global average economic loss, which were derived through the second-order Taylor expansion across the whole disturbance and uncertainty space, rather than a single nominally optimal point in local methods. The CV selection problem was formulated as a nonlinear programming (NLP) problem and an efficient short-cut algorithm [12], [13] was proposed to solve the problem by introducing an extra constraint to force a unit Hessian matrix at a reference point.

C. RELATED WORKS ABOUT CONSTRAINT HANDLING IN SOC

However, all the above approaches are under the invariant constraint activeness assumption, *i.e.* the set of active constraints does not change. This assumption is acceptable for local SOC approaches as only a small neighborhood of a reference point is considered in these methods. However, in practice, varying activeness constraints commonly exist, particularly when disturbance ranges are large. The gSOC approach inherited the theoretic basis from the local SOC to restrict itself for invariant constraint activeness cases. It significantly restricts the operation range for the gSOC to be applicable.

The constraint activeness varying issue has been studied since the local SOC theory was developed. To solve this issue within the local SOC theory, the main difficulty is the varying degrees of freedom for CVs to be selected. To address this issue, in one of these approaches, the entire operation space was divided into several regions such that within each region, the active constraints are invariant, hence the null space method is applicable to derive a set of specific CVs for a specific region [14]. With this approach, a switching strategy is also needed, which results in a complicated control structure. As an alternative, since the cascade control structure is widely used in many chemical processes such as integrating [15] and unstable processes [16], [17], it can be also applied to cope with conditionally active constraints [18]. The constrained variables are controlled in the inner loop, whose setpoints are determined by the self-optimizing CVs (selected as the gradient of the cost function with respect to the manipulated variable) in the outer loop. The cascade control structure ensures

that the constrained variables are controlled at their setpoints when the constraints are active while the self-optimizing CVs are maintained at zero setpoints to keep optimality when the constraints are inactive. Nevertheless, due to the limitation of the structure, the number of CVs should be no less than the number of constrained variables which may vary between active and inactive. To simplify the implementation policy, the explicit constraint handling approach [19] was proposed to find optimal CVs that make sure the constrained variables are always within their feasible ranges. Based on the same idea, the constraint estimation method [20] which incorporates the constrained variables with the gSOC formulation was proposed for the globally optimal CVs. However, these two approaches are both conservative because of off tracking optimal values of constrained variables, which often leads to a relatively large loss. Recently, an intelligent approach [21], [22] was presented to find globally optimal CVs through machine learning to make sure the optimal values of constrained variables are tracked as well as the economic loss is acceptable. Whereas, it is restricted to deal with the case that the number of remaining degrees of freedom and constrained variables are the same.

D. PROPOSED IDEA IN THIS WORK

Due to the limitation of local SOC theory, the above methods are not able to solve the varying constraint activeness problem satisfactorily either over complicated or over conservative. Such a local restriction is inherited even by the gSOC. To overcome this problem, the theoretical basis has to be rebuilt to get rid of the invariant constraint activeness assumption. For this purpose, by adopting the Lagrange function and the well known Karush-Kuhn-Tucker (KKT) conditions for constrained optimization problems, this paper derives a new Lagrange-based quadratic economic loss function applicable to both unconstrained and constrained gSOC problems. It ensures that the economic loss is always greater than zero regardless of whether the constraints are active or not. When the activeness of all constraints is invariant, the new Lagrange-based loss is equivalent to the conventional loss, hence it is a generalization of the original global average loss. More importantly, the Lagrange-based loss provides a unified criterion of optimality of a set of CVs so that the optimal CVs can be obtained by minimizing the Lagrange-based loss function over the entire operational space. However, this optimization problem is complicated and non-convex. To facilitate solving it, the short-cut algorithm of the original gSOC [12] is adopted to work with the new Lagrange-based loss function. An extra constraint is enforced that the Hessian of the Lagrange function with respect to the CVs is constant at a selected reference point. Then the explicit expression for optimal CVs can be obtained. The proposed Lagrange-based gSOC (LgSOC) method provides a systematic way to find globally optimal CVs applicable to both constraint activeness varying or invariant cases. To further reduce the extra loss caused by the short-cut algorithm, the LgSOC method

is retrofitted by enforcing multiple constraints at multiple points. Furthermore, the existing cascade SOC structure for a single constraint proposed in [18] is generalized in this work to ensure all constraints are satisfied in the whole operational space. A numerical example and a benchmark evaporator process are studied to show the effectiveness of these methods.

E. CONTRIBUTIONS

The contributions to knowledge made in this paper are:

1) a general Lagrange-based quadratic loss function is proposed to provide a unified criterion of optimality of a single set of CVs for both varying and invariant constraint activeness cases;
2) the LgSOC method is introduced to facilitate solving the non-convex NLP problem which is formulated to find optimal CVs by minimizing the Lagrange-based quadratic loss;
3) the one degree of freedom cascade SOC structure is generalized to guarantee that all constraints are satisfied.

The paper is organized as follows. Section II derives the Lagrange-based loss function in detail and in Section III, the LgSOC method and its retrofitted version are proposed to find optimal CVs efficiently. Section IV introduces the generalized cascade constrained SOC structure. A numerical example and an evaporator case study of the proposed methods are given in Section V, and finally, the work is concluded in Section VI.

II. DERIVATION OF LAGRANGE-BASED LOSS FUNCTION

A general constrained nonlinear optimization problem is considered as follows

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

with available measurements

$$\mathbf{y} = \mathbf{f}(\mathbf{u}, \mathbf{d}) \tag{2}$$
$$\mathbf{y}_{\mathrm{m}} = \mathbf{y} + \mathbf{n} \tag{3}$$

where J is a scalar cost function to be minimized, $\mathbf{g}$ are the constraints related to operational safety and product quality requirements, and $\mathbf{f}$ is the input-output model function, and $\mathbf{u}$, $\mathbf{d}$, $\mathbf{y}$, $\mathbf{y}_{\mathrm{m}}$ and $\mathbf{n}$ are manipulated variables, disturbances, theoretical measurements, actual measurements and measurement noise, respectively. The objective cost function and constraints are both indirectly related to the measurements by using the manipulated variables to control the corresponding controlled variables under different disturbances.

For the inequality constrained optimization problem (1), the necessary conditions of the problem are the Karush-Kuhn-Tucker (KKT) conditions [23] presented as

follows

$$\mathcal{L}_u(\mathbf{u}, \mathbf{d}, \lambda) = \mathbf{J}_u(\mathbf{u}, \mathbf{d}) + \sum_{j=1}^{n_g} \lambda^j \mathbf{g}_u^j(\mathbf{u}, \mathbf{d}) \qquad = \mathbf{0} \qquad (4)$$

$$\lambda^j g^j(\mathbf{u}, \mathbf{d}) = 0 \qquad (5)$$

$$\lambda^j \geq 0 \qquad (6)$$

$$g^j(\mathbf{u}, \mathbf{d}) \leq 0 \qquad (7)$$

where $\mathcal{L}(\mathbf{u}, \mathbf{d}, \lambda) = J(\mathbf{u}, \mathbf{d}) + \sum_{j=1}^{n_g} \lambda^j g^j(\mathbf{u}, \mathbf{d})$ is the Lagrange function, n_g is the number of constraints and λ^j is the jth Lagrange multiplier corresponding to the jth constraint g^j. $\mathcal{L}_u$, $\mathbf{J}_u$ and $\mathbf{g}_u^j$ represent the gradient of $\mathcal{L}$, J and jth element of $\mathbf{g}$ respectively, all with respect to $\mathbf{u}$. $\mathbf{u}_{\text{opt}}(\mathbf{d})$ and $\lambda_{\text{opt}}^j(\mathbf{d})$, $j = 1, \ldots, n_g$ are the optimal solution to optimization problem (1) satisfying the KKT conditions (4)-(7).

SOC method tries to find an optimal combination of measurements $\mathbf{c} = \mathbf{Hy}$, where $\mathbf{H}$ is a linear combination matrix, as CVs such that when $\mathbf{c}$ is kept at constant setpoint $\mathbf{c}_s$ by adjusting $\mathbf{u}$, the optimal steady-state operation (1) can be achieved. Note that measurements $\mathbf{y}$ can be expanded with a constant 1, e.g. $\hat{\mathbf{y}} = \begin{bmatrix} 1 & \mathbf{y}^T \end{bmatrix}^T$, and at the same time, $\mathbf{H}$ can be augmented to be $\hat{\mathbf{H}} = \begin{bmatrix} \mathbf{h}_0 & \mathbf{H} \end{bmatrix}$, where $\mathbf{h}_0 = -\mathbf{c}_s$. In this case, $\hat{\mathbf{c}}$ should be controlled at zero setpoints. For simplicity, $\mathbf{y}$ and $\mathbf{H}$ will represent $\hat{\mathbf{y}}$ and $\hat{\mathbf{H}}$ respectively in the remaining part of the paper.

Remark 1: Since the economic operation of the plant often occurs at steady-state for most continuous processes, the objective in most continuous processes is simply to find the economically optimal steady-state operating point. Therefore, this paper mainly focuses on achieving asymptotic optimal operation by finding optimal CVs, and improves dynamic performances by tuning parameters of the feedback controllers.

