Shaping of molecular weight distribution using self-optimizing control based on moments

Shaping of molecular weight distribution using self-optimizing control based on moments

Hongxin Su,[†] Chenchen Zhou,[†] Xinhui Tang,[†] Yi Cao,[†,‡] and Shuang-Hua Yang[*,†,‡,¶]

†*College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, Zhejiang, 310058, China*

‡*Institute of Zhejiang University-Quzhou, Quzhou, Zhejiang, 324000, China*

¶*University of Reading, Department of Computer Science, RG6 6UR, UK*

E-mail: shuang-hua.yang@reading.ac.uk

Abstract

It is important but also challenging to control the full shape of the molecular weight distribution in polymerization processes since it is an infinite dimensional probability density function (PDF). In this work, a self-optimizing control (SOC) strategy is adopted to achieve the aim of PDF-shaping by maintaining some elaborately selected controlled variables (CVs) at constant setpoints through online feedback control, even in the presence of uncertainties. To find optimal CVs, finite moments rather than the full shape, which corresponds to an infinite-dimensional space, of the PDF are adopted as elements to parameterize CVs, whilst the optimization problem is to minimize the distance between the actual PDF and the target PDF. The proposed SOC-PDF method is demonstrated more effective than the existing stochastic distribution control method through a pilot semi-batch styrene polymerization case study.

1 Introduction

The molecular weight distribution (MWD) of a polymer has a significant influence on its end-use properties, such as rheological (e.g. viscosity, melt index, etc.), mechanical (e.g. tensile strength, toughness, etc.) and physical and thermal properties (e.g. density, melting point, etc.)[1–4] Therefore, controlling the MWD is a crucial aspect in process control of industrial polymerization processes, which is a well-established fact in the polymer science literature. The objective is to find appropriate control methods that can shape the MWD according to the end-use quality specifications of a polymer.

Nomenclature

$\boldsymbol{\rho}(\tilde{\mathbf{d}})$ Probability density function of $\tilde{\mathbf{d}}$

$\boldsymbol{\xi}$ Augmented vector with $\mathbf{y}$ and $\mathbf{u}$

$\boldsymbol{\xi}_{\mathrm{m}}$ Measured $\boldsymbol{\xi}$ with noise

$\delta_{\mathrm{m}}(i,k)$ Measurement noise of MWD

$\Delta_n(k)$ nth moments of $\delta_{\mathrm{m}}(i,k)$

$\gamma(i,t)$ or $\gamma(i,k)$ Molecular weight distribution (MWD) with respect to chain length i at continuous time t or discrete time k

$\gamma_{\mathrm{m}}(i,k)$ Measured MWD with noise

$\lambda_K(t)$ Kth moments of $C_{\mathrm{P}}(i,t)$ at time t

$\mathbf{c}$ Self-optimizing controlled variables

$\mathbf{c}_{\mathrm{s}}$ Setpoints of self-optimizing controlled variables

$\mathbf{d}$ Exogenous disturbances

$\mathbf{e}_{\tilde{c}}$ Deviation of $\tilde{\mathbf{c}}$ and its optimal trajectory

$\mathbf{H}$ Linear combination matrix

$\mathbf{J}^*_{\tilde{c}\tilde{c}}$ Optimal Hessian matrix of J with respect to $\tilde{\mathbf{c}}$ along the optimal trajectory

$\mathbf{J}^*_{\tilde{c}}$ Optimal gradients of J with respect to $\tilde{\mathbf{c}}$ along the optimal trajectory

$\mathbf{J}^*_{\tilde{u}\tilde{u}}$ Optimal Hessian matrix of J with respect to $\tilde{\mathbf{u}}$ along the optimal trajectory

$\mathbf{n}$ Augmented noise with $\mathbf{n}_y$ and $\mathbf{n}_u$

$\mathbf{n}_u$ Measurement noise of $\mathbf{u}$

$\mathbf{n}_y$ Measurement noise of $\mathbf{y}$

$\mathbf{u}$ Control input

$\mathbf{W}^2$ Covariance matrix of measurement noise $\mathbf{n}$

$\mathbf{x}$ System states

$\mathbf{x}_0$ Initial states

$\mathbf{y}$ Measurement candidate

$\mu_n(k)$ nth moments of $\gamma(i,k)$

$\tilde{\boldsymbol{\xi}}$ Augmented vector with stacked $\boldsymbol{\xi}$ along the time steps and constant 1

$\tilde{\boldsymbol{\xi}}_{\mathrm{m}}$ Augmented vector with stacked $\boldsymbol{\xi}_{\mathrm{m}}$ along the time steps and constant 1

$\tilde{\mathbf{B}}$ Stacked scaling diagonal matrix

$\tilde{\mathbf{c}}$ Stacked self-optimizing controlled variables along the time steps

$\tilde{\mathbf{d}}$ Stacked disturbances along the time steps

$\tilde{\mathbf{G}}^*_{\tilde{\xi}}$ Stacked gradients of $\tilde{\boldsymbol{\xi}}$ with respect to $\tilde{\mathbf{u}}$ along the optimal trajectory

$\tilde{\mathbf{H}}$ Augmented matrix with $\mathbf{H}$ and $\mathbf{c}_\mathrm{s}$

$\tilde{\mathbf{n}}$ Augmented noise with stacked $\mathbf{n}$ along the time steps and constant 0

$\tilde{\mathbf{u}}$ Stacked control input vector along the time steps

$\tilde{\mathbf{W}}^2$ Stacked block diagonal matrix of $\mathbf{W}^2$

$C_\mathrm{D}(i,t)$ Concentration of dead polymer chains with chain length i at time t [mol/L]

$C_\mathrm{i}(t)$ Concentration of initiator at time t [mol/L]

$C_\mathrm{m}(t)$ Concentration of monomer at time t [mol/L]

$C_\mathrm{P}(i,t)$ Concentration of live radicals with chain length i at time t [mol/L]

$C_\mathrm{s}(t)$ Concentration of solvent at time t [mol/L]

D_i Dead polymer chains with a degree of polymerization i

$E_\mathrm{d}, E_\mathrm{i}, E_\mathrm{p}, E_\mathrm{trm}, E_\mathrm{trs}, E_\mathrm{tc}, E_\mathrm{td}$ Activation energy for decomposition, initiation, propagation, chain transfer to monomer and solvent, termination reactions by combination and disproportionation, respectively [J/mol]

f Initiation efficiency

I Initiator

J Objective function

J^* Optimal value of the objective function

k_d Rate coefficient for decomposition [min^{-1}]

k_i Rate coefficient for initiation [min^{-1}]

k_p Rate coefficient for propagation [L/(mol·min)]

k_{tc} Rate coefficient for termination reactions by combination [L/(mol·min)]

k_{td} Rate coefficient for termination reactions by disproportionation [L/(mol·min)]

k_{trm} Rate coefficient for chain transfer to monomer [L/(mol·min)]

k_{trs} Rate coefficient for chain transfer to solvent [L/(mol·min)]

L Loss between the actual and optimal value of the objective function value

L^{d} Loss related to dusturbances

L^{n} Loss related to measurement noise

L_{gav} Global average loss

M Monomer

$m_K(t)$ Kth moments of $C_{\mathrm{D}}(i,t)$ at time t

N_{f} Maximum chain length

P_i Live radicals with a chain length i

Q Weighting factor

R Universal gas constant [cal/(mol·K)]

R^* Primary radicals

S Solvent

T Reaction temperature [K]

Different methods have been devised to control the MWD of polymerization processes in recent decades. Traditional control methods for polymerization processes can be broadly categorized into two types: two-step methods and stochastic distribution control. In two-step methods,[2,5–8] offline optimization is first performed to determine optimal trajectories

for certain measurement variables (optimization layer), such as reactor temperature and/or feed flow rate. These optimal trajectories are then tracked online (control layer). An optimal control solution was proposed to produce a linear emulsion polymer with the desired MWD within a minimum time using online reaction calorimetry.[9] The optimal feed profiles of monomer and chain-transfer agent were calculated through dynamic programming and implemented experimentally for MWD control in a semi-batch nonlinear emulsion copolymerization system.[10] An online two-step method was developed to get the reactor temperature so as to produce the polymer with the desired number-average degree of polymerization.[11] The optimum values for the initiator concentration, chain transfer agent concentration and addition time, and reactor temperature were determined to achieve the desired MWD and then the global linearizing control strategy was applied online to achieve the optimal reactor temperature trajectory.[12] These approaches simplify the control problem by focusing on specific measurable variables, hence it indirectly controls the MWD by manipulating variables such as reactor temperature. This indirect control does not directly influence the shape of the MWD, which can limit its effectiveness in achieving desired polymer properties.

Stochastic distribution control (SDC),[13,14] on the other hand, directly enables the shape of MWD to closely approach a target distribution through online optimization. The MWD is approximated through a linear combination of B-spline basis functions, weighted appropriately. The weights associated with the B-spline approximations are dynamically linked with the control input via ordinary differential equations, and consequently the MWD can be dynamically controlled by the optimal control input calculated through online dynamic optimization. This method provides a more direct approach to achieving the desired MWD but comes with significant computational load for online optimization. Since SDC, through online numerical optimization, directly obtains control inputs, it cannot necessarily guarantee the stability and robustness of the closed-loop system.[15] Therefore, the absence of a well-defined structure for the controller can hamper practical realization. Additionally, SDC lacks clear physical interpretation and suffers from arising complexity due to intricate MWD

shapes.

In order to address these challenges, a control structure design strategy known as self-optimizing control (SOC)[16,17] is first introduced into such field of controlling the shape of a probability density function (PDF). SOC enables the attainment of near-optimal operation by maintaining constant setpoint values for the controlled variables, even in the presence of uncertainties. It effectively transforms the complex online optimization problem into a simpler online feedback control problem by shifting the computational burden to the offline task of selecting appropriate controlled variables, which results in minimal online calculations and computational load. Furthermore, since SOC places emphasis on the meticulous selection of suitable controlled variables rather than directly computing the control input, any feedback control structure such as a Proportional-Integral-Derivative (PID) controller can be chosen to achieve self-optimizing control. Besides, uncertainties are taken into account in the selection of control variables, hence SOC is robust.

The above advantages make SOC different from real-time optimization (RTO) methods. The following comparative analysis highlights the key differences and similarities between self-optimizing control (SOC) and other RTO method such as traditional RTO and model-free RTO based on Bayesian optimization in Table 1.

Nevertheless, making SOC applicable to such PDF-shaping problems, several challenges have to be addressed. Firstly, it is noteworthy that existing SOC methods primarily focus on selecting measurement combinations as controlled variables through the minimization of economic objective functions.[18–30] However, when addressing the PDF-shaping problem, the objective function undergoes a fundamental shift, with economic considerations giving way to minimizing the tracking error between the actual output PDF and the desired target distribution. Furthermore, a critical challenge emerges in determining what variables to control, given the inherent nature of the output PDF as an infinite-dimensional entity that cannot be directly controlled. Drawing inspiration from the moment problem[31] in the field of mathematical statistics, it is recognized that the truncation problem of the MWD function

Table 1: Comparison of SOC, Traditional RTO, and Model-Free RTO with Bayesian Optimization

Aspect	**Self-Optimizing Control (SOC)**	**Traditional Real-Time Optimization (RTO)**	**Model-Free RTO with Bayesian Optimization**
Approach	Model-based with offline optimization	Model-based with repeated real-time optimization	Model-free using probabilistic surrogate models
Optimization Focus	Offline optimization to minimize global average loss	Repeated real-time optimization	Black-box optimization balancing exploration and exploitation
Model Requirements	Requires a detailed mechanistic model	Requires a detailed mechanistic model	Does not require explicit mechanistic models
Online computational Load	Low during real-time control due to offline pre-computation	High due to real-time model updates and optimization	High due to iterative updates and solving optimization problems
Real-Time Performance	High efficiency with quick response time	Potentially slower response due to repeated gradient-based optimization	Converges faster but computationally demanding
Implementation Complexity	Low due to simple online feedback control	High due to repeated model updates and optimization	High due to iterative nature and computational demands
Scalability	Scalable to various systems with shape control requirements	Scalable but complex with large-scale systems	Limited by computational complexity of Gaussian processes

in polymerization reactions has a unique solution, meaning that the MWD can be uniquely determined by its truncated moments.[32] Typically, for computational simplicity, it can be assumed that the MWD follows a specific distribution, such as the Schultz-Zimm distribution, which contains two unknown parameters. Consequently, the determination of the two lower-order moments uniquely defines this MWD. With a finite number of lower-order moments, it is possible to deduce all moment values of the MWD, thereby enabling control of the MWD by controlling a finite number of lower-order moments. Consequently, finite moments (or their combinations), rather than infinite PDF, can be regarded as the controlled variables of

interest.

Therefore, in this work, a novel approach is proposed wherein finite moments of the output PDF are incorporated as candidate measurements. Offline dynamic optimization is solved to minimize the tracking error of the PDF, leading to the derivation of an optimal solution. Concurrently, sensitivity analysis is conducted to gain insights into the system's response. Then, since the existing exact local method[18,29] only identifies locally optimal controlled variables, considering uncertainties arising from exogenous disturbances or/and measurement noise across the entire operational space, the local method may exhibit limited performance. To overcome this limitation, the global SOC (gSOC) method[28] for static systems is generalized to dynamic cases which is employed to identify globally optimal combinations of moments as controlled variables. A comprehensive comparison is then conducted between the proposed global dynamic SOC method for PDF-shaping problems (SOC-PDF) and the SDC method. The effectiveness and efficacy of the SOC-PDF method are validated through a case study involving a pilot styrene polymerization process. For clarity, the differences among existing methods (e.g., Two-step methods and SDC) and the proposed SOC method for PDF-shaping (SOC-PDF) are compared in Table 2.

Table 2: Comparison of Two-Step Methods, SDC, and SOC-PDF

	Two-Step Methods	**SDC**	**SOC-PDF**
Control Approach	Offline optimization, online tracking	Online optimization	Offline CV selection, online feedback control
MWD shape control	Indirect	Direct	Direct
Online computational Load	Low	High	Low
Physical Interpretation	Clear	Unclear	Moderate
Robustness	Moderate	Low	High

This paper makes significant contributions to the field of knowledge in several ways. 1) self-optimizing control is for the first time introduced to address PDF-shaping problems; 2) finite moments are adopted as the controlled variables, replacing the inherently complex and

infinite-dimensional PDF.

The paper is organized as follows. In Section 2, an example for modeling of molecular weight distribution is introduced, and Section 3 derives the dynamic gSOC method for PDF-shaping problems in detail. A case study of the proposed methods is given in Section 4, and finally, the work is concluded in Section **??**section:Conclusion.

2 Molecular weight distribution modeling

For the shape control of molecular weight distribution (MWD), it is necessary that it be measurable, either through direct or indirect means. Various techniques exist for direct measurement of MWD, including the high-temperature gel permeation chromatography (HT-GPC) method,[33] Fourier Transform Infrared Spectroscopy (FTI),[34] and Multi-Angle Light Scattering (MALS),[35] the latter offering real-time insights into molecular weight, size, particle concentration, and more. In cases where sophisticated instrumentation is lacking, soft sensors[36] can serve to estimate the dynamic MWD online based on the equation-oriented (EO) approach.[37] Above approaches ensure that our control methodology remains feasible in practical industrial settings. Therefore, in order to highlight the focus of this article (which is to extend self-optimizing control to the full shape control of a distribution), this paper assumes that the MWD can be measured online, with an exemplar of MWD modeling provided subsequently.

Assume that the kinetic mechanism of a general free-radical polymerization comprises

the following elementary reactions.

Initiator decomposition	$I \xrightarrow{k_\mathrm{d}} 2R^*$
Chain initiation	$R^* + M \xrightarrow{k_\mathrm{i}} P_1$
Propagation	$P_i + M \xrightarrow{k_\mathrm{p}} P_{i+1}$
Chain transfer to monomer	$P_i + M \xrightarrow{k_\mathrm{trm}} P_1 + D_i$
Chain transfer to solvent (chain transfer agent, CTA)	$P_i + S \xrightarrow{k_\mathrm{trs}} P_1 + D_i$
Termination by combination	$P_i + P_j \xrightarrow{k_\mathrm{tc}} D_{i+j}$
Termination by disproportionation	$P_i + P_j \xrightarrow{k_\mathrm{td}} D_i + D_j$

where I, R^*, M and S denote the initiator, primary radicals, monomer and solvent, respectively, and P_i and D_i are the corresponding live and dead polymer chains, with a degree of polymerization i (the number of monomer units in a polymer chain). $K_j = k_j e^{-\frac{E_j}{RT}}$ ($j = \{\mathrm{d,i,p,trm,trs,tc,td}\}$) denotes the reaction constant for decomposition, initiation, propagation, chain transfer to monomer and solvent, and termination reactions by combination and disproportionation respectively, and $K_\mathrm{i} = fK_\mathrm{d}$, where f, k_j, E_j, R and T is the initiation efficiency, rate coefficient, activation energy, universal gas constant and reaction temperature respectively.

The standard model assumptions are listed as follows.[38]

- All of the reaction steps are elementary and irreversible.
- Quasi-steady-state approximation (QSSA) and long-chain hypothesis (LCH) are satisfied.
- The rates of each reaction phase are independent of the live polymer chain length.

Based on the above kinetic mechanism and the model assumptions, the population balance equations for the concentration of live radicals, $C_\mathrm{P}(i,t)$, and dead, $C_\mathrm{D}(i,t)$, polymer

chains of length i are as follows.

$$\begin{aligned}\frac{dC_{\mathrm{P}}(i,t)}{dt} = &\left\{2fk_{\mathrm{d}}C_{\mathrm{i}} + k_{\mathrm{trm}}C_{\mathrm{m}}\sum_{i=1}^{\infty}C_{\mathrm{P}}(i,t) + k_{\mathrm{trs}}C_{\mathrm{s}}\sum_{i=1}^{\infty}C_{\mathrm{P}}(i,t)\right\}\delta_1(i)\\ &+ k_{\mathrm{p}}C_{\mathrm{m}}\left(C_{\mathrm{P}}(i-1,t) - C_{\mathrm{P}}(i,t)\right) - \left(k_{\mathrm{tc}}+k_{\mathrm{td}}\right)C_{\mathrm{P}}(i,t)\sum_{i=1}^{\infty}C_{\mathrm{P}}(i,t)\\ &- \left(k_{\mathrm{trm}}C_{\mathrm{m}} + k_{\mathrm{trs}}C_{\mathrm{s}}\right)C_{\mathrm{P}}(i,t) \qquad (1)\end{aligned}$$

$$\frac{dC_{\mathrm{D}}(i,t)}{dt} = k_{\mathrm{td}}C_{\mathrm{P}}(i,t)\sum_{j=1}^{\infty}C_{\mathrm{P}}(j,t) + \frac{1}{2}k_{\mathrm{tc}}\sum_{j=1}^{i-1}C_{\mathrm{P}}(j,t)C_{\mathrm{P}}(i-j,t) + \left(k_{\mathrm{trm}}C_{\mathrm{m}} + k_{\mathrm{trs}}C_{\mathrm{s}}\right)C_{\mathrm{P}}(i,t) \qquad (2)$$

where C_{i}, C_{m}, C_{s}, C_{P} and C_{D} are the concentration of initiator, monomer, solvent, live radicals and dead polymer chains respectively. $\delta_1(i)$ is the Kronecker delta function (*i.e.*, $\delta_1(i) = 1$ if $i = 1$ and $\delta_1(i) = 0$ if $i \neq 1$). The degree of polymerization i is assumed to only take discrete values ($i = 1, 2, \cdots, N_{\mathrm{f}}$). The maximum number of monomer units N_{f} can range from hundreds to thousands. Therefore, solving the full set of differential equations requires a lot of computation, which is impractical for most relevant scenarios and prevents the real-time use of such a model. Several methods have been developed to address this high-dimensionality problem by reducing the infinite system of differential equations to a low-order system of differential algebra equations such as the method of moments.[39] Thus, the Kth moments of $C_{\mathrm{P}}(i,t)$ and $C_{\mathrm{D}}(i,t)$ are defined as follows.

$$\lambda_K(t) = \sum_{i=1}^{\infty} i^K C_{\mathrm{P}}(i,t), \quad m_K(t) = \sum_{i=2}^{\infty} i^K C_{\mathrm{D}}(i,t), \quad K = 0, 1, \cdots \qquad (3)$$

Then apply Equation (3) into (1) and (2) to obtain the following equations.

$$\begin{aligned}\frac{d\lambda_K(t)}{dt} =& 2fk_{\mathrm{d}}C_{\mathrm{i}} + k_{\mathrm{p}}C_{\mathrm{m}}\sum_{n=0}^{K}\begin{pmatrix} K \\ n \end{pmatrix}\lambda_{K-n} - k_{\mathrm{p}}C_{\mathrm{m}}\lambda_K \\ &+ \left(k_{\mathrm{trm}}C_{\mathrm{m}} + k_{\mathrm{trs}}C_{\mathrm{s}}\right)\lambda_0 - \left(k_{\mathrm{tc}} + k_{\mathrm{td}}\right)\lambda_0\lambda_K - \left(k_{\mathrm{trm}}C_{\mathrm{m}} + k_{\mathrm{trs}}C_{\mathrm{s}}\right)\lambda_K\end{aligned} \tag{4}$$

$$\frac{dm_K(t)}{dt} = k_{\mathrm{td}}\lambda_0\lambda_K + \frac{1}{2}k_{\mathrm{tc}}\sum_{n=0}^{K}\begin{pmatrix} K \\ n \end{pmatrix}\lambda_n\lambda_{K-n} + \left(k_{\mathrm{trm}}C_{\mathrm{m}} + k_{\mathrm{trs}}C_{\mathrm{s}}\right)\lambda_K \tag{5}$$