The Lagrange-based economic loss is selected as the criterion of CV selection and it is defined as the difference between the actual value of the Lagrange function and its optimum shown as follows

$$L = \mathcal{L}(\mathbf{u}, \mathbf{d}, \lambda) - \mathcal{L}_{\text{opt}}(\mathbf{d}) \qquad (8)$$

However, it can be difficult to calculate the Lagrange-based economic loss straightforward from (8) since the computation is cumbersome when the controlled system is operated in closed-loop. To simplify this problem, the loss can be approximated through the second-order Taylor expansion of $\mathcal{L}$ with respect to $\mathbf{u}$ at the optimal point

$$\mathcal{L}(\mathbf{u}, \mathbf{d}, \lambda) \approx \mathcal{L}_{\text{opt}}(\mathbf{d}) + \mathcal{L}_u \mathbf{e}_u + \frac{1}{2} \mathbf{e}_u^T \mathcal{L}_{uu} \mathbf{e}_u \qquad (9)$$

where $\mathbf{e}_u \triangleq \mathbf{u}^{\text{fb}} - \mathbf{u}_{\text{opt}}$ is the input deviation from the optimum in the closed loop and $(\cdot)^{\text{fb}}$ represents the terms of actual values after feedback control. $\mathcal{L}_u$ and $\mathcal{L}_{uu}$ are the gradient and Hessian matrix of $\mathcal{L}$ with respect to $\mathbf{u}$ evaluated at the optimum respectively. $\mathcal{L}_u$ is obviously equal to zero according to the KKT condition (4) and $\mathcal{L}_{uu}$ is symmetric

positive definite since it is evaluated at the optimum and it can be calculated from

$$\mathcal{L}_{uu} = \mathbf{J}_{uu} + \sum_{j=1}^{n_g} \lambda_{\text{opt}}^j \mathbf{g}_{uu}^j \qquad (10)$$

where $\mathbf{J}_{uu}$ and $\mathbf{g}_{uu}^j$ are the Hessian matrix of J and jth element of $\mathbf{g}$ with respect to $\mathbf{u}$ at the optimum respectively.

Therefore, the Lagrange-based economic loss function can be expressed in a quadratic form as indicated below

$$L = \mathcal{L}(\mathbf{u}, \mathbf{d}, \lambda) - \mathcal{L}_{\text{opt}}(\mathbf{d}) \approx \frac{1}{2} \mathbf{e}_u^T \mathcal{L}_{uu} \mathbf{e}_u \qquad (11)$$

Since the global SOC method does not rely on a linearized model like local SOC methods but the original nonlinear model (2), the values of $\mathbf{u}^{\text{fb}}$ and $\mathbf{y}^{\text{fb}}$ are determined by the following two equations

$$\mathbf{y}^{\text{fb}} = \mathbf{f}(\mathbf{u}^{\text{fb}}, \mathbf{d}) \qquad (12)$$

$$\mathbf{c}_m^{\text{fb}} = \mathbf{Hy}_m^{\text{fb}} = \mathbf{H}(\mathbf{y}^{\text{fb}} + \mathbf{n}) = \mathbf{0} \qquad (13)$$

where $\mathbf{c}_m^{\text{fb}}$ means the measured values of CVs under closed-loop control, which should be controlled exactly at zero at steady state without implementation error when an integral is included in the feedback controller. However, minimizing L in (11) requires solving a set of nonlinear model equations for a given $\mathbf{H}$ because L is implicitly related with $\mathbf{H}$ through $\mathbf{u}^{\text{fb}}$, resulting in cumbersome and time-consuming calculation.

To address this issue, L can be similarly expressed in the quadratic form in terms of $\mathbf{c}$ as

$$L \approx \frac{1}{2} \mathbf{e}_c^T \mathcal{L}_{cc} \mathbf{e}_c \qquad (14)$$

where $\mathbf{e}_c \triangleq \mathbf{c}^{\text{fb}} - \mathbf{c}_{\text{opt}}$ is the CV deviation from the optimum. $\mathcal{L}_{cc}$ is the Hessian matrix of $\mathcal{L}$ with respect to $\mathbf{c}$ evaluated at the optimal point, which is also symmetric positive definite and has the following relationship with $\mathcal{L}_{uu}$ [12].

$$\mathcal{L}_{cc} = (\mathbf{Hf}_u)^{-T} \mathcal{L}_{uu} (\mathbf{Hf}_u)^{-1} \qquad (15)$$

where $\mathbf{f}_u$ is the sensitivity matrix of $\mathbf{y}$ with respect to $\mathbf{u}$.

Since $\mathbf{c}^{\text{fb}} = \mathbf{Hy}^{\text{fb}} = \mathbf{H}(\mathbf{y}_m^{\text{fb}} - \mathbf{n})$ and $\mathbf{c}_m^{\text{fb}} = \mathbf{Hy}_m^{\text{fb}} = \mathbf{0}$ through feedback control, $\mathbf{c}^{\text{fb}} = -\mathbf{Hn}$. Because $\mathbf{c}_{\text{opt}} = \mathbf{Hy}_{\text{opt}}$, $\mathbf{e}_c = \mathbf{c}^{\text{fb}} - \mathbf{c}_{\text{opt}} = -\mathbf{Hn} - \mathbf{Hy}_{\text{opt}} = -\mathbf{H}(\mathbf{y}_{\text{opt}} + \mathbf{n})$. Thus, the Lagrange-based economic loss can be explicitly with $\mathbf{H}$ and expressed as follows

$$L = \frac{1}{2}(\mathbf{y}_{\text{opt}} + \mathbf{n})^T \mathbf{H}^T \mathcal{L}_{cc} \mathbf{H}(\mathbf{y}_{\text{opt}} + \mathbf{n}) \qquad (16)$$

Remark 2: Since $\mathbf{c}_s$ has been incorporated in the formulation of the combination matrix, the setpoint for $\mathbf{c}_m^{\text{fb}}$ is zero. When an integral is included in the feedback controller, $\mathbf{c}_m^{\text{fb}}$ should be exactly equal to zero at steady state. Whereas, $\mathbf{c}^{\text{fb}}$ does not necessarily equal zero due to measurement noise. In addition, $\mathbf{c}_{\text{opt}}$ is the optimal value of $\mathbf{c}$ for a given $\mathbf{d}$, which is also not necessarily equal to zero since the objective is to minimize L in (16).

Note that in the unconstrained gSOC method, the original economic loss is approximated through second-order Taylor

expansion of J with respect to $\mathbf{c}$ at the optimum shown as follows [4]

$$J(\mathbf{c}, \mathbf{d}) = J_{\text{opt}}(\mathbf{d}) + \mathbf{J}_c \mathbf{e}_c + \frac{1}{2} \mathbf{e}_c^{\mathrm{T}} \mathbf{J}_{cc} \mathbf{e}_c \quad (17)$$

where $\mathbf{J}_c$ and $\mathbf{J}_{cc}$ are the first derivative and the Hessian matrix of J with respect to $\mathbf{c}$ at the optimum, and $\mathbf{J}_{cc}$ satisfies the following relationship

$$\mathbf{J}_{cc} = (\mathbf{Hf}_u)^{-\mathrm{T}} \mathbf{J}_{uu} (\mathbf{Hf}_u)^{-1} \quad (18)$$

Since the active constraint set is invariant in the unconstrained case, $\mathbf{J}_c = \mathbf{0}$. The original global economic loss is then $L_0 = \frac{1}{2} \mathbf{e}_c^{\mathrm{T}} \mathbf{J}_{cc} \mathbf{e}_c > 0$ since $\mathbf{J}_{cc}$ is positive definite at the optimal point. However, under the circumstance that constraint activeness varies, the loss can not be guaranteed to be greater than 0 because $\mathbf{J}_c$ is not necessarily equal to zero and thus the loss can not be expressed in the quadratic form. In this case, if $L_0 = J(\mathbf{c}, \mathbf{d}) - J_{\text{opt}}(\mathbf{d}) = \mathbf{J}_c \mathbf{e}_c + \frac{1}{2} \mathbf{e}_c^{\mathrm{T}} \mathbf{J}_{cc} \mathbf{e}_c < 0$, it indicates that the operational constraints in the system are violated. Therefore, the CV selection criterion in the unconstrained gSOC method is no longer applicable to constrained gSOC problems. As a contrast, $\mathcal{L}_c = \mathcal{L}_u (\mathbf{Hf}_u)^{-1} = \mathbf{0}$ always holds no matter whether the active set changes or not according to (4), hence the Lagrange-based economic loss can always be described as a quadratic form in (16) and is always greater than zero regardless of whether the active set changes or not since $\mathcal{L}_{cc}$ is positive definite at the optimal point.

To sum up, the Lagrange-based global economic loss can be regarded as an extended form of the original global economic loss, and the Lagrange-based global economic loss degenerates into the original one when active set does not change. In addition, the former is suitable for both constraint activeness varying or invariant cases while the latter only applies to the invariant active set case.