Finite molecular weight moments have been suggested as a way to calculate MWD in free radical polymerization.[39] Statistical methods such as Markov chain,[40–42] Weibull distribution[43] and Schultz-Zimm distribution[44] have also been used to model some polymerization systems. For many practical problems of linear polymerization under steady-state or quasi-steady state conditions, the generalized Schulz-Flory distribution[45,46] can adequately describe the MWD of polymer chains. Therefore, the full MWD $\gamma(i,t)$ $(i = 1, 2, \cdots, N_{\mathrm{f}})$ can be obtained based on the above methods. Take the Schultz-Zimm distribution as an example and the MWD can be then calculated by

$$\gamma(i,t) = \frac{a(t)^{a(t)}i^{a(t)-1}e^{\frac{-a(t)i}{b(t)}}}{b(t)^{a(t)}\Gamma(a(t))} \tag{6}$$

where

$$\Gamma(a(t)) = \int_0^\infty i^{a(t)-1}e^{-i}di \tag{7}$$

$$a(t) = \frac{m_1^2(t)}{m_0(t)m_2(t) - m_1^2(t)} \tag{8}$$

$$b(t) = \frac{m_1(t)}{m_0(t)} \tag{9}$$

Remark 1 *It is essential to emphasize that the primary focus of the proposed SOC approach is not on the exact calculation of the MWD but rather on assuming that the MWD is known and then selecting the optimal linear combination of the moments of the MWD as controlled variables. This shift in focus means that while accurate modeling of the MWD is important, the strength of the SOC approach lies in its ability to use selected controlled variables to achieve robust control.*

Remark 2 *It should be noted that self-optimizing control possesses the capability to realize the shape control of general distributions, provided they are measurable, with molecular weight distribution being just one particular instance. It is worth highlighting that the SOC approach can be extended beyond the control of MWD in polymerization processes. The proposed SOC method is versatile and can be applied to various systems where the shape of the distribution function needs to be controlled, including particle size distributions (e.g., microscopic particles), pore size distributions, energy-related distributions (e.g., temperature distributions), composition distributions in materials (e.g., alloys) and etc. Therefore, the scope of applications for SOC extends far beyond, encompassing domains such as paper web formation processes,*[47,48] *combustion processes,*[49] *powder manufacture,*[50,51] *traffic flow networks,*[52] *crystallization processes,*[53] *and others.*

3 Dynamic global self-optimizing control of molecular weight distribution

3.1 Introduction of moments

In the context of SOC, an issue emerges that what measurement variables should be controlled. Since MWD is an infinite-dimensional function, making it difficult to directly control it with finite manipulated variables, dimensionality reduction is required such as using the moments of MWD instead of itself. In this case, Herein lies the conundrum: whether the

MWD γ can be uniquely determined by its moments $\mu_n = \sum_{i=1}^{N_\mathrm{f}} i^n \gamma(i)$ $(n = 0, 1, \cdots)$ and the optimal finite number of moments to be preserved.

The first question can be answered using the moment problem in mathematics, which is whether a sequence of real numbers s_n $(n = 0, 1, \cdots)$ can uniquely determine a probability density function on an interval of $I \in \mathbb{R}$. The moment problem consists of two aspects: existence and uniqueness, which are that does there exist a positive measure on I with moments s_n $(n = 0, 1, \cdots)$, and in the affirmative, is this positive measure uniquely determined by the moments s_n $(n = 0, 1, \cdots)$. The probability density function on I with moments s_n $(n = 0, 1, \cdots)$ (if exists) is designated as a solution to the moment problem. If the solution to the moment problem is unique, it is termed determinate; otherwise, it is considered indeterminate. There exist three fundamentally distinct types of (closed) intervals: those with two finite endpoints, those with one finite endpoint, and those with no finite endpoints. In the latter case, the interval is simply $\mathbb{R}$, while in the former two cases, one can envisage $[0, 1]$ and $[0, \infty)$. For historical reasons, the moment problem on $[0, \infty)$ is denoted as the Stieltjes moment problem, and the moment problem on $\mathbb{R}$ is designated as the Hamburger moment problem. Additionally, the moment problem on $[0, 1]$ is known as the Hausdorff moment problem. Should the Stieltjes moment problems and the Hamburger moment problems prove resolvable, they may exhibit an abundance of solutions, rendering them indeterminate moment problems. Whereas, it is elementary in linear algebra to demonstrate that a positive measure with finite support is uniquely characterized by its moments. By employing the approximation theorem of Weierstrass and the Riesz representation theorem, one can extend this outcome to encompass probability density functions with compact support. Consequently, the Hausdorff moment problem is invariably determinate. Concerning existence, Hausdorff[54] established in 1921 that the moment problem possesses a solution on $[0, 1]$ (any finite interval $[a, b]$ can be normalized as $[0, 1]$) if and only if the sequence s_n $(n = 0, 1, \cdots)$, is completely monotonic, which is stated in the following theorem.[55]

Theorem 1 *A necessary and sufficient condition for the Hausdorff moment problem to have*

a solution is that the sequence s_n $(n = 0, 1, \cdots)$ *satisfies the inequalities*

$$(-1)^j \Delta^j s_n = (-1)^j \sum_{i=1}^{j} (-1)^i \binom{j}{i} s_{n+i} \geq 0 \text{ for all } j, n = 0, 1, \cdots \tag{10}$$

If a solution exists, then it is unique.

Since the MWD is generally bounded in the finite support due to the limit of maximum chain length, it is a type of the Hausdorff moment problem. Since μ_n $(n = 0, 1, \cdots)$ is the sequence of moments calculated from the MWD, the existence condition (10) is automatically satisfied. Besides, due to the uniqueness of the Hausdorff moment problem, the MWD γ can be uniquely determined by its moments μ_n $(n = 0, 1, \cdots)$.

However, only finite moments can be available in engineering contexts. The task of reconstructing a density function when armed solely with a finite set of its moments $\boldsymbol{\mu}_m = [\mu_1, \cdots, \mu_m]^{\mathrm{T}}$ is recognized as the truncated moment problem. Many approaches[55–57] have been proposed to solve the problem. It is noteworthy that within the framework of SOC, there is no necessity to contemplate the precise reconstruction of the original distribution using finite moments. Rather, the focus lies on attaining the optimal combination of moments that minimizes the objective loss function. Hence, high-order moments may not be requisite; instead, controlling the combination of lower-order moments at constant values may suffice to steer the system towards optimal operation. Therefore, the optimal number of preserved moments is case-dependent and determined through minimization of the objective loss.

3.2 Problem formulation and dimensionality reduction based on moments

The following dynamic discrete-time optimization problem for PDF-shaping in polymerization systems is considered.

$$\min_{\tilde{\mathbf{u}}} \quad J = \sum_{i=1}^{N_\mathrm{f}} [\gamma(i,N) - \gamma_\mathrm{t}(i)]^2 + \sum_{k=0}^{N-1} \mathbf{u}(k)^\mathrm{T} \mathbf{Q} \mathbf{u}(k) \tag{11}$$

$$\text{s.t.} \quad \mathbf{x}(k+1) = f(\mathbf{x}(k), \mathbf{u}(k), \mathbf{d}(k)) \tag{12}$$

$$\gamma(i,k) = f_\gamma(\mathbf{x}(k), \mathbf{u}(k), \mathbf{d}(k)) \tag{13}$$

$$\gamma_\mathrm{m}(i,k) = \gamma(i,k) + \delta(i,k) \tag{14}$$

$$\mathbf{u}_\mathrm{m}(k) = \mathbf{u}(k) + \mathbf{n}_u(k) \tag{15}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \tag{16}$$

where $\mathbf{x}(k) \in \mathbb{R}^{n_x}$ and $\mathbf{x}_0 \in \mathbb{R}^{n_x}$ denote the system states (such as C_i, C_m, C_s, T, etc) and initial states. $\mathbf{u}(k) \in \mathbb{R}^{n_u}$ and $\mathbf{u}_\mathrm{m}(k) \in \mathbb{R}^{n_u}$ are the true control input and measured control input with implementation error. $\mathbf{d}(k) \in \mathbb{R}^{n_d}$, $\mathbf{n}_u(k) \in \mathbb{R}^{n_u}$ and $\delta(i,k)$ are exogenous disturbances, the implementation error added on control input and measurement noise on the MWD respectively, and they all follow specific distributions. $\gamma(i,k)$ is the MWD with chain length of i ($i = 1, 2, \cdots, N_\mathrm{f}$) at time step k ($k = 0, \cdots, N-1$). $\gamma_\mathrm{t}(i)$ is the target MWD and $\gamma_\mathrm{m}(i,k)$ is the actual measured MWD with noise. f is the nonlinear system function and f_γ is the function to calculate MWD such as the function of Schultz-Zimm distribution (6). J is the objective function to be minimized, where the first term represents the squared error between the actual MWD and the target at the final time, and the second term is the input effort with a weighting factor $\mathbf{Q}$. $\tilde{\mathbf{u}} = \left[\mathbf{u}(0)^\mathrm{T}, \cdots, \mathbf{u}(N-1)^\mathrm{T}\right]^\mathrm{T} \in \mathbb{R}^{Nn_u}$ is the stacked input vector. Since the input variables have been parameterized as piece-wise linear, the optimization problem (11)-(16) can be solved through nonlinear programming to obtain the optimal solution $\tilde{\mathbf{u}}^*(\tilde{\mathbf{d}}) = \left[\mathbf{u}^*(0)^\mathrm{T}, \cdots, \mathbf{u}^*(N-1)^\mathrm{T}\right]^\mathrm{T}$ (where $(\cdot)^*$ denotes the

optimal value of the term) and the optimal value of the objective function $J^*(\tilde{\mathbf{d}})$ (where $\tilde{\mathbf{d}} = \left[\mathbf{d}(0)^{\mathrm{T}}, \cdots, \mathbf{d}(N-1)^{\mathrm{T}}\right]^{\mathrm{T}}$ is the stacked disturbance vector along the time step).

Since the chain length i can be enormous, it is not clear what should be controlled and also difficult to control nearly infinite dimensional distribution functions with finite control input variables. Based on the introduction in the above section, finite moments of $\gamma(i,k)$ (or $\gamma_{\mathrm{m}}(i,k)$) are a promising measurement candidate which can be expressed as follows.

$$\mathbf{y}(k) = \left[\mu_1(k), \cdots, \mu_{n_y}(k)\right]^{\mathrm{T}} \tag{17}$$

$$\mathbf{n}_y(k) = \left[\Delta_1(k), \cdots, \Delta_{n_y}(k)\right]^{\mathrm{T}} \tag{18}$$

$$\mathbf{y}_{\mathrm{m}}(k) = \mathbf{y}(k) + \mathbf{n}_y(k) \tag{19}$$

where $\mu_n(k)$ and $\Delta_n(k)$ are the nth moment of the MWD $\gamma(i,k)$ and the MWD noise $\delta(i,k)$ respectively.

$$\mu_n(k) = \sum_{i=1}^{N_{\mathrm{f}}} i^n \gamma(i,k),\ n = 1, 2, \cdots, n_y \tag{20}$$

$$\Delta_n(k) = \sum_{i=1}^{N_{\mathrm{f}}} i^n \delta(i,k),\ n = 1, 2, \cdots, n_y \tag{21}$$

$\mathbf{y}(k)$ is the set of measurement variables containing moments up to order n_y and $\mathbf{y}_{\mathrm{m}}(k)$ is the measured moment vector with measurement noise $\mathbf{n}_y(k)$ following a certain distribution. Therefore, the high-dimensional MWD $\gamma(i,k)$ is reduced to finite n_y-dimensional moment vector $\mathbf{y}(k)$ which can be used for the design of the self-optimizing controlled variable.

Note that it is crucial to determine the number of finite moments and the criterion is based on the minimization of the global average loss (37) which will be introduced in Section 3.3. The process of determining the number can be summarized as follows.

- Initially, we specify an initial number of moments and minimize the global average loss to obtain the optimal global average loss. We then incrementally increase the number of moments and repeat the optimization process. The optimal number of moments is

determined by analyzing the curve of the optimal global average loss versus the number of moments.

- The turning point on this curve, where increasing the number of moments no longer significantly reduces the optimal global average loss, indicates the optimal number of moments to be used. This method balances the trade-off between the computational complexity and the accuracy of representing the MWD information.

Remark 3 *Since the measurable probability density function is generally discrete with respect to the random variable in practice, $\gamma(i,k)$ is expressed in a discrete form.*

Remark 4 *Note that the set of measurement variables used to design self-optimizing controlled variables is not limited to moments; other process variables can also be included such as temperature, pressure and etc.*

3.3 Offline selection of dynamic global self-optimizing controlled variables

Generally, in most static SOC methods, the primary goal of self-optimizing control is to carefully choose measurement combinations offline, represented as self-optimizing controlled variables (CVs), *i.e* $\mathbf{c} = [\mathbf{c}_\mathrm{s}, \mathbf{H}] \begin{bmatrix} 1 \\ \boldsymbol{\xi} \end{bmatrix}$ (where $\mathbf{H} \in \mathbb{R}^{n_u \times (n_y+n_u)}$ is the combination matrix, $\boldsymbol{\xi} = \left[\mathbf{y}^\mathrm{T}, \mathbf{u}^\mathrm{T}\right]^\mathrm{T} \in \mathbb{R}^{n_y+n_u}$ is the extended measurement with $\mathbf{u}$ included and $\mathbf{c}_\mathrm{s} \in \mathbb{R}^{n_u}$ is the setpoint of CVs $\mathbf{c} \in \mathbb{R}^{n_u}$). The self-optimizing CVs are selected through minimization of the steady-state closed-loop loss function which is the error between the actual value of the objective function and its optimum. The near-optimal operation will be achieved automatically by ensuring that the CVs are maintained at zero setpoints through online feedback control since $\mathbf{c}_\mathrm{s}$ has been included in $\mathbf{c}$.

When extended to dynamic cases, all the variables may vary during the batch and can be stacked to form new stacked matrices or vectors ($\tilde{\mathbf{c}}$, $\tilde{\mathbf{H}}$, $\tilde{\boldsymbol{\xi}}$) along the time step. Hence the

dynamic self-optimizing CVs within a batch can be expressed as

$$\tilde{\mathbf{c}} = \tilde{\mathbf{H}}\tilde{\boldsymbol{\xi}} \tag{22}$$

where

$$\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{c}_\mathrm{s} & \mathbf{H} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{c}_\mathrm{s} & \mathbf{0} & \mathbf{H} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_\mathrm{s} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H} \end{bmatrix} \in \mathbb{R}^{Nn_u \times (N(n_y+n_u)+1)} \tag{23}$$

$$\tilde{\boldsymbol{\xi}} = \left[1, \boldsymbol{\xi}(0)^\mathrm{T}, \cdots, \boldsymbol{\xi}(N-1)^\mathrm{T}\right]^\mathrm{T} \in \mathbb{R}^{N(n_y+n_u)+1} \tag{24}$$

$$\tilde{\mathbf{c}} = \left[\mathbf{c}(0)^\mathrm{T}, \mathbf{c}(1)^\mathrm{T}, \cdots, \mathbf{c}(N-1)^\mathrm{T}\right]^\mathrm{T} \in \mathbb{R}^{Nn_u} \tag{25}$$

$$\mathbf{c}(k) = \mathbf{c}_\mathrm{s} + \mathbf{H}\boldsymbol{\xi}(k) \in \mathbb{R}^{n_u},\ k = 0, 1, \cdots, N-1 \tag{26}$$

$$\boldsymbol{\xi}(k) = \left[\mathbf{y}(k)^\mathrm{T}, \mathbf{u}(k)^\mathrm{T}\right]^\mathrm{T} \in \mathbb{R}^{n_y+n_u} \tag{27}$$

$\tilde{\mathbf{H}}$ is a matrix combined by a time-invariant block diagonal matrix with $\mathbf{H}$ on the diagonal and a column vector with N constant setpoints $\mathbf{c}_\mathrm{s}$, which indicates that only current measurements $\boldsymbol{\xi}(k)$ are utilized to calculate current CVs $\mathbf{c}(k)$. Although there are other structures of $\tilde{\mathbf{H}}$ such as the lower-block triangular matrix (all the measurements up to the present time are used and setpoints are time-varying) and varying block diagonal matrix (sub-matrix on the diagonal and setpoints are both time-varying), it is proved that the constant block diagonal structure is the most robust and easiest to implement,[29,58] and hence it is adopted in this paper.

This dynamic self-optimizing CVs are selected by minimizing the final closed-loop loss function $L = J(\tilde{\mathbf{u}}^\mathrm{fb}, \tilde{\mathbf{d}}) - J^*(\tilde{\mathbf{d}})$ (where the superscript $(\cdot)^\mathrm{fb}$ denotes the closed-loop value of the term). Since $\tilde{\mathbf{u}}^\mathrm{fb}$ is determined after the implementation of the online feedback control for a given $\tilde{\mathbf{H}}$, the evaluation of the loss function is not straightforward. Thus the second-order Taylor expansion in terms of the free variable $\tilde{\mathbf{c}}$ is employed to approximate the closed-loop

loss.

$$L = J(\tilde{\mathbf{u}}^{\mathrm{fb}}, \tilde{\mathbf{d}}) - J^*(\tilde{\mathbf{d}}) \approx \mathbf{J}_{\tilde{c}}^{*\mathrm{T}} \mathbf{e}_{\tilde{c}} + \mathbf{e}_{\tilde{c}}^{\mathrm{T}} \mathbf{J}_{\tilde{c}\tilde{c}}^{*} \mathbf{e}_{\tilde{c}} = \mathbf{e}_{\tilde{c}}^{\mathrm{T}} \mathbf{J}_{\tilde{c}\tilde{c}}^{*} \mathbf{e}_{\tilde{c}} \tag{28}$$

where $\mathbf{e}_{\tilde{c}} = \tilde{\mathbf{c}}^{\mathrm{fb}} - \tilde{\mathbf{c}}^*$ is the CV deviation of $\tilde{\mathbf{c}}$ around its optimal trajectory $\tilde{\mathbf{c}}^* = \left[\mathbf{c}^*(0)^{\mathrm{T}}, \cdots, \mathbf{c}^*(N-1)^{\mathrm{T}}\right]^{\mathrm{T}}$. $\mathbf{J}_{\tilde{c}}^* = \frac{\partial J}{\partial \tilde{\mathbf{c}}}|_{\tilde{\mathbf{c}}=\tilde{\mathbf{c}}^*} \in \mathbb{R}^{Nn_u}$ is the gradient of J with respect to $\tilde{\mathbf{c}}$ along the optimal trajectory, hence $\mathbf{J}_{\tilde{c}}^* = \mathbf{0}$. $\mathbf{J}_{\tilde{c}\tilde{c}}^* = \frac{\partial^2 J}{\partial \tilde{\mathbf{c}}^2}|_{\tilde{\mathbf{c}}=\tilde{\mathbf{c}}^*} \in \mathbb{R}^{Nn_u \times Nn_u}$ is the Hessian matrix of J with respect to $\tilde{\mathbf{c}}$ along the optimal trajectory, which can be obtained by

$$\mathbf{J}_{\tilde{c}\tilde{c}}^* = (\tilde{\mathbf{H}}\tilde{\mathbf{G}}_{\tilde{\xi}}^*)^{-\mathrm{T}} \mathbf{J}_{\tilde{u}\tilde{u}}^* (\tilde{\mathbf{H}}\tilde{\mathbf{G}}_{\tilde{\xi}}^*)^{-1} \in \mathbb{R}^{Nn_u \times Nn_u} \tag{29}$$

where

$$\tilde{\mathbf{G}}_{\tilde{\xi}}^* = \left.\begin{bmatrix} \mathbf{0}_{1\times n_u} & \mathbf{0}_{1\times n_u} & \cdots & \mathbf{0}_{1\times n_u} \\ \frac{\partial \boldsymbol{\xi}(0)}{\partial \mathbf{u}(0)} & \mathbf{0}_{(n_y+n_u)\times n_u} & \cdots & \mathbf{0}_{(n_y+n_u)\times n_u} \\ \frac{\partial \boldsymbol{\xi}(1)}{\partial \mathbf{u}(0)} & \frac{\partial \boldsymbol{\xi}(1)}{\partial \mathbf{u}(1)} & \cdots & \mathbf{0}_{(n_y+n_u)\times n_u} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \boldsymbol{\xi}(N-1)}{\partial \mathbf{u}(0)} & \frac{\partial \boldsymbol{\xi}(N-1)}{\partial \mathbf{u}(1)} & \cdots & \frac{\partial \boldsymbol{\xi}(N-1)}{\partial \mathbf{u}(N-1)} \end{bmatrix}\right|_{\mathbf{u}(k)=\mathbf{u}^*(k)} \in \mathbb{R}^{(N(n_y+n_u)+1)\times Nn_u} \tag{30}$$

$$\mathbf{J}_{\tilde{u}\tilde{u}}^* = \left.\begin{bmatrix} \frac{\partial J^2(0)}{\partial \mathbf{u}^2(0)} & \frac{\partial J^2(0)}{\partial \mathbf{u}(0)\mathbf{u}(1)} & \cdots & \frac{\partial J^2(0)}{\partial \mathbf{u}(0)\mathbf{u}(N-1)} \\ \frac{\partial J^2(1)}{\partial \mathbf{u}(1)\mathbf{u}(0)} & \frac{\partial J^2(1)}{\partial \mathbf{u}^2(1)} & \cdots & \frac{\partial J^2(1)}{\partial \mathbf{u}(1)\mathbf{u}(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial J^2(N-1)}{\partial \mathbf{u}(N-1)\mathbf{u}(0)} & \frac{\partial J^2(N-1)}{\partial \mathbf{u}(N-1)\mathbf{u}(1)} & \cdots & \frac{\partial J^2(N-1)}{\partial \mathbf{u}^2(N-1)} \end{bmatrix}\right|_{\mathbf{u}(k)=\mathbf{u}^*(k)} \in \mathbb{R}^{Nn_u \times Nn_u} \tag{31}$$

$\tilde{\mathbf{G}}_{\tilde{\xi}}^*$ is the stacked gradients of $\tilde{\boldsymbol{\xi}}$ with respect to $\tilde{\mathbf{u}}$ along the optimal trajectory, which is a lower-block triangular matrix due to causality. $\mathbf{J}_{\tilde{u}\tilde{u}}^*$ is the optimal Hessian matrix of J with respect to $\tilde{\mathbf{u}}$ along the optimal trajectory, which is positive semi-definite.