Following the idea in gSOC method [12], the Lagrange-based global average economic loss can also be expressed as two parts

$$L_{\text{gav}} = \mathrm{E}(\frac{1}{2}(\mathbf{y}_{\text{opt}} + \mathbf{n})^{\mathrm{T}} \mathbf{H}^{\mathrm{T}} \mathcal{L}_{cc} \mathbf{H}(\mathbf{y}_{\text{opt}} + \mathbf{n}))$$
$$= \mathrm{E}(L^d) + \mathrm{E}(L^n) \quad (19)$$

where

$$L^d = \frac{1}{2} \mathbf{y}_{\text{opt}}^{\mathrm{T}} \mathbf{H}^{\mathrm{T}} \mathcal{L}_{cc} \mathbf{H} \mathbf{y}_{\text{opt}}, \quad L^n = \frac{1}{2} \mathrm{tr}(\mathbf{W}^2 \mathbf{H}^{\mathrm{T}} \mathcal{L}_{cc} \mathbf{H}) \quad (20)$$

$\mathrm{tr}(\cdot)$ represents the trace of a matrix, $\mathrm{E}(\cdot)$ stands for the expectation and $\mathbf{W}^2 = \mathrm{E}(\mathbf{n}\mathbf{n}^{\mathrm{T}})$ is a diagonal matrix if $\mathbf{n}$ are mutually independent. L^d and L^n denote the contribution of the disturbance and measurement noise to the economic loss respectively.

III. LAGRANGE-BASED gSOC METHOD

For a nonlinear process, the expectations of L^d and L^n in (20) are generally hard to be analytically calculated. To overcome the difficulty, the Lagrange-based global average economic

loss can be approximated over a set of sampling points through Monte Carlo simulation as below

$$L_{\text{gav}} \approx \overline{L}_{\text{gav}} \triangleq \frac{1}{N} \sum_{i=1}^{N} \left[L_{(i)}^d + L_{(i)}^n \right] \quad (21)$$

where N is the number of sampled optimal points over the whole operation region, and the subscript $(\cdot)_{(i)}$ represents the terms corresponding to the ith disturbance scenario $\mathbf{d}_{(i)}$, obtained by sampling which follows the random distribution specified to the disturbance.

SOC method intends to find the optimal $\mathbf{H}$ by minimizing the Lagrange-based global average economic loss. Note that the solution $\mathbf{H}$ to the optimization problem of minimizing $\overline{L}_{\text{gav}}$ in (21) is not unique, given in the Appendix.

However, minimizing $\overline{L}_{\text{gav}}$ in (21) is a non-convex optimization problem, hence could not be solved efficiently. In order to develop an efficient algorithm to approximate the optimal solution, a short-cut approach [12] is adopted here. By assuming $\mathcal{L}_{cc}$ constant over the entire disturbance region, then, $\overline{L}_{\text{gav}}$ becomes a quadratic function, hence easy to be optimized. It is worth noting that the assumption of $\mathcal{L}_{cc} = \mathbf{I}$, equivalent to $\mathbf{Hf}_u = \mathcal{L}_{uu}^{1/2}$, is directly imposed on the loss function $\overline{L}_{\text{gav}}$ but not explicitly involved in the optimization problem. Therefore, to guarantee the uniqueness of the optimal CVs as well as enforcing $\mathcal{L}_{cc,(r)} = \mathbf{I}$ at the reference point, the extra constraint of $\mathbf{Hf}_{u,(r)} = \mathcal{L}_{uu,(r)}^{1/2}$ imposed on a selected reference point is required, where the subscript $(\cdot)_{(r)}$ represents the terms corresponding to the selected reference point.

Consequently, (20) can be simplified as

$$L^d = \frac{1}{2} \mathbf{y}_{\text{opt}}^{\mathrm{T}} \mathbf{H}^{\mathrm{T}} \mathbf{H} \mathbf{y}_{\text{opt}}, \quad L^n = \frac{1}{2} \mathrm{tr}(\mathbf{W}^2 \mathbf{H}^{\mathrm{T}} \mathbf{H}) \quad (22)$$

and the short-cut global average economic loss can be written as

$$\overline{\overline{L}}_{\text{gav}} = \mathrm{E}(\frac{1}{2} \mathbf{y}_{\text{opt}}^{\mathrm{T}} \mathbf{H}^{\mathrm{T}} \mathbf{H} \mathbf{y}_{\text{opt}}) + \mathrm{E}(\frac{1}{2} \mathrm{tr}(\mathbf{W}^2 \mathbf{H}^{\mathrm{T}} \mathbf{H}))$$
$$= \frac{1}{2N} \left\| \mathbf{Y}^{\mathrm{T}} \mathbf{H}^{\mathrm{T}} \right\|_{\mathrm{F}}^2 + \frac{1}{2} \left\| \mathbf{W} \mathbf{H}^{\mathrm{T}} \right\|_{\mathrm{F}}^2 \quad (23)$$
$$= \frac{1}{2} \left\| \tilde{\mathbf{Y}} \mathbf{H}^{\mathrm{T}} \right\|_{\mathrm{F}}^2$$

where

$$\mathbf{Y} = \left[\mathbf{y}_{\text{opt},(1)} \cdots \mathbf{y}_{\text{opt},(N)} \right], \quad \tilde{\mathbf{Y}} = \begin{bmatrix} \frac{1}{\sqrt{N}} \mathbf{Y}^{\mathrm{T}} \\ \mathbf{W} \end{bmatrix} \quad (24)$$

$\mathbf{y}_{\text{opt},(i)}$ is the optimal measurement vector at ith sampling point.

Thus the original optimization problem can be reformulated as

$$\min_{\mathbf{H}} \overline{\overline{L}}_{\text{gav}} = \frac{1}{2} \left\| \tilde{\mathbf{Y}} \mathbf{H}^{\mathrm{T}} \right\|_{\mathrm{F}}^2 \quad (25)$$
$$\text{s.t. } \mathbf{Hf}_{u,(r)} = \mathcal{L}_{uu,(r)}^{1/2}$$

Then **H** can be explicitly derived as below [6]

$$\mathbf{H}^T = \left(\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}\right)^{-1}\mathbf{f}_{u,(r)}\left(\mathbf{f}_{u,(r)}^T\left(\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}\right)^{-1}\mathbf{f}_{u,(r)}\right)^{-1}\mathcal{L}_{\text{uu},(r)}^{1/2} \tag{26}$$

To sum up, the detailed procedure for implementing the proposed Lagrange-based gSOC (LgSOC) method to find the globally optimal CVs is shown as follows

---

Algorithm 1 LgSOC Method

1: Sample the whole operation space using Monte Carlo simulation to generate N different disturbance scenarios
2: Obtain optimal data such as $\mathbf{y}_{\text{opt},(i)}$, $\mathbf{f}_{u,(i)}$, $\mathbf{J}_{uu,(i)}$, $\mathbf{g}_{uu,(i)}^j$ and $\lambda_{\text{opt},(i)}^j$ for each $\mathbf{d}_{(i)}$ $(i = 1 \cdots N, j = 1 \cdots n_g)$ through minimizing J (offline optimization), and then calculate $\mathcal{L}_{uu,(i)}$ from (10).
3: Choose **W**. Construct **Y** and $\tilde{\mathbf{Y}}$ in (24).
4: Compute the solution **H** from (26).

Remark 3: Note that it is different from the gradient-based methods [7], [8] that use non-optimal data for regression, in which case the values of Lagrange multipliers λ^j $(j = 1 \cdots n_g)$ are undetermined. The proposed Lagrange-based method is based on optimal data inherited from the original gSOC method [12]. The optimal data such as $\mathbf{y}_{\text{opt}}$ and λ_{opt}^j can be obtained through offline solving the original optimization problem (1). Then the optimal **H** in (26) can be calculated offline based on these optimal data. For an online application, since the exact value of **H** is already known, the online values of CVs can be simply calculated using $\mathbf{c} = \mathbf{H}\mathbf{y}$ based on online measured values of **y** and offline calculated **H**. Hence, instead of detecting the values of λ^j online, only offline obtained optimal values λ_{opt}^j are needed when calculating optimal **H** offline.

The short-cut algorithm is derived based on the assumption $\mathcal{L}_{cc} = \mathbf{I}$ over the entire region, while in practice only at a reference point this constraint can be imposed. This mismatch determines the result obtained through the short-cut approach is not optimal. To alleviate the deficiency, it is desired that the extra constraints can be imposed on multiple selected points. Based on the method of Lagrange multipliers, the expression of **H** explicitly related to the multipliers can be similarly obtained as

$$\tilde{\mathbf{H}}^T = \left(\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}\right)^{-1}\tilde{\mathbf{f}}_u\left(\tilde{\mathbf{f}}_u^T\left(\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}\right)^{-1}\tilde{\mathbf{f}}_u\right)^{-1}\tilde{\mathcal{L}}_{uu}^{1/2} \tag{27}$$

where $\tilde{\mathbf{f}}_u = \sum_{k=1}^m \mu_k \mathbf{f}_{u,(k)}$, $\tilde{\mathcal{L}}_{uu}^{1/2} = \sum_{k=1}^m \mu_k \mathcal{L}_{uu,(k)}^{1/2}$, μ_k is the kth Lagrange multiplier and m is the number of selected reference points. Then (27) can be substituted into (21) and get

$$\min_{\mu_1 \cdots \mu_m} \quad \frac{1}{N}\sum_{i=1}^N \left[\frac{1}{2}(\mathbf{y}_{\text{opt},(i)} + \mathbf{n}_i)^T\tilde{\mathbf{H}}^T\mathcal{L}_{cc,(i)}\tilde{\mathbf{H}}(\mathbf{y}_{\text{opt},(i)} + \mathbf{n}_i)\right] \tag{28}$$

Then the optimal $\tilde{\mathbf{H}}$ can be finally found once optimal values of Lagrange multipliers μ_k $(k = 1 \cdots m)$ are obtained by solving the nonlinear optimization problem (28) through numerical optimization methods such as the sequential quadratic programming (SQP), interior-point, etc. Since the reference point in the LgSOC method can be selected as the initial point in the retrofitted LgSOC method, this retrofitted version would achieve better performance than LgSOC method.