Since $\mathbf{y}(k)$ and $\mathbf{u}(k)$ are measured with noise of $\mathbf{n}_y(k)$ and $\mathbf{n}_u(k)$, it can be obtained as

$$\boldsymbol{\xi}_{\mathrm{m}}(k) = \boldsymbol{\xi}(k) + \mathbf{n}(k) \in \mathbb{R}^{n_y+n_u} \tag{32}$$

$$\mathbf{n}(k) = \left[\mathbf{n}_y(k)^{\mathrm{T}}, \mathbf{n}_u(k)^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{n_y+n_u} \tag{33}$$

$$\tilde{\boldsymbol{\xi}}_{\mathrm{m}} = \left[1, \boldsymbol{\xi}_{\mathrm{m}}(0)^{\mathrm{T}}, \cdots, \boldsymbol{\xi}_{\mathrm{m}}(N-1)^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{N(n_y+n_u)+1} \tag{34}$$

$$\tilde{\mathbf{n}} = \left[0, \mathbf{n}(0)^{\mathrm{T}}, \cdots, \mathbf{n}(N-1)^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{N(n_y+n_u)+1} \tag{35}$$

After feedback control with integral function, the measured CVs are supposed to be exactly maintained at zero, thus $\tilde{\mathbf{c}}_{\mathrm{m}}^{\mathrm{fb}} = \tilde{\mathbf{H}}\tilde{\boldsymbol{\xi}}_{\mathrm{m}}^{\mathrm{fb}} = \tilde{\mathbf{H}}(\tilde{\boldsymbol{\xi}}^{\mathrm{fb}} + \tilde{\mathbf{n}}) = \mathbf{0}_{Nn_u\times 1}$ and then $\tilde{\mathbf{H}}\tilde{\boldsymbol{\xi}}^{\mathrm{fb}} = -\tilde{\mathbf{H}}\tilde{\mathbf{n}}$. Thus $\tilde{\mathbf{c}}^{\mathrm{fb}} = \tilde{\mathbf{H}}\tilde{\boldsymbol{\xi}}^{\mathrm{fb}} = -\tilde{\mathbf{H}}\tilde{\mathbf{n}}$. While the optimal value can be obtained as $\tilde{\mathbf{c}}^{*} = \tilde{\mathbf{H}}\tilde{\boldsymbol{\xi}}^{*}$, the deviation can be finally obtained as $\mathbf{e}_{\tilde{c}} = \tilde{\mathbf{c}}^{\mathrm{fb}} - \tilde{\mathbf{c}}^{*} = -\tilde{\mathbf{H}}(\tilde{\boldsymbol{\xi}}^{*} + \tilde{\mathbf{n}})$. Therefore, the closed-loop loss can be simplified to be explicitly related with $\tilde{\mathbf{H}}$ as follows.

$$L \approx (\tilde{\boldsymbol{\xi}}^{*} + \tilde{\mathbf{n}})^{\mathrm{T}}\tilde{\mathbf{H}}^{\mathrm{T}}\mathbf{J}_{\tilde{c}\tilde{c}}^{*}\tilde{\mathbf{H}}(\tilde{\boldsymbol{\xi}}^{*} + \tilde{\mathbf{n}}) \tag{36}$$

where $\mathbf{J}_{\tilde{c}\tilde{c}}^{*}$ can be calculated through (29).

Since it is considered in a global space spanned by $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{n}}$, the average loss should be minimized which can be decomposed into two parts as in the static case.[28]

$$L_{\mathrm{gav}} = \mathbb{E}(L) = \mathbb{E}(L^{d}) + \mathbb{E}(L^{n}) \tag{37}$$

where $\mathbb{E}(\cdot)$ calculates the expected value of the term. L^{d} and L^{n} represent the effect of disturbances and measurement noise on the global average loss respectively.

$$L^{d} = \frac{1}{2}\tilde{\boldsymbol{\xi}}^{*\mathrm{T}}\tilde{\mathbf{H}}^{\mathrm{T}}\mathbf{J}_{\tilde{c}\tilde{c}}^{*}\tilde{\mathbf{H}}\tilde{\boldsymbol{\xi}}^{*},\ L^{n} = \frac{1}{2}\mathrm{tr}(\tilde{\mathbf{W}}^{2}\tilde{\mathbf{H}}^{\mathrm{T}}\mathbf{J}_{\tilde{c}\tilde{c}}^{*}\tilde{\mathbf{H}}) \tag{38}$$

where $\mathrm{tr}(\cdot)$ calculates the trace of a matrix. $\mathbf{W}^{2}$ is the covariance matrix of measurement

noise $\mathbf{n}(k)$ and $\tilde{\mathbf{W}}^2$ is a stacked block-diagonal matrix expressed as follows.

$$\tilde{\mathbf{W}}^2 = \mathrm{diag}\left(0, \underbrace{\mathbf{W}^2, \cdots, \mathbf{W}^2}_{N}\right) \in \mathbb{R}^{(N(n_y+n_u)+1)\times(N(n_y+n_u)+1)} \tag{39}$$

$$\mathbf{W}^2 = \mathbb{E}(\mathbf{n}(k)\mathbf{n}(k)^{\mathrm{T}}) \in \mathbb{R}^{(n_y+n_u)\times(n_y+n_u)},\ k = 0, \cdots, N-1 \tag{40}$$

$$= \mathbb{E}\left(\begin{bmatrix} \Delta_1^2(k) & \Delta_1(k)\Delta_2(k) & \cdots & \Delta_1(k)\Delta_{n_y}(k) & \mathbf{0}_{n_u\times n_u} \\ \Delta_2(k)\Delta_1(k) & \Delta_2^2(k) & \cdots & \Delta_2(k)\Delta_{n_y}(k) & \mathbf{0}_{n_u\times n_u} \\ \vdots & \vdots & \ddots & \vdots & \mathbf{0}_{n_u\times n_u} \\ \Delta_{n_y}(k)\Delta_1(k) & \Delta_{n_y}(k)\Delta_2(k) & \cdots & \Delta_{n_y}^2(k) & \mathbf{0}_{n_u\times n_u} \\ \mathbf{0}_{n_u\times n_u} & \mathbf{0}_{n_u\times n_u} & \mathbf{0}_{n_u\times n_u} & \mathbf{0}_{n_u\times n_u} & \mathbf{n}_u(k)\mathbf{n}_u(k)^{\mathrm{T}} \end{bmatrix}\right) \tag{41}$$

$$= \begin{bmatrix} \sum_{i=a}^{b} i^2\mathbb{E}[\delta^2(i,k)] & \sum_{i=a}^{b} i^3\mathbb{E}[\delta^2(i,k)] & \cdots & \sum_{i=a}^{b} i^{n_y+1}\mathbb{E}[\delta^2(i,k)] & \mathbf{0}_{n_u\times n_u} \\ \sum_{i=a}^{b} i^3\mathbb{E}[\delta^2(i,k)] & \sum_{i=a}^{b} i^4\mathbb{E}[\delta^2(i,k)] & \cdots & \sum_{i=a}^{b} i^{n_y+2}\mathbb{E}[\delta^2(i,k)] & \mathbf{0}_{n_u\times n_u} \\ \vdots & \vdots & \ddots & \vdots & \mathbf{0}_{n_u\times n_u} \\ \sum_{i=a}^{b} i^{n_y+1}\mathbb{E}[\delta^2(i,k)] & \sum_{i=a}^{b} i^{n_y+2}\mathbb{E}[\delta^2(i,k)] & \cdots & \sum_{i=a}^{b} i^{2n_y}\mathbb{E}[\delta^2(i,k)] & \mathbf{0}_{n_u\times n_u} \\ \mathbf{0}_{n_u\times n_u} & \mathbf{0}_{n_u\times n_u} & \mathbf{0}_{n_u\times n_u} & \mathbf{0}_{n_u\times n_u} & \boldsymbol{\sigma}_u^2 \end{bmatrix} \tag{42}$$

$$= \sum_{i=a}^{b} \begin{bmatrix} i^2 & i^3 & \cdots & i^{n_y+1} & \mathbf{0}_{n_u\times n_u} \\ i^3 & i^4 & \cdots & i^{n_y+2} & \mathbf{0}_{n_u\times n_u} \\ \vdots & \vdots & \ddots & \vdots & \mathbf{0}_{n_u\times n_u} \\ i^{n_y+1} & i^{n_y+2} & \cdots & i^{2n_y} & \mathbf{0}_{n_u\times n_u} \\ \mathbf{0}_{n_u\times n_u} & \mathbf{0}_{n_u\times n_u} & \mathbf{0}_{n_u\times n_u} & \mathbf{0}_{n_u\times n_u} & \overline{\boldsymbol{\sigma}}_u^2 \end{bmatrix} \sigma_\gamma^2 \tag{43}$$

$\mathrm{diag}(\cdot)$ represents to create a block-diagonal matrix with the elements in parentheses as the values of the diagonal lines of the matrix. The covariance matrix in (43) is obtained by substituting (18) (21) and (33) into (40). Since the measurement noise added on the true MWD is assumed to be independent and uniformly distributed in time, the variance of $\delta(i,k)$ is equal for all $i \in [a,b]$, *i.e.* $\mathbb{E}(\delta^2(i,k)) = \sigma_\gamma^2$ $(\forall i \in [a,b])$. In addition, since different

manipulated variables are independent of each other, the variance of $\mathbf{u}(k)$ is a diagonal matrix, *i.e.* $\mathbb{E}(\mathbf{n}_u(k)\mathbf{n}_u(k)^{\mathrm{T}}) = \boldsymbol{\sigma}_u^2 \in \mathbb{R}^{n_u \times n_u}$. $\overline{\boldsymbol{\sigma}}_u^2$ is also a diagonal matrix with the elements on its diagonal being half of the elements on the diagonal of $\boldsymbol{\sigma}_u^2$ divided by σ_γ^2. Note that the moments of different orders $\Delta_1(k), \cdots, \Delta_{n_y}(k)$ are correlated and thus the first block matrix on the diagonal of the block diagonal matrix $\mathbf{W}^2$ is full rather than diagonal.

However, since the order of moments of the MWD $(\mu_1(k), \cdots, \mu_{n_y}(k))$ increases exponentially as their order increases, numerical problems will be encountered when obtaining optimal $\tilde{\mathbf{H}}$. Therefore, the moments (and all the available measurements) can be appropriately scaled to resolve the problem. Suppose $\tilde{\mathbf{B}}$ is the stacked scaling diagonal matrix shown as

$$\tilde{\mathbf{B}} = \mathrm{diag}\left(1, \underbrace{\mathbf{B}, \cdots, \mathbf{B}}_{N}\right) \in \mathbb{R}^{(N(n_y+n_u)+1)\times(N(n_y+n_u)+1)} \tag{44}$$

$$\mathbf{B} = \mathrm{diag}\left(b_1, \cdots, b_{n_y}, \mathbf{1}_{1\times n_u}\right) \in \mathbb{R}^{(n_y+n_u)\times(n_y+n_u)} \tag{45}$$

$$\mathbf{b} = \left[b_1, \cdots, b_{n_y}, \mathbf{1}_{1\times n_u}\right] \in \mathbb{R}^{1\times(n_y+n_u)} \tag{46}$$

where $b_1, \cdots, b_{n_y}$ is the scaling factors corresponding to $\mathbf{y}(k)$, while $\mathbf{u}(k)$ remains unscaled for simplicity. Then it can be obtained that

$$\overline{\tilde{\boldsymbol{\xi}}}^* = \tilde{\mathbf{B}}\tilde{\boldsymbol{\xi}}^* \tag{47}$$

$$\overline{\tilde{\mathbf{n}}} = \tilde{\mathbf{B}}\tilde{\mathbf{n}} \tag{48}$$

$$\overline{\tilde{\mathbf{G}}}_{\tilde{\xi}}^* = \tilde{\mathbf{B}}\tilde{\mathbf{G}}_{\tilde{\xi}}^* \tag{49}$$

$$\overline{\tilde{\mathbf{W}}}^2 = \mathrm{diag}\left(0, \underbrace{\overline{\mathbf{W}}^2, \cdots, \overline{\mathbf{W}}^2}_{N}\right) \tag{50}$$

$$\overline{\mathbf{W}}^2 = \mathbf{W}^2 \cdot \mathrm{repmat}(\mathbf{b}) \tag{51}$$

where $\overline{\tilde{\boldsymbol{\xi}}}^*$, $\overline{\tilde{\mathbf{n}}}$, $\overline{\tilde{\mathbf{G}}}_{\tilde{\xi}}^*$, $\overline{\tilde{\mathbf{W}}}^2$ and $\overline{\mathbf{W}}^2$ are the scaled optimal extended measurements (including moments and control inputs), scaled measurement noise, optimal gradients of $\overline{\tilde{\boldsymbol{\xi}}}$ with respect

to $\tilde{\mathbf{u}}$, scaled stacked covariance matrix and scaled covariance matrix, respectively. Note that repmat$(\mathbf{b}) \in \mathbb{R}^{(n_y+n_u)\times(n_y+n_u)}$ indicates that $\mathbf{b}$ is copied $n_y + n_u$ times by row and $\mathbf{W}^2 \cdot \text{repmat}(\mathbf{b})$ in (51) denotes the dot product of $\mathbf{W}^2$ and repmat$(\mathbf{b})$. Then substituting (47) (48) (49) and (29) into (36), it can be obtained that

$$\begin{aligned}
L &\approx (\tilde{\boldsymbol{\xi}}^* + \tilde{\mathbf{n}})^{\mathrm{T}}\tilde{\mathbf{H}}^{\mathrm{T}}(\tilde{\mathbf{H}}\tilde{\mathbf{G}}_{\tilde{\xi}}^*)^{-\mathrm{T}}\mathbf{J}_{\tilde{u}\tilde{u}}^*(\tilde{\mathbf{H}}\tilde{\mathbf{G}}_{\tilde{\xi}}^*)^{-1}\tilde{\mathbf{H}}(\tilde{\boldsymbol{\xi}}^* + \tilde{\mathbf{n}}) \\
&= (\tilde{\mathbf{B}}^{-1}(\overline{\tilde{\boldsymbol{\xi}}}^* + \overline{\tilde{\mathbf{n}}}))^{\mathrm{T}}\tilde{\mathbf{H}}^{\mathrm{T}}(\tilde{\mathbf{H}}\tilde{\mathbf{B}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{G}}_{\tilde{\xi}}^*)^{-\mathrm{T}}\mathbf{J}_{\tilde{u}\tilde{u}}^*(\tilde{\mathbf{H}}\tilde{\mathbf{B}}^{-1}\tilde{\mathbf{B}}\tilde{\mathbf{G}}_{\tilde{\xi}}^*)^{-1}\tilde{\mathbf{H}}(\tilde{\mathbf{B}}^{-1}(\overline{\tilde{\boldsymbol{\xi}}}^* + \overline{\tilde{\mathbf{n}}})) \\
&= (\overline{\tilde{\boldsymbol{\xi}}}^* + \overline{\tilde{\mathbf{n}}})^{\mathrm{T}}(\tilde{\mathbf{H}}\tilde{\mathbf{B}}^{-1})^{\mathrm{T}}(\tilde{\mathbf{H}}\tilde{\mathbf{B}}^{-1}\overline{\tilde{\mathbf{G}}}_{\tilde{\xi}}^*)^{-\mathrm{T}}\mathbf{J}_{\tilde{u}\tilde{u}}^*(\tilde{\mathbf{H}}\tilde{\mathbf{B}}^{-1}\overline{\tilde{\mathbf{G}}}_{\tilde{\xi}}^*)^{-1}(\tilde{\mathbf{H}}\tilde{\mathbf{B}}^{-1})(\overline{\tilde{\boldsymbol{\xi}}}^* + \overline{\tilde{\mathbf{n}}}) \\
&\xlongequal{\overline{\tilde{\mathbf{H}}}=\tilde{\mathbf{H}}\tilde{\mathbf{B}}^{-1}} (\overline{\tilde{\boldsymbol{\xi}}}^* + \overline{\tilde{\mathbf{n}}})^{\mathrm{T}}\overline{\tilde{\mathbf{H}}}^{\mathrm{T}}(\overline{\tilde{\mathbf{H}}\tilde{\mathbf{G}}}_{\tilde{\xi}}^*)^{-\mathrm{T}}\mathbf{J}_{\tilde{u}\tilde{u}}^*(\overline{\tilde{\mathbf{H}}\tilde{\mathbf{G}}}_{\tilde{\xi}}^*)^{-1}\overline{\tilde{\mathbf{H}}}(\overline{\tilde{\boldsymbol{\xi}}}^* + \overline{\tilde{\mathbf{n}}}) \\
&= \overline{L}
\end{aligned} \tag{52}$$

Thus it is indicated in (52) that scaling the measurements is equivalent to a linear transformation of $\tilde{\mathbf{H}}$, *i.e.* $\overline{\tilde{\mathbf{H}}} = \tilde{\mathbf{H}}\tilde{\mathbf{B}}^{-1}$, which does not influence the optimal value of the loss function. Therefore, $\overline{\tilde{\mathbf{H}}}$ can be adopted to achieve the same self-optimizing control performance as with $\tilde{\mathbf{H}}$, while $\overline{\tilde{\mathbf{H}}}$ has better numerical performance. Hence, $\overline{\tilde{\mathbf{H}}}$ should be obtained through minimization of the global average loss $\mathbb{E}(\overline{L})$. To achieve this, the global average loss can be approximated through Monte Carlo simulation by sampling disturbances within the predefined set. Then the optimal $\overline{\tilde{\mathbf{H}}}$ is obtained by solving the following structure-constrained SOC problem.

$$\begin{aligned}
\min_{\overline{\tilde{\mathbf{H}}}} \quad & \overline{L}_{\text{gav}} = \mathbb{E}(\overline{L}) = \int \boldsymbol{\rho}(\tilde{\mathbf{d}})(\overline{L}^d + \overline{L}^n)d\tilde{\mathbf{d}} \approx \frac{1}{N_d}\sum_{j=1}^{N_d}(\overline{L}_j^d + \overline{L}_j^n) \\
\text{s.t.} \quad & \overline{\tilde{\mathbf{H}}} \text{ is constant block diagonal in (23).}
\end{aligned} \tag{53}$$

where

$$\overline{L}^d = \frac{1}{2}\overline{\tilde{\boldsymbol{\xi}}}^{*\mathrm{T}}\overline{\tilde{\mathbf{H}}}^{\mathrm{T}}\mathbf{J}_{\tilde{c}\tilde{c}}^*\overline{\tilde{\mathbf{H}}\tilde{\boldsymbol{\xi}}}^*, \ \overline{L}^n = \frac{1}{2}\text{tr}(\overline{\tilde{\mathbf{W}}}^2\overline{\tilde{\mathbf{H}}}^{\mathrm{T}}\mathbf{J}_{\tilde{c}\tilde{c}}^*\overline{\tilde{\mathbf{H}}}) \tag{54}$$

Algorithm 1: Dynamic global SOC method for PDF-shaping problems (SOC-PDF)

Input: Disturbance sampling set: $\mathcal{D}$; Scaling matrix: $\mathbf{B}$; Scaled covariance matrix of measurement noise: $\overline{\mathbf{W}}^2$; Weighting factor: $\mathbf{Q}$.

Output: Optimal combination matrix: $\overline{\tilde{\mathbf{H}}}$.

foreach $\tilde{\mathbf{d}}_j$ *in* $\mathcal{D}$ **do**
  $\tilde{\mathbf{u}}_j^*$, γ_j^*, $J^* \leftarrow$ solve the optimization problem (11)-(16);
  Dimensionality reduction: $\mu_{n,j}^*$ $(n = 1, \cdots, n_y) \leftarrow \gamma_j^*$ based on (20) and then $\mathbf{y}_j^* \leftarrow \mu_{n,j}^*$ using (17);
  Obtain scaled stacked vectors $\overline{\tilde{\boldsymbol{\xi}}}^*$ and $\overline{\tilde{\mathbf{n}}}$ based on (24) (27) (47) and (18) (33) (35) (48), and $\overline{\tilde{\mathbf{G}}}_{\tilde{\xi}}^*$ and obtain $\mathbf{J}_{\tilde{u}\tilde{u}}^*$ through sensitivity analysis based on (30) (49) and (31), respectively;
  Compute $\overline{L}_j^d$ and $\overline{L}_j^n$ based on (54) and (29);
end
$\overline{\tilde{\mathbf{H}}} \leftarrow$ solve the structure-constrained SOC problem (53) numerically.

$\overline{L}^d$ and $\overline{L}^n$ in (54) are the scaled ones in (38). $\boldsymbol{\rho}(\tilde{\mathbf{d}})$ is the probability density function of $\tilde{\mathbf{d}}$. N_d is the number of total sampling points and the subscript $(\cdot)_j$ denotes the term associated with the jth sampled disturbance scenario. Therefore, the optimal scaled stacked combination matrix $\overline{\tilde{\mathbf{H}}}$ can be obtained by solving (53) numerically.

Unfortunately, such structure-constrained dynamic SOC problem is usually non-convex and has no closed-form solution in general.[59] It can be solved by some numerical optimization solvers such as sequential quadratic programming (SQP), interior-point method and etc.