Remark 4: Note that λ and μ are both Lagrange multipliers. However, λ is used in the KKT conditions to construct the Lagrange-based economic loss function while μ is used to combine extra constraints at multiple reference points in the retrofitted LgSOC method.

In summary, the procedure for implementing the retrofitted LgSOC method is presented as follows.

Algorithm 2 retrofitted LgSOC Method

1: Sample the whole operation space using Monte Carlo simulation to generate N different disturbance scenarios
2: Obtain optimal data such as $\mathbf{y}_{\text{opt},(i)}$, $\mathbf{f}_{u,(i)}$, $\mathbf{J}_{uu,(i)}$, $\mathbf{g}_{uu,(i)}^j$ and $\lambda_{\text{opt},(i)}^j$ for each $\mathbf{d}_{(i)}$ $(i = 1 \cdots N, j = 1 \cdots n_g)$ through minimizing J (offline optimization), and then calculate $\mathcal{L}_{uu,(i)}$ from (10).
3: Choose **W**. Construct **Y** and $\tilde{\mathbf{Y}}$ in (24).
4: Solve the optimization problem (28) to obtain optimal values of Lagrange multipliers μ_k.
5: Compute the optimal **H** from (27).

IV. GENERALIZED CASCADE SOC STRUCTURE

Although all constraints are satisfied in solutions to the original optimization problem (1) due to the KKT conditions, it is the case for optimal but not perfect CVs. As long as the loss is not zero, the corresponding operating condition is not optimal but only near optimal, hence, it is not guaranteed that all constraints are satisfied online, which is illustrated by the numerical example in Section V. Therefore, extra measures should be taken to ensure feasibility of operations.

To tackle the constraint activeness varying issue, a cascade control structure, as shown in Fig. 2 was proposed in 2005 [18] to effectively achieve automatic switching between an active constraint and a self-optimizing CV under disturbances. However, only the simplest case of one degree of freedom was considered in the original work. Since then, the idea was laid aside as the community considering it is difficult to apply it to multivariate cases [2]. In this section, this idea is revisited to propose a systematic procedure for multivariate cases.

A. DESCRIPTION OF GENERALIZED STRUCTURE

A generalized cascade control structure is illustrated in Fig. 3. There may be more similar loops since multiple degrees of freedom cases are considered and only one of them is shown here without loss of generality. It is clear that the
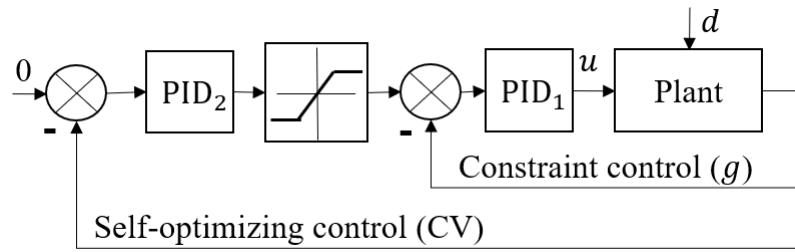
FIGURE 2. Cascade control structure for one degree of freedom cases [18].

manipulated variable (MV) u_1 should work in the innermost loop, the self-optimizing CV c_1 obtained based on LgSOC method proposed in Section III in the outermost loop and conditionally active constraints $g_1, g_2, \cdots, g_{n_g}$ in between. Saturation blocks are added to restrict their limits.

Note that there is at most one of them that can be active at the same time. Specifically speaking, if the MV becomes active, it means that it is set exactly at its limit value and consumes one degree of freedom, leading to an open-loop control structure. If one of the conditionally active constraints becomes active, it should be controlled at its limit value by adjusting MV. When the self-optimizing CV comes to be active, the system is unconstrained and the CV is maintained at constant zero setpoint by adjusting MV similarly.

B. VARIABLE PAIRING

Since there must be only one active variable in the same cascade loop, constraints which can become active simultaneously should not be put together in the same loop. Examining data obtained from offline optimization cover the entire operation range, simultaneously active constraints can be identified and they should be placed separately in different cascade loops. However, if the case that they are in the same loop cannot be avoided, infrequently active constraints or constraints that contribute less to the economic loss can be put in the same loop and they can be chosen following either of the rules:

1) Select constraints from the optimal data which become active less frequently.
2) Select constraints whose corresponding $\lambda_{\text{opt}}^j (j = 1 \cdots n_g)$ are small.

After separating simultaneously active constraints in different cascade loops, here comes the pairing problem, that is, which MV should be selected to control which constraint. Based on useful tools such as the relative gain array (RGA), the pairings must be chosen such that the steady-state RGA of the resulting transfer matrix is non-negative and close to the identity matrix at crossover frequencies [24]. Additionally, it is assumed that the control problem is feasible, in other words, there exist solutions of MVs within their limits satisfying all the constraints in the whole disturbance ranges.

Sometimes there may be more constraints than the number of degrees of freedom that will switch between active and inactive, but the number of simultaneously active constraints should be no more than the number of degrees of freedom at the same time. Therefore, after the pairing is complete based on the steady-state RGA rule, there may be remaining constraints, which can be paired with corresponding MVs in the same way to form a nested cascade control structure.

To be specific, as shown in Fig. 4, assume that there are two MVs u_1 and u_2 and three conditionally active constraints g_1, g_2 and g_3 but only two of them will be active at a given time. Two self-optimizing CVs c_1 and c_2 have been obtained based on the proposed LgSOC method. If g_1 and g_3 can become active at the same time, they should be separated in two cascade loops. Then it can be assumed that g_1 and g_3 are paired with u_1 and u_2 respectively according to the steady-state RGA rule. Then the remaining constraint g_2 can also be paired similarly. Assume g_2 is paired with u_1, hence g_1 and g_2 can be nested in cascade with either g_1 or g_2 in the inner or outer loop.

The systematic design procedure of the generalized cascade SOC structure can be summarized by the following steps:

Step 1. Separate simultaneously active constraints in different cascade loops.

Step 2. Pair constraints with suitable MVs based on the steady-state RGA rules.

Step 3. Check that all constraints can be satisfied by adjusting corresponding MVs to ensure feasibility.

Step 4. Remaining constraints (if exist) can be nested in cascade with corresponding MVs paired in the same way.

Step 5. Put self-optimizing CVs in the outermost loop with zero setpoint.

Step 6. Design SISO controllers to compute setpoints for inner loops.

Remark 5: The sequence in which conditionally active constraints are cascaded can be arbitrary in terms of the steady-state sense. Considering the dynamic characteristics of cascade control, variables with fast response should be placed in the inner loop while those with slow response in the outer loop.

Remark 6: As for self-optimizing CVs pairing with MVs, since left multiplication of **H** by a non-singular matrix such as **B** does not change the steady-state economic performance as mentioned in Section III, it can be tactfully selected as
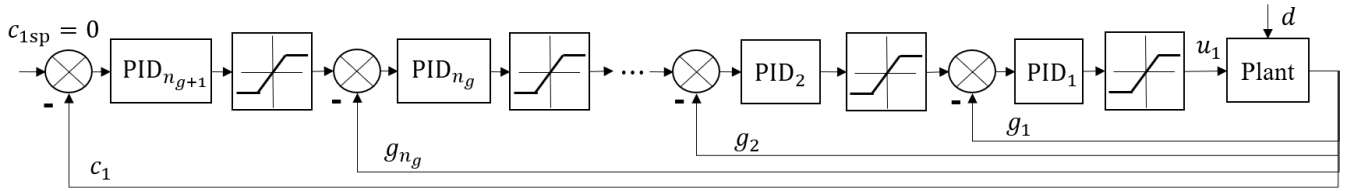
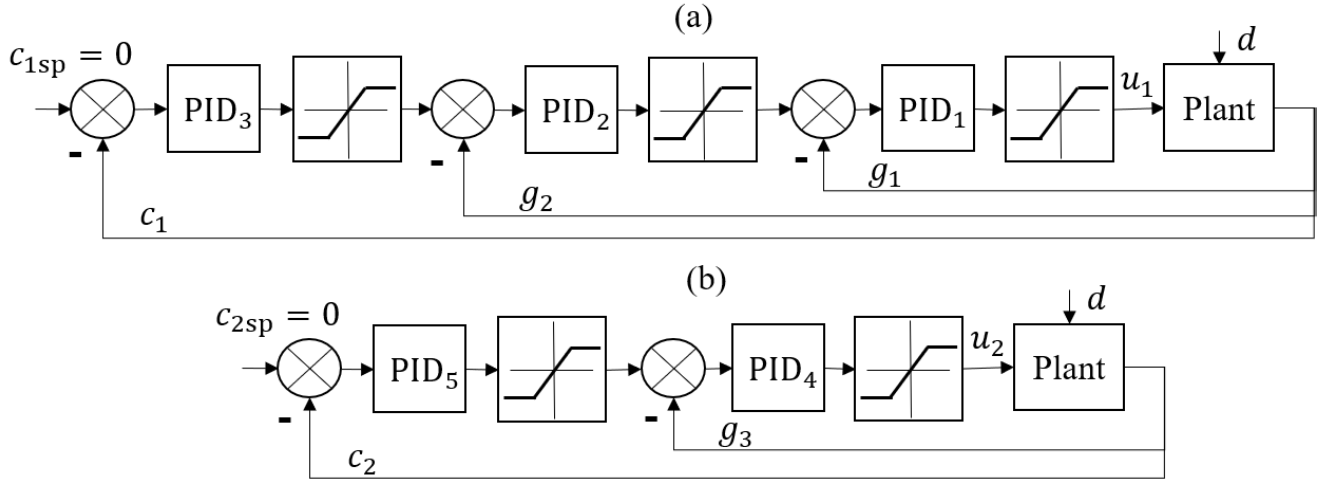**FIGURE 3.** Generalized cascade control structure.