In summary, the dynamic global self-optimizing control method for the PDF-shaping problems (SOC-PDF) is concluded in Algorithm 1.

3.4 Online implementation of SOC

After obtaining optimal $\overline{\tilde{\mathbf{H}}}$ offline based on Algorithm 1, the online control law can be obtained directly from the optimal CVs under the condition that $\mathbf{u}(k)$ is included in the extended measurement set $\boldsymbol{\xi}(k)$ and the measured CVs at time k should be maintained at

zero which can be written as

$$\overline{\mathbf{c}}_{\mathrm{m}}^{\mathrm{fb}}(k) = \overline{\mathbf{c}}_{\mathrm{s}} + \overline{\mathbf{H}}\overline{\boldsymbol{\xi}}_{\mathrm{m}}^{\mathrm{fb}}(k) = \overline{\mathbf{c}}_{\mathrm{s}} + \underbrace{\begin{bmatrix} \overline{\mathbf{H}}^{y} & \overline{\mathbf{H}}^{u} \end{bmatrix}}_{\overline{\mathbf{H}}} \underbrace{\begin{bmatrix} \overline{\mathbf{y}}_{\mathrm{m}}^{\mathrm{fb}}(k) \\ \overline{\mathbf{u}}_{\mathrm{m}}^{\mathrm{fb}}(k) \end{bmatrix}}_{\overline{\boldsymbol{\xi}}_{\mathrm{m}}^{\mathrm{fb}}(k)} = \mathbf{0}_{n_u \times 1} \tag{55}$$

where $\overline{(\cdot)}$ denotes the scaled value of the term. $\overline{\mathbf{H}}^{y} \in \mathbb{R}^{n_u \times n_y}$ and $\overline{\mathbf{H}}^{u} \in \mathbb{R}^{n_u \times n_u}$ are the sub-matrices corresponding to $\overline{\mathbf{y}}_{\mathrm{m}}^{\mathrm{fb}}(k)$ and $\overline{\mathbf{u}}_{\mathrm{m}}^{\mathrm{fb}}(k)$ in $\overline{\mathbf{H}} \in \mathbb{R}^{(n_y+n_u)\times(n_y+n_u)}$. Then the closed-loop control inputs at time k can be calculated by

$$\mathbf{u}_{\mathrm{m}}^{\mathrm{fb}}(k) = -(\overline{\mathbf{H}}^{u})^{-1}\left(\overline{\mathbf{c}}_{\mathrm{s}} + \overline{\mathbf{H}}^{y}\overline{\mathbf{y}}_{\mathrm{m}}^{\mathrm{fb}}(k)\right) \tag{56}$$

Note that $\overline{\mathbf{H}}^{u}$ is invertible and an extra constraint can be enforced as $\overline{\mathbf{H}}^{u} = \mathbf{I}_{n_u \times n_u}$ for simplicity (where $\mathbf{I}$ is the identity matrix of $n_u \times n_u$). Furthermore, since $\overline{\mathbf{H}}^{u}$, $\overline{\mathbf{H}}^{y}$ and $\overline{\mathbf{c}}_{\mathrm{s}}$ are elements in $\overline{\tilde{\mathbf{H}}}$ and have been obtained offline based on Algorithm 1, $\mathbf{u}_{\mathrm{m}}^{\mathrm{fb}}(k)$ can be easily obtained in real-time once $\overline{\mathbf{y}}_{\mathrm{m}}^{\mathrm{fb}}(k)$ is measured online.

4 Case study

4.1 Model description

A semi-batch styrene polymerization process model[44] is studied as shown in Figure 1. Styrene is the monomer for polymerization and azobisisobutyronitrile is the initiator. The total feed flow rate F is the sum of the monomer flow rate F_{m} and the initiator flow rate F_{i}. These two streams are injected into a continuous stirring tank reactor (CSTR) of 3.927 L with a certain ratio of $u = \frac{F_{\mathrm{m}}}{F_{\mathrm{m}}+F_{\mathrm{i}}} = \frac{F_{\mathrm{m}}}{F} \in [0.3, 0.7]$, which is assumed as control input. The energy is provided by the heated oil in the CSTR jacket whose temperature is controlled by a temperature control loop. For simplicity, the reaction temperature is assumed constant at 353 K. The molecular weight distribution is supposed to be measurable or mathematically

calculated online.

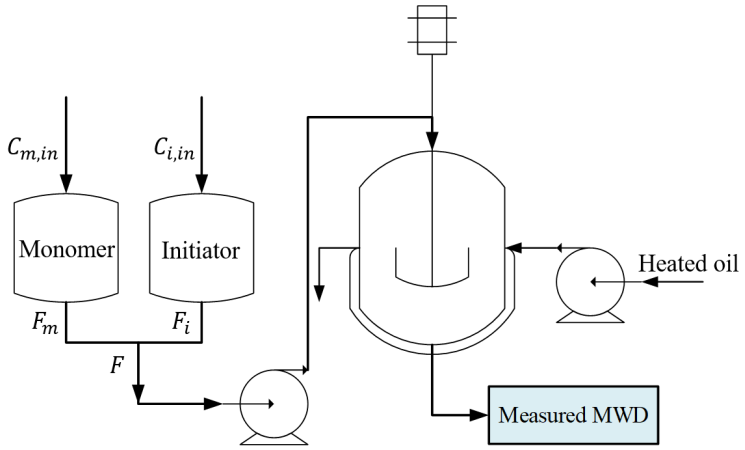

Figure 1: Semi-batch styrene polymerization model in a continuous stirring tank reactor (CSTR)

The overall ordinary differential equations[44] are given as follows, and the nominal values of process variables are presented in Table 3.

$$\frac{dC_{\mathrm{i}}}{dt} = \frac{F}{V}\left(\left(1-u\right)C_{\mathrm{i,in}} - C_{\mathrm{i}}\right) - K_{\mathrm{d}}C_{\mathrm{i}} \tag{57}$$

$$\frac{dC_{\mathrm{m}}}{dt} = \frac{F}{V}\left(uC_{\mathrm{m,in}} - C_{\mathrm{m}}\right) - 2K_{\mathrm{i}}C_{\mathrm{i}} - \left(K_{\mathrm{p}} + K_{\mathrm{trm}}\right)C_{\mathrm{m}}\psi_0 \tag{58}$$

$$\frac{d\psi_0}{dt} = -\frac{F}{V}\psi_0 + 2K_{\mathrm{i}}C_{\mathrm{i}} - K_{\mathrm{t}}\psi_0^2 \tag{59}$$

$$\frac{d\psi_1}{dt} = -\frac{F}{V}\psi_1 + 2K_{\mathrm{i}}C_{\mathrm{i}} + K_{\mathrm{p}}\psi_0 C_{\mathrm{m}} - K_{\mathrm{t}}\psi_0\psi_1 + K_{\mathrm{trm}}C_{\mathrm{m}}\left(\psi_0 - \psi_1\right) \tag{60}$$

$$\frac{d\psi_2}{dt} = -\frac{F}{V}\psi_2 + 2K_{\mathrm{i}}C_{\mathrm{i}} + K_{\mathrm{p}}C_{\mathrm{m}}\left(2\psi_1 + \psi_0\right) - K_{\mathrm{t}}\psi_0\psi_2 + K_{\mathrm{trm}}C_{\mathrm{m}}\left(\psi_0 - \psi_2\right) \tag{61}$$

$$\frac{dZ_0}{dt} = -\frac{F}{V}Z_0 + K_{\mathrm{trm}}C_{\mathrm{m}}\psi_0 + \frac{K_{\mathrm{t}}}{2}\psi_0^2 \tag{62}$$

$$\frac{dZ_1}{dt} = -\frac{F}{V}Z_1 + K_{\mathrm{trm}}C_{\mathrm{m}}\psi_1 + K_{\mathrm{t}}\psi_0\psi_1 \tag{63}$$

$$\frac{dZ_2}{dt} = -\frac{F}{V}Z_2 + K_{\mathrm{trm}}C_{\mathrm{m}}\psi_2 + K_{\mathrm{t}}\psi_0\psi_2 + K_{\mathrm{t}}\psi_1^2 \tag{64}$$

where $K_j = k_j e^{-\frac{E_j}{RT}}$ $(j = \{\mathrm{d,i,p,trm,t}\})$ denotes the reaction constant for decomposition, initiation, propagation, chain transfer or termination respectively, and $K_{\mathrm{i}} = fK_{\mathrm{d}}$. C_{i} and C_{m} are concentration of initiator and monomer in the CSTR. ψ_0, ψ_1, ψ_2 and Z_0, Z_1, Z_2 are the leading moments of living radicals and dead polymer respectively. The initial

Table 3: Process Variables and Their Nominal Values

Variable	Description	Value	Unit
f	Efficiency of the initiator	0.6	-
k_{d}	Rate coefficient for decomposition	9.480×10^{16}	min^{-1}
k_{p}	Rate coefficient for propagation	6.306×10^{8}	$\mathrm{L}/(\mathrm{mol}\cdot\mathrm{min})$
k_{trm}	Rate coefficient for chain transfer to monomer	1.386×10^{8}	$\mathrm{L}/(\mathrm{mol}\cdot\mathrm{min})$
k_{t}	Rate coefficient for termination	3.765×10^{10}	$\mathrm{L}/(\mathrm{mol}\cdot\mathrm{min})$
E_{d}	Activation energy for decomposition	30798.5	J/mol
E_{p}	Activation energy for propagation	7067.8	J/mol
E_{trm}	Activation energy for chain transfer to monomer	12671.1	J/mol
E_{t}	Activation energy for termination	1680	J/mol
V	Volume of the reacting mixture	3.927	L
F	Total feed flow rate	0.0238	L/min
T	Reaction temperature	353	K
$C_{\mathrm{i,in}}$	Inlet initiator concentration	0.0106	mol/L
$C_{\mathrm{m,in}}$	Inlet monomer concentration	4.81	mol/L
R	Universal gas constant	1.987	$\mathrm{cal}/(\mathrm{mol}\cdot\mathrm{K})$

values of all the states are $\mathbf{x}_0 = [C_{\mathrm{i}}(0), C_{\mathrm{m}}(0), \psi_0(0), \psi_1(0), \psi_2(0), Z_0(0), Z_1(0), Z_2(0)]^{\mathrm{T}} = [0.002, 2.262, 0, 0, 0.0079, 0.0014, 0.624, 425.805]^{\mathrm{T}}$

Then to obtain molecular weight distribution, the following well-known normalized Schultz–Zimm distribution function[60] is utilized to describe MWD of polystyrene.

$$\gamma(i) = \frac{h^h i^{h-1} e^{\frac{-hi}{M_{\mathrm{n}}}}}{M_{\mathrm{n}}^h \Gamma(h)} \tag{65}$$

where

$$\Gamma(h) = \int_0^\infty i^{h-1} e^{-i} di \tag{66}$$

$$h = \frac{Z_1^2}{Z_0 Z_2 - Z_1^2} \tag{67}$$

$$M_{\mathrm{n}} = \frac{Z_1}{Z_0} \tag{68}$$

$i \geq 0$ is the chain length while the maximum chain length is set as 2000 in this case. h denotes the distribution breadth, $\Gamma(h)$ is the gamma function and M_{n} is the number average chain length. When $h = 1$, the Schultz–Zimm distribution reduces to another commonly used distribution for MWD, the exponential Flory distribution.