FIGURE 4. Generalized cascade SOC structure for more constraints than the number of degrees of freedom.

$\mathbf{B} = (\mathbf{H}\mathbf{f}_u)^{-1}$ so that it is decoupled from self-optimizing CVs to MVs. Therefore, any MV can be chosen to pair based on the decoupling calculation.

Remark 7: Note that the cascade structure can ensure all constraints satisfied at steady-state. However, the limits in the saturation block should be considered with a suitable back-off to avoid possible dynamic violation under worst case, which introduces conservatism.

V. CASE STUDY

A. TOY EXAMPLE

Consider the following toy example

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

$$\text{s.t. } g = \frac{1}{4}u^2 - \frac{1}{2}d \le 0 \qquad (29)$$

with available measurements

$$y_1 = u \qquad (30)$$

$$y_2 = \frac{1}{4}u^2 - \frac{1}{2}d \qquad (31)$$

where u and d are both scalars. The disturbance range is uniformly distributed between 0.5 to 4, *i.e.*, $d \in [0.5, 4]$.

Through solving the optimization problem (29), the optimal value of the constraint g is conditionally active as shown

in Fig. 5 and the whole disturbance region can be divided into two parts accordingly. Region I ($d \in [0.5, 2]$) is fully unconstrained with $g < 0$ while Region II ($d \in (2, 4]$) is fully constrained with $g = 0$.

To find the optimal self-optimizing CV based on the proposed LgSOC method, 100 optimal data are obtained in the whole disturbance region through Monte Carlo simulation, and then the optimal CV can be computed

$$c = \mathbf{H}\mathbf{y} = [0.3773, -0.1487, 1.1377][1, y_1, y_2]^{\mathrm{T}} \qquad (32)$$

The closed-loop steady-state performance by controlling c at zero setpoint is illustrated in Fig. 6. It can be seen that the steady-state value of g after self-optimizing control almost tracks its optimal value in both regions, and the optimality of the system after self-optimizing control is evaluated by the Lagrange-based global average economic loss $\overline{L}_{gav} = 0.0051$. Note that the constraint $g \le 0$ is sometimes violated in Region II, hence it is needed to apply cascade control structure to enforce the constraint satisfied in both regions.

The dynamic SOC performance using cascade control structure is shown in Fig. 7. The disturbance changes from 1 to 3.5 at time=5 s, and at the same time the constraint changes from inactive to active while the self-optimizing CV the opposite with no constraint violation when the system reaches steady-state.
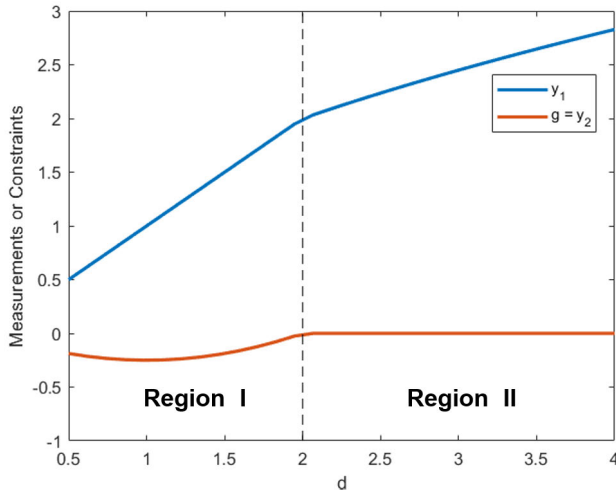
FIGURE 5. Optimal values of measurements as a function of the disturbance.
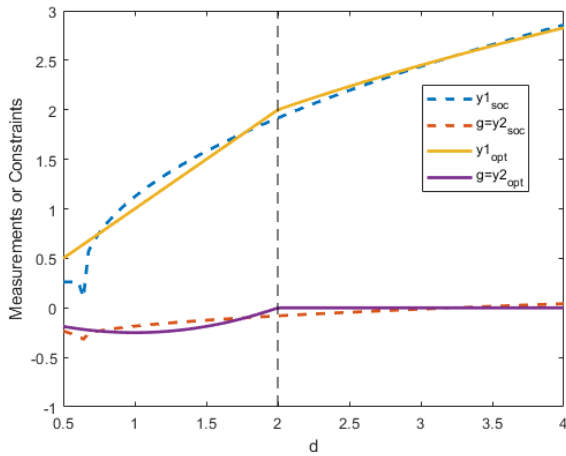

FIGURE 6. Closed-loop steady-state performance of SOC.

B. EVAPORATOR MODEL

1) MODEL DESCRIPTION

A benchmark process, the forced-circulation evaporator [25], is investigated for testing the proposed approach as shown in Fig. 8.

There are in total 20 process variables involved in the process. The description of the 20 process variables and their nominal optimal values are listed in Table 1. Detailed model equations are given in the appendix.

The objective cost function comprises five terms related to steam, water and pumping, as well as the raw material cost and the product value, which is defined as [26]

$$J = 600F_{100} + 0.6\,F_{200} + 1.009\,(F_2 + F_3) + 0.2\,F_1 - 4800F_2 \tag{33}$$

There are 3 state variables, 5 manipulated variables and 3 disturbance variables listed as below

$$\mathbf{x} = [L_2\ X_2\ P_2]^{\mathrm{T}} \tag{34}$$

TABLE 1. Process variables and their nominal optimal values.

Variable	Description	Value	Unit
F_1	Feed Flow Rate	9.4697	kg/min
F_2	Product Flow Rate	1.3338	kg/min
F_3	Circulating Flow Rate	24.7273	kg/min
F_4	Vapor Flow Rate	8.1360	kg/min
F_5	Condensate Flow Rate	8.1360	kg/min
X_1	Feed Composition	5	%
X_2	Product Composition	35.5	%
T_1	Feed Temperature	40	°C
T_2	Product Temperature	88.4013	°C
T_3	Vapor Temperature	81.0695	°C
L_2	Separator Level	1	%
P_2	Operating Pressure	51.4138	kPa
F_{100}	Steam Flow Rate	9.4350	kg/min
T_{100}	Steam Temperature	151.5134	°C
P_{100}	Steam Pressure	400	kPa
Q_{100}	Heat Duty	345.3194	kW
F_{200}	Cooling Water Flowrate	217.7531	kg/min
T_{200}	Inlet Temperature of Cooling Water	25	°C
T_{201}	Outlet Temperature of Cooling Water	45.5498	°C
Q_{200}	Condenser Duty	313.2350	kW

$$\mathbf{u} = [F_{200}\ F_1\ F_2\ F_3\ P_{100}]^{\mathrm{T}} \tag{35}$$

$$\mathbf{d} = [X_1\ T_1\ T_{200}]^{\mathrm{T}} \tag{36}$$

where the ranges of disturbances are defined as $\pm 5\,\%$ for X_1 and $\pm 20\,\%$ for both T_1 and T_{200} of their corresponding nominal optimal values. There are 10 available measurement variables in total shown as follows

$$\mathbf{y} = [F_5\ T_2\ T_3\ F_{100}\ T_{201}\ P_2\ F_2\ F_{200}\ F_3\ F_1]^{\mathrm{T}} \tag{37}$$

The measurement noise for the pressure and flow measurements are taken to be $\pm 2.5\,\%$ and $\pm 2\,\%$, respectively, of the nominal operating values. For temperature measurements, it is considered to be within $\pm 1\,°C$.

Several constraints related to operational safety and product quality are listed below

$$X_2 \geq 35.5\,\% \tag{38}$$

$$40\,\mathrm{kPa} \leq P_2 \leq 80\,\mathrm{kPa} \tag{39}$$

$$P_{100} \leq 400\,\mathrm{kPa} \tag{40}$$

$$0\,\mathrm{kg/min} \leq F_{200} \leq 400\,\mathrm{kg/min} \tag{41}$$

$$0\,\mathrm{kg/min} \leq F_1 \leq 20\,\mathrm{kg/min} \tag{42}$$

$$0\,\mathrm{kg/min} \leq F_3 \leq 100\,\mathrm{kg/min} \tag{43}$$

In order for the plant to work optimally, active constraint control is always required to maintain the active constraints at their limits [1]. In this case, there are two active constraints $X_2 = 35.5\,\%$ and $P_{100} = 400\,\mathrm{kPa}$ at the optimal point. From a physical perspective, it requires more steam, water and pumping cost to achieve a higher product composition, so X_2 is active at its minimum. For the second active constraint, reducing P_{100} will increase F_3 due to energy balance, thus increasing the objective cost since the sensitivity of F_3 to the objective is larger than that of P_{100} [18]. Therefore, P_{100} should be controlled at its maximum. According to the rule of active constraint control, these two active constraints consume two degrees of freedom. Since L_2 is unstable,
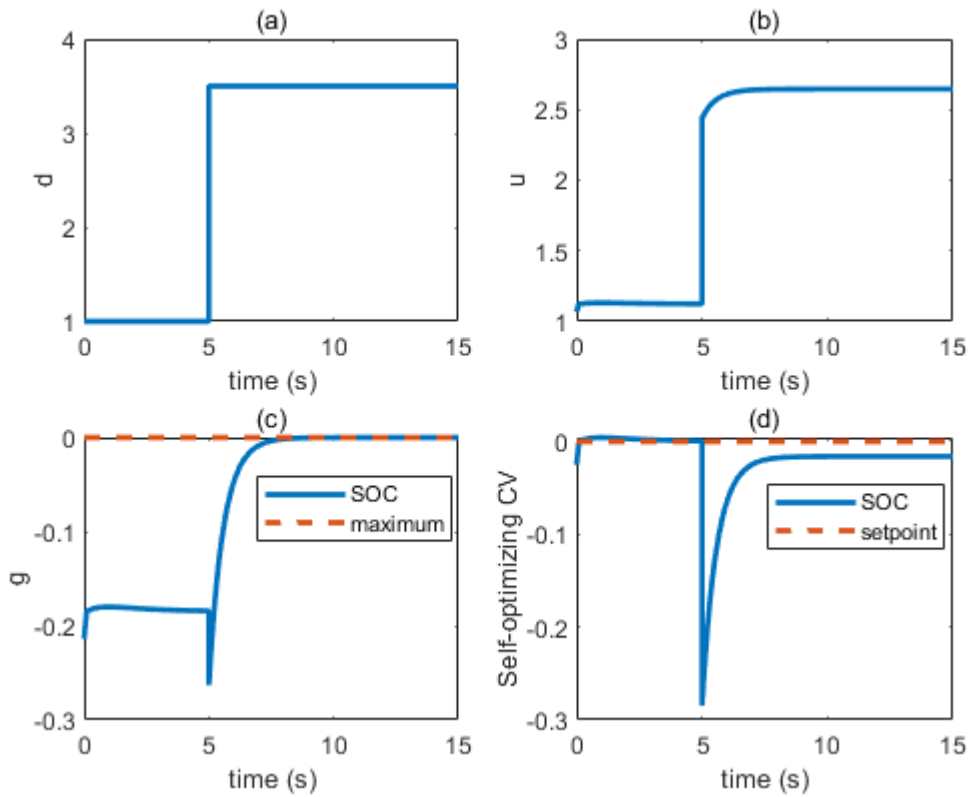
FIGURE 7. Dynamic performance using cascade constrained SOC structure.
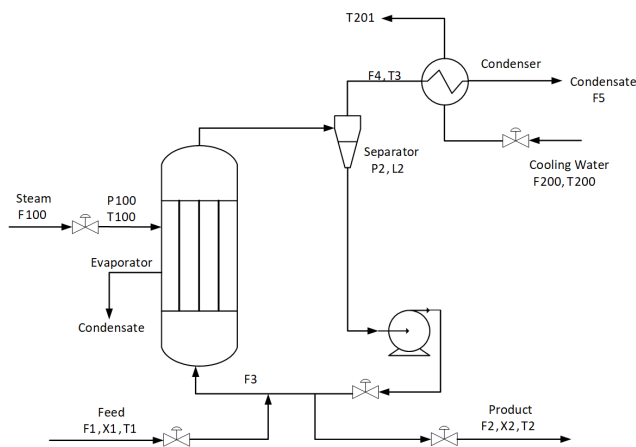

FIGURE 8. Evaporator model.

it should also be controlled, thus consuming another degree of freedom. Therefore, the remaining available manipulated variables are $\mathbf{u} = [F_{200} \; F_1]^T$ without loss of generality.

Besides, F_{200}, F_1 and F_3 are always inactive within the predefined disturbance range, which means there is no input saturation, hence they can be ignored and need not to be controlled. As for the constraint (39), the optimal value of the operating pressure P_2 varies from 40 kPa to 80 kPa depending on different disturbances and maintaining P_2 at a constant

setpoint cannot necessarily minimize the economic objective function.

2) RESULTS AND DISCUSSIONS
a: STEADY-STATE EVALUATIONS
The whole disturbance region is randomly sampled into $N = 2000$ scenarios. Optimal data including measurements and manipulated variables are obtained through minimizing the cost function under these disturbances. Note that the optimal values of P_2 in the constraint (39) vary between active and inactive within the whole disturbance scenarios. Then the matrices such as $\mathbf{W}, \mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{f}_u$ and $\mathcal{L}_{uu}$ can be obtained to find optimal CVs.

The optimal measurement subsets corresponding to the number of measurements from 2 to 10 are selected using LgSOC method based on searching algorithm such as exhaustive search method. As shown in Fig. 9, with more measurement variables involved in CV selection, the minimal average economic loss gradually decreases and it reaches the minimum of 2.2268 when all measurement variables are included. It can be seen that when only 2 measurements are chosen, it gives the average loss of 18.6362 and there is a drop when $n_y = 3$ (n_y denotes the number of measurements selected). It is clear that there is no need to use all of the measurements considering the cost of sensors and installation, and $n_y = 3$ or 4 can achieve relatively good performance.

TABLE 2. Nonlinear model evaluated loss using 100 random samples with measurement noise ($n_Y = 4$).

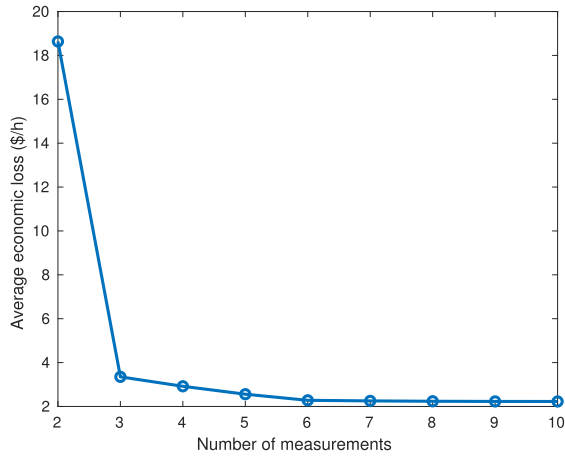| Method	Measurement subset	Optimal $\mathbf{H}$					Average	Maximum	Standard Deviation
constraint estimation method	$[F_5\ P_2\ F_2\ F_{200}]$	-10.7685 -20.2978	6.8298 7.2582	0.0882 0.3428	-49.2050 -43.6534	0.0756 -0.0002	6.4309	23.9875	6.3261
LgSOC method	$[F_5\ P_2\ F_2\ F_{200}]$	-10.7513 -22.4888	6.7779 13.8746	0.0889 0.2514	-48.9480 -76.4355	0.0756 -0.0087	1.6869	6.3133	1.7029
LgSOC method	$[F_5\ F_{100}\ F_2\ F_{200}]$	-10.2683 -24.3848	2.5153 6.0142	4.2310 8.4050	-49.0691 -74.1862	0.0713 -0.0241	1.3176	8.7723	1.6075
retrofitted LgSOC method	$[F_5\ F_{100}\ F_2\ F_{200}]$	-9.8799 -23.7703	2.0495 5.0443	4.5567 9.1989	-49.0398 -74.4186	0.0727 -0.0238	1.2876	8.5796	1.5535

FIGURE 9. Minimal average economic loss against different measurement subset size.

It is further investigated when choosing $n_y = 4$ as measurement subset by 100 random samples through Monte Carlo simulation. For demonstration, another gSOC method handling variant active constraint set, namely the constraint estimation method [20] (which is denoted as method 2 in the reference), is compared with the proposed methods. The nonlinear model evaluated economic loss and the corresponding optimal linear combination matrix $\mathbf{H}$ are tabulated in Table 2. More specifically, for the first one, the best measurement subset when $n_y = 4$ is selected using constraint estimation method through exhaustive search, which is $[F_5\ P_2\ F_2\ F_{200}]$ and meanwhile the corresponding optimal $\mathbf{H}$ is obtained using constraint estimation method. For the second one, the subset is the same as the first one, but the corresponding optimal $\mathbf{H}$ is obtained using LgSOC method. As for the last two ones, the best measurement subset when $n_y = 4$ is found using LgSOC method through exhaustive search, which is $[F_5\ F_{100}\ F_2\ F_{200}]$ and the corresponding optimal $\mathbf{H}$ is obtained using LgSOC method and its retrofitted version respectively. The difference between the two optimal measurement subsets lies in P_2 and F_{100}, which is probably because J is more sensitive to F_{100} than P_2, hence including F_{100} instead of P_2 may perform better.

Comparing these four cases, the average economic loss decreases in order. More specifically, the average, maximum and standard deviation of the economic loss of the second case are largely reduced compared with those of the first one

and the economic performance of the last two cases are better in further, which indicates that given the same subset, the LgSOC method performs better than the constraint estimation method, and furthermore, the LgSOC method can give even better performance when the optimal subset is chosen using the LgSOC method. In addition, by comparing the last two, the retrofitted LgSOC method can achieve even better performance than the LgSOC method, which demonstrates the effectiveness of considering multiple reference points. Therefore, through the comparison, it is clear to see that the proposed LgSOC method can achieve better steady-state economic performance than what the existing method achieves.

b: DYNAMIC SIMULATIONS
As for online implementation of the proposed LgSOC method, the proposed cascade control structure is applied in the dynamic simulation under disturbances and measurement noises. The optimal matrix $\mathbf{H} =$

$$\begin{bmatrix} -10.2683 & 2.5153 & 4.2310 & -49.0691 & 0.0713 \\ -24.3848 & 6.0142 & 8.4050 & -74.1862 & -0.0241 \end{bmatrix}$$

is applied in the dynamic simulation. To decouple at the reference point, the combination matrix $\mathbf{H}$ calculated above was left multiplied by a non-singular matrix $\mathbf{B}$, such as $\mathbf{B} = \left(\mathbf{H}\mathbf{f}_{u,(r)}\right)^{-1}$, which leads to $\mathbf{B}\mathbf{H}\mathbf{f}_{u,(r)} = \mathbf{I}$. Therefore, the combination matrix used in dynamic simulations is modified into $\mathbf{H}^* = \mathbf{B}\mathbf{H}$. Note that this will not affect the steady-state economic properties but can improve the dynamic performance.

To avoid constraint violation, the proposed generalized cascade control is adopted to close the self-optimizing CVs in the outer loop while close the constrained variables in the inner loop. Since there are two self-optimizing CVs, namely $\mathbf{c}_1$ and $\mathbf{c}_2$, and only one constraint P_2 which may vary between active and inactive, there should be one cascade control loop and another simple feedback control loop. As shown in Fig. 10, P_2 is closed in the inner loop, whose setpoint is determined by the outer loop controller, while $\mathbf{c}_1$ is controlled in the outer loop at zero setpoint by adjusting F_{200}. The saturation block restricts P_2 not to exceed its limits and a suitable back-off is also needed to avoid dynamic constraint violation. The other self-optimizing CV $\mathbf{c}_2$ is simply controlled at zero by F_1. Since P_2 is more sensitive to F_{200} than F_3, P_2 is cascaded by F_{200}. In this way, the self-optimizing control loop and the constraint control loop will automatically switch
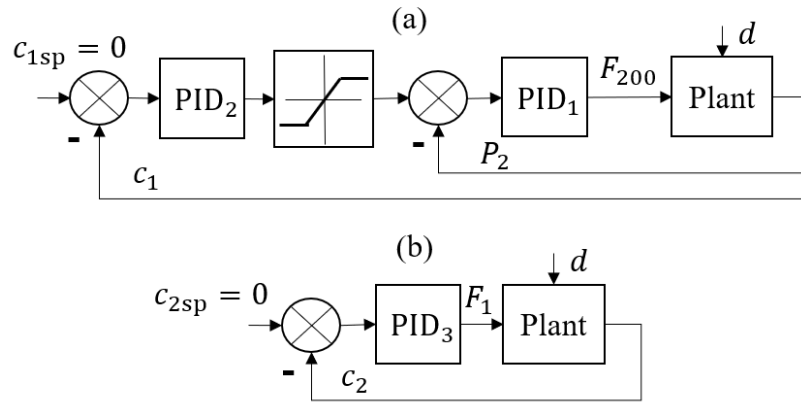
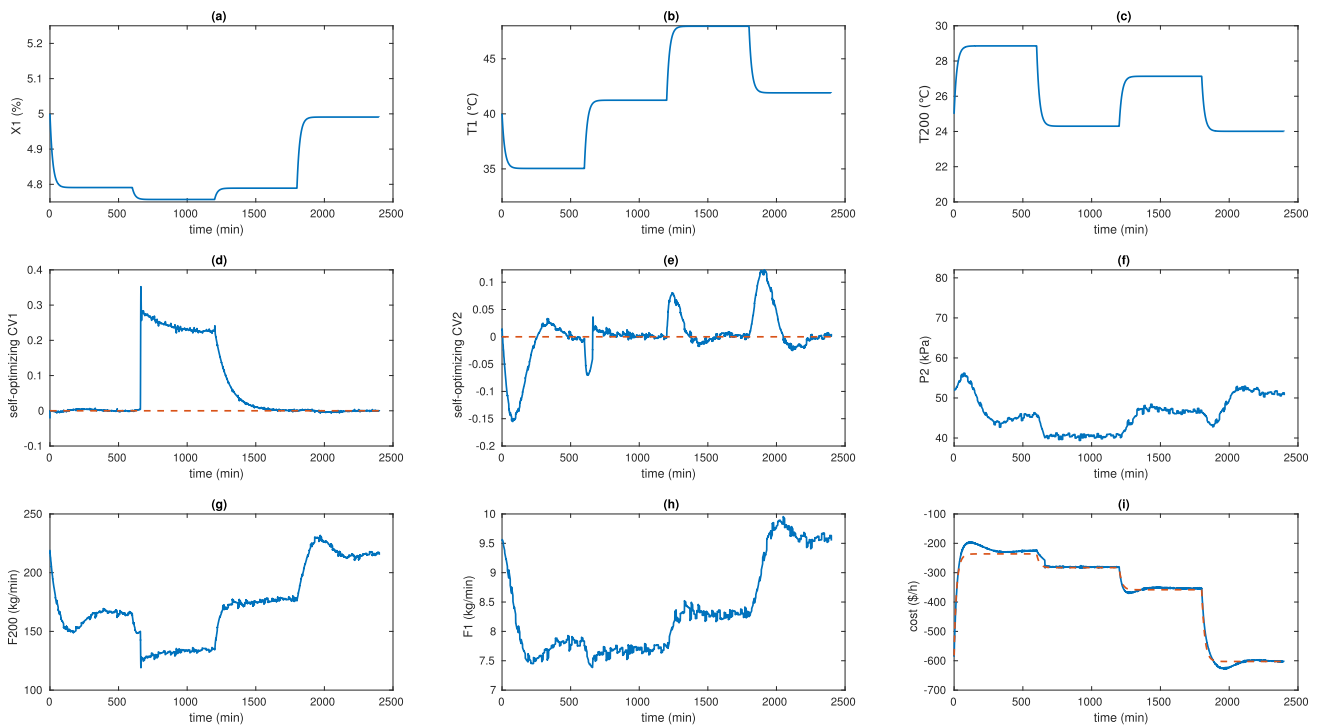**FIGURE 10.** Control structure for self-optimizing controlled variables.

FIGURE 11. Simulation results of self-optimizing control (blue solid line: actual values; red dashed line: setpoints or optimal values): (a) Feed composition, (b) Feed temperature, (c) Cooling water inlet temperature, (d) Self-optimizing controlled variable 1, (e) Self-optimizing controlled variable 2, (f) Operating pressure, (g) Cooling water flowrate, (h) Feed flow rate, (i) Economic cost.

between active and inactive to ensure system optimality all the time. As for the configuration of rest control loops, L_2 is controlled by F_3, X_2 is controlled by F_2 and P_{100} is controlled by itself [18]. PI controllers are adopted in all closed loops and the parameters are tuned according to the SIMC rule [27].

The dynamic simulation results of a 2400-min operation are shown in Fig. 11. Three disturbances X_1, T_1 and T_{200} vary in their predefined ranges, and their changing scenarios are shown in Fig. 11(a-c) with a time interval of 600 min. Note that the disturbances do not change abruptly and sharply but gradually. Gaussian noise within the predefined variation ranges is added to all the measurement variables and

manipulated variables. The self-optimizing CV $\mathbf{c}_1$ is tightly controlled at zero except for 700 to 1200 min in Fig. 11(d), during which time the constrained variables P_2 is active at 40 kPa as indicated in Fig. 11(f). This is because that during that time F_{200} is used to control P_2 at its minimum while $\mathbf{c}_1$ is no longer under control hence does not equal zero. From Fig. 11(e), the deviation of $\mathbf{c}_2$ to zero is relatively small by manipulation F_1. It is clear that in Fig. 11(i) the actual economic cost tracks the optimal values well and it achieves good steady-state economic performance, which indicates that the proposed LgSOC method works quite well in the dynamic case.

VI. CONCLUSION

This paper rebuilds the theoretical basis so that the constraint activeness varying gSOC problem can be solved as easy as the invariant constraint activeness problems. The Lagrange-based global average economic loss is derived to unify the optimality evaluation when only a single set of self-optimizing CVs is applied in the whole operational region. Also, it can be used to obtain optimal set of CVs by minimizing the new Lagrange-based loss function over the entire operational space. To facilitate the optimization process, the Lagrange-based gSOC method and its retrofitted version are proposed based on the short-cut algorithm of gSOC. Furthermore, the one degree of freedom cascade SOC structure is generalized to multiple degrees of freedom cases to ensure all constraints satisfied in the whole region. As demonstrated in the numerical example and the evaporator benchmark, the proposed methods achieve both good steady-state and dynamic performance.

APPENDIX

A. PROOF OF THE NON-UNIQUENESS OF **H**

Suppose $\overline{\mathbf{H}} = \mathbf{BH}$, where $\mathbf{H}$ is the solution and $\mathbf{B}$ is an arbitrary non-singular matrix. Then $(\mathbf{BH})^{\mathrm{T}}\overline{\mathcal{L}}_{cc}(\mathbf{BH}) = (\mathbf{BH})^{\mathrm{T}}(\mathbf{BHf}_u)^{-\mathrm{T}}\mathcal{L}_{uu}(\mathbf{BHf}_u)^{-1}(\mathbf{BH}) = \mathbf{H}^{\mathrm{T}}\mathcal{L}_{cc}\mathbf{H}$, so $\overline{L}_{\mathrm{gav}}(\overline{\mathbf{H}}) = \overline{L}_{\mathrm{gav}}(\mathbf{H})$, indicating that $\overline{\mathbf{H}} = \mathbf{BH}$ is another solution.

B. DETAILED MODEL EQUATIONS OF THE EVAPORATOR

$$\frac{\mathrm{d}L_2}{\mathrm{d}t} = \frac{F_1 - F_4 - F_2}{20}$$
$$\frac{\mathrm{d}X_2}{\mathrm{d}t} = \frac{F_1 X_1 - F_2 X_2}{20}$$
$$\frac{\mathrm{d}P_2}{\mathrm{d}t} = \frac{F_4 - F_5}{4}$$
$$T_2 = 0.5616\,P_2 + 0.3126\,X_2 + 48.43$$
$$T_3 = 0.507\,P_2 + 55.0$$
$$F_4 = \frac{Q_{100} - 0.07\,F_1\,(T_2 - T_1)}{38.5}$$
$$T_{100} = 0.1538\,P_{100} + 90.0$$
$$Q_{100} = 0.16\,(F_1 + F_3)\,(T_{100} - T_2)$$
$$F_{100} = \frac{Q_{100}}{36.6}$$
$$Q_{200} = \frac{0.9576\,F_{200}\,(T_3 - T_{200})}{0.14\,F_{200} + 6.84}$$
$$T_{201} = T_{200} + \frac{13.68\,(T_3 - T_{200})}{0.14\,F_{200} + 6.84}$$

REFERENCES

[1] S. Skogestad, "Plantwide control: The search for the self-optimizing control structure," *J. Process Control*, vol. 10, no. 5, pp. 487–507, Oct. 2000.

[2] J. Jäschke, Y. Cao, and V. Kariwala, "Self-optimizing control—A survey," *Annu. Rev. Control*, vol. 43, pp. 199–223, Jan. 2017.

[3] C. Y. Chen and B. Joseph, "On-line optimization using a two-phase approach: An application study," *Ind. Eng. Chem. Res.*, vol. 26, no. 9, pp. 1924–1930, Sep. 1987.

[4] I. J. Halvorsen, S. Skogestad, J. C. Morud, and V. Alstad, "Optimal selection of controlled variables," *Ind. Eng. Chem. Res.*, vol. 42, no. 14, pp. 3273–3284, Jul. 2003.

[5] V. Alstad and S. Skogestad, "Null space method for selecting optimal measurement combinations as controlled variables," *Ind. Eng. Chem. Res.*, vol. 46, no. 3, pp. 846–853, Jan. 2007.

[6] V. Alstad, S. Skogestad, and E. S. Hori, "Optimal measurement combinations as controlled variables," *J. Process Control*, vol. 19, no. 1, pp. 138–148, Jan. 2009.

[7] L. Ye, Y. Cao, Y. Li, and Z. Song, "Approximating necessary conditions of optimality as controlled variables," *Ind. Eng. Chem. Res.*, vol. 52, no. 2, pp. 798–808, Jan. 2013.

[8] L. Ye, Y. Cao, Y. Li, and Z. Song, "A data-driven approach for selecting controlled variables," *IFAC Proc. Volumes*, vol. 45, no. 15, pp. 904–909, 2012.

[9] L. Ye, Y. Cao, X. Ma, and Z. Song, "A novel hierarchical control structure with controlled variable adaptation," *Ind. Eng. Chem. Res.*, vol. 53, no. 38, pp. 14695–14711, Sep. 2014.

[10] J. Jäschke and S. Skogestad, "NCO tracking and self-optimizing control in the context of real-time optimization," *J. Process Control*, vol. 21, no. 10, pp. 1407–1416, Dec. 2011.

[11] S. A. Girei, Y. Cao, A. S. Grema, L. Ye, and V. Kariwala, "Data-driven self-optimizing control," *Comput. Aided Chem. Eng.*, vol. 33, pp. 649–654, Jan. 2014.

[12] L. Ye, Y. Cao, and X. Yuan, "Global approximation of self-optimizing controlled variables with average loss minimization," *Ind. Eng. Chem. Res.*, vol. 54, no. 48, pp. 12040–12053, Dec. 2015.

[13] L. Ye, Y. Cao, X. Yuan, and Z. Song, "Retrofit self-optimizing control: A step forward toward real implementation," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 4662–4670, Jun. 2017.

[14] H. Manum and S. Skogestad, "Self-optimizing control with active set changes," *J. Process Control*, vol. 22, no. 5, pp. 873–883, Jun. 2012.

[15] D. Mukherjee, G. L. Raja, and P. Kundu, "Optimal fractional order IMC-based series cascade control strategy with dead-time compensator for unstable processes," *J. Control, Autom. Electr. Syst.*, vol. 32, no. 1, pp. 30–41, Feb. 2021.

[16] M. A. Siddiqui, M. N. Anwar, S. H. Laskar, and M. R. Mahboob, "A unified approach to design controller in cascade control structure for unstable, integrating and stable processes," *ISA Trans.*, vol. 114, pp. 331–346, Aug. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019057820305577

[17] G. L. Raja and A. Ali, "Enhanced tuning of Smith predictor based series cascaded control structure for integrating processes," *ISA Trans.*, vol. 114, pp. 191–205, Aug. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019057820305644

[18] Y. Cao, "Constrained self-optimizing control via differentiation," *IFAC Proc. Volumes*, vol. 37, no. 1, pp. 63–70, Jan. 2004.

[19] W. Hu, L. M. Umar, G. Xiao, and V. Kariwala, "Local self-optimizing control of constrained processes," *J. Process Control*, vol. 22, no. 2, pp. 488–493, Feb. 2012.

[20] L. Ye, Y. Cao, and S. Skogestad, "Global self-optimizing control for uncertain constrained process systems," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 4672–4677, 2017.

[21] H. Su, C. Zhou, Y. Cao, S.-H. Yang, and Z. Ji, "An intelligent approach of controlled variable selection for constrained process self-optimizing control," *Syst. Sci. Control Eng.*, vol. 10, no. 1, pp. 65–72, Dec. 2022.

[22] H. Su, C. Zhou, Y. Cao, and S. Yang, "Machine learning for constrained self-optimizing control," *Comput. Aided Chem. Eng.*, vol. 50, pp. 1203–1208, Jan. 2021.

[23] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Traces and Emergence of Nonlinear Programming*. Cham, Switzerland: Springer, 2014, pp. 247–258.

[24] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*. Chichester, U.K.: Wiley, 2005.

[25] P. L. Lee, R. Newell, and G. Sullivan, "Generic model control—A case study," *Can. J. Chem. Eng.*, vol. 67, no. 3, pp. 478–484, 1989.

[26] V. Kariwala, Y. Cao, and S. Janardhanan, "Local self-optimizing control with average loss minimization," *Ind. Eng. Chem. Res.*, vol. 47, no. 4, pp. 1150–1158, Feb. 2008.

[27] S. Skogestad, "Simple analytic rules for model reduction and PID controller tuning," *J. Process Control*, vol. 13, no. 4, pp. 291–309, Jun. 2003.

HONGXIN SU received the B.S. degree in automation from the China University of Petroleum (East China), Qingdao, China, in 2020. She is currently pursuing the Ph.D. degree in chemical engineering and technology with the College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, China. Her research interests include self-optimizing control and machine learning.

SHUANG-HUA YANG (Senior Member, IEEE) received the B.S. degree in instrument and automation and the M.S. degree in process control from the China University of Petroleum, Beijing, China, in 1983 and 1986, respectively, and the Ph.D. degree in intelligent systems from Zhejiang University, Hangzhou, China, in 1991.

He is currently the Director of the Intelligent Systems Safety, Optimization and Control Laboratory, Zhejiang University, and a Professor and the Head of the Department of Computer Science, University of Reading, U.K. His research interests include industrial process modeling, optimization control, chemical process safety, and chemical system engineering.

Prof. Yang is a fellow of IET and InstMC, U.K. He is an Associate Editor of *IETCyber-Physical Systems: Theory and Applications*, the *International Journal of Computing and Automation*, and the *Journal of InstMC*.
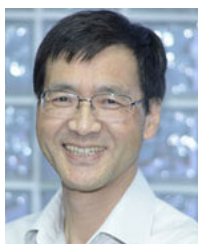
YI CAO (Senior Member, IEEE) received the M.Sc. degree in control engineering from Zhejiang University, Hangzhou, China, in 1985, and the Ph.D. degree in engineering from the University of Exeter, Exeter, U.K., in 1996.

His research interests include advanced process control, plant-wide control, nonlinear system identification, nonlinear model predictive control, and process monitoring.

LINGJIAN YE (Member, IEEE) received the B.Eng. degree in chemical engineering and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2006 and 2011, respectively.

He was with the School of Information Science and Engineering, Ningbo Institute of Technology, Zhejiang University, from 2011 to 2020. He has been a Professor with the School of Engineering, Huzhou University, Huzhou, China, since 2020. His research interests include data-based process monitoring, plant-wide control, real-time optimization, and their applications in process systems.

● ● ●