Remark 5 *The leading moments of dead polymer Z_0, Z_1, Z_2 are simply used to calculate the molecular weight distribution γ based on the normalized Schultz–Zimm distribution* (65)*. In fact, γ is assumed to be measurable and it is the moments of γ that are applied in self-optimizing control.*

The objective function of SOC for the PDF-shaping problem is defined as

$$J = \sum_{i=1}^{2000} [\gamma(i, N) - \gamma_{\mathrm{t}}(i)]^2 + \sum_{k=0}^{N-1} u(k)^{\mathrm{T}} Q u(k) \tag{69}$$

with disturbances and moments up to order of 7 (not all of them have to be used in this case) as

$$\mathbf{d} = [C_{\mathrm{i,in}}, C_{\mathrm{m,in}}, k_{\mathrm{p}}]^{\mathrm{T}} \tag{70}$$

$$\overline{\boldsymbol{\xi}}(k) = [\overline{\mu}_1(k), \overline{\mu}_2(k), \cdots, \overline{\mu}_7(k), u(k)]^{\mathrm{T}} \tag{71}$$

where $\overline{\mu}_1(k), \cdots, \overline{\mu}_7(k)$ are the 1st to 7th scaled moments of the MWD $\gamma(i, k)$ at time k, *i.e* $\overline{\boldsymbol{\xi}}(k) = \boldsymbol{\xi}(k) \cdot \mathbf{b}^{\mathrm{T}}$, where $\mathbf{b} = [10^{-2}, 10^{-5}, 10^{-8}, 10^{-11}, 10^{-14}, 10^{-17}, 10^{-20}, 1]$. μ_0 is constant 1 since the MWD is a PDF hence not included in $\overline{\boldsymbol{\xi}}(k)$. The total batch time 400 min is divided into $N = 20$ intervals. The penalty weighting factor is set as $Q = 0$. The disturbances are assumed unchanged within the batch and varied batch to batch. The variation ranges for disturbances are defined as ± 20 % of their nominal values. The measurement noise on $\gamma(i, k)$ and $u(k)$ are both normal distributions with the same zero mean while different variances $\sigma_\gamma^2 = 10^{-10}$ and $\sigma_u^2 = 10^{-4}$ respectively. Therefore, the scaled covariance matrix can be

obtained as

$$\overline{\mathbf{W}}^2 = \begin{bmatrix} 2.669\times10^{-5} & 4.004\times10^{-5} & 6.408\times10^{-5} & 1.068\times10^{-4} & 1.832\times10^{-4} & 3.206\times10^{-4} & 5.702\times10^{-4} & 0 \\ 4.004\times10^{-5} & 6.408\times10^{-5} & 1.068\times10^{-4} & 1.832\times10^{-4} & 3.206\times10^{-4} & 5.702\times10^{-4} & 0.001 & 0 \\ 6.408\times10^{-5} & 1.068\times10^{-4} & 1.832\times10^{-4} & 3.206\times10^{-4} & 5.702\times10^{-4} & 0.001 & 0.002 & 0 \\ 1.068\times10^{-4} & 1.832\times10^{-4} & 3.206\times10^{-4} & 5.702\times10^{-4} & 0.001 & 0.002 & 0.003 & 0 \\ 1.832\times10^{-4} & 3.206\times10^{-4} & 5.702\times10^{-4} & 0.001 & 0.002 & 0.003 & 0.006 & 0 \\ 3.206\times10^{-4} & 5.702\times10^{-4} & 0.001 & 0.002 & 0.003 & 0.006 & 0.012 & 0 \\ 5.702\times10^{-4} & 0.001 & 0.002 & 0.003 & 0.006 & 0.012 & 0.022 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 10^{-4} \end{bmatrix} \tag{72}$$

Remark 6 *Since the order of magnitude of moments μ_n becomes larger and larger as the order increases, the raw moments are scaled to an identical order of magnitude to avoid numerical issues.*

The whole disturbance space is randomly sampled by $N_d = 500$ Gaussian distributed scenarios. Through offline minimization of J subjected to the dynamic model equations and input constraint, optimal data can be obtained and then the optimal $\overline{\tilde{\mathbf{H}}}$ can be acquired following Algorithm 1. All the computations were executed on a desktop computer with an Intel Core i5-9400 CPU at 2.90 GHz, 8 GB RAM at 2.90GHz under 64-bit Windows 10 operating system and the algorithm is implemented on Matlab 2021b.

4.2 Results and Discussions

To test the proposed SOC-PDF method, another 100 normally distributed disturbance scenarios are generated. The closed-loop global average loss of the objective function for SOC PDF-shaping problem is applied to quantify the control performance.

To find how many moments should be included in self-optimizing control, different number of moments are selected to find optimal $\overline{\tilde{\mathbf{H}}}$ and then closed-loop average loss is evaluated

through the polymerization model as depicted in Figure 2. It is clear that the closed-loop global average loss decreases with the number of moments ranging from 0 to 7. The performance of not using any moment is not satisfactory, which gives an average loss of 1.014×10^{-5}. This may be due to the fact that no moment information is used, and the control input u remains constant after online calculation $u(k) = 0.5779$, thus unable to resist the influence of disturbances. It shows that choosing one moment ($\overline{\mu}_1$) is good enough without need of more higher moments. It is probably because that the MWD is unimodal, and the initial MWD is similar to the target. Hence, controlling only the first moment can make the actual MWD close to the target distribution. The resulting optimal self-optimizing CVs at each time step are $c(k) = -2.3851 + [0.4309, 1]\,[\overline{\mu}_1(k), u(k)]^{\mathrm{T}}$ which should be maintained at zero, and therefore the online control input is directly calculated by $u(k) = 2.3851 - 0.4309\overline{\mu}_1(k)$.

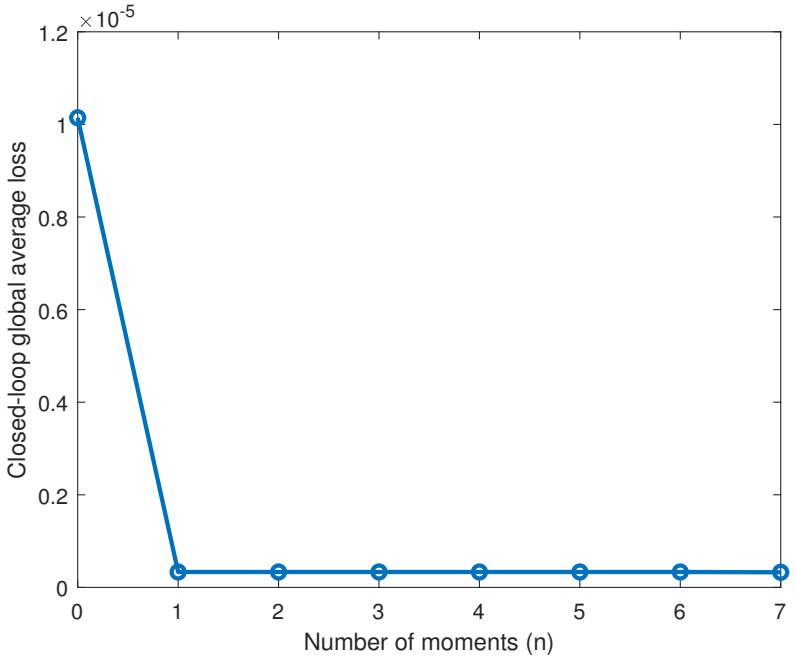

Figure 2: Closed-loop global average loss obtained through nonlinear model evaluation using 0 to 7 moments.

When $\mathbf{d}$ changes by 20 % (which is the maximum range) from the nominal point (e.g. $\mathbf{d} = [0.0127, 5.7720, 7.5672 \times 10^8]^{\mathrm{T}}$), the trajectories of initiator concentration C_{i}, monomer concentration C_{m}, objective function J and manipulated variable u using the first order

moment ($n = 1$) are depicted in Figure 3. In addition, the final MWD curves after self-optimizing control using no moment (u keeps at constant 0.5779) and using the first order moment ($n = 1$) are compared with the target MWD as shown in Figure 4. It is indicated that under large disturbances, the actual MWD is closer to the target distribution by using the first order moment ($J = 3.625 \times 10^{-7}$) than by using no moments. The trajectory of the actual MWD after self-optimizing control using the first order moment within the batch is shown in Figure 5.

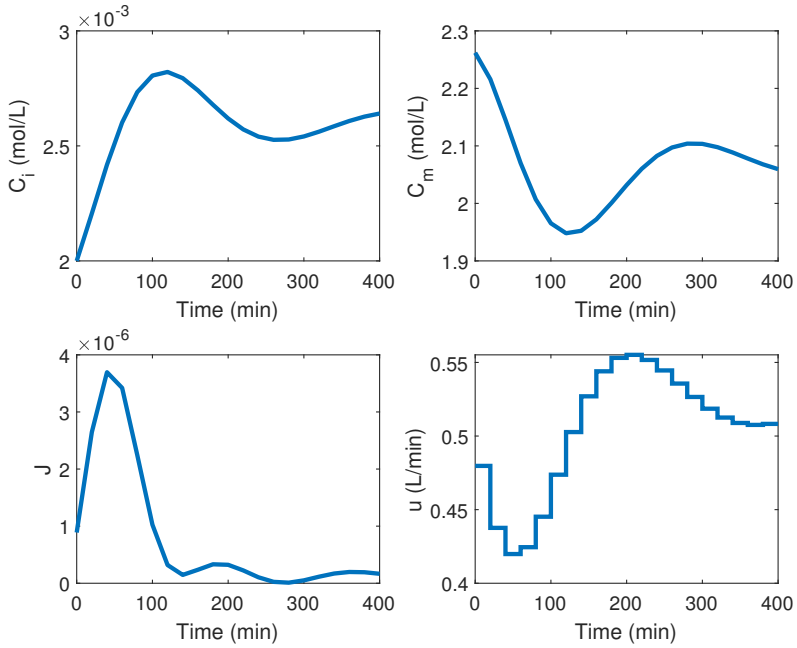

Figure 3: Performance of SOC-PDF when $\mathbf{d}$ changes by 20 % from the nominal point.

For comparison, the stochastic distribution control (SDC) method (standard output PDF control in[14]) is also implemented on the styrene polymerization process for the same 100 disturbance scenarios. The results are summarized in Table 4 and it can be seen that the proposed SOC-PDF method (using only the first order moment $\overline{\mu}_1$) outperforms the SDC method, whose average, maximum and standard deviation of performance index is about one to two orders of magnitude smaller than the SDC method. Because the SDC method adopts the optimal control strategy and does not consider the influence of disturbances and noise, it has poor performance in the global variation range of disturbances. In the contrast,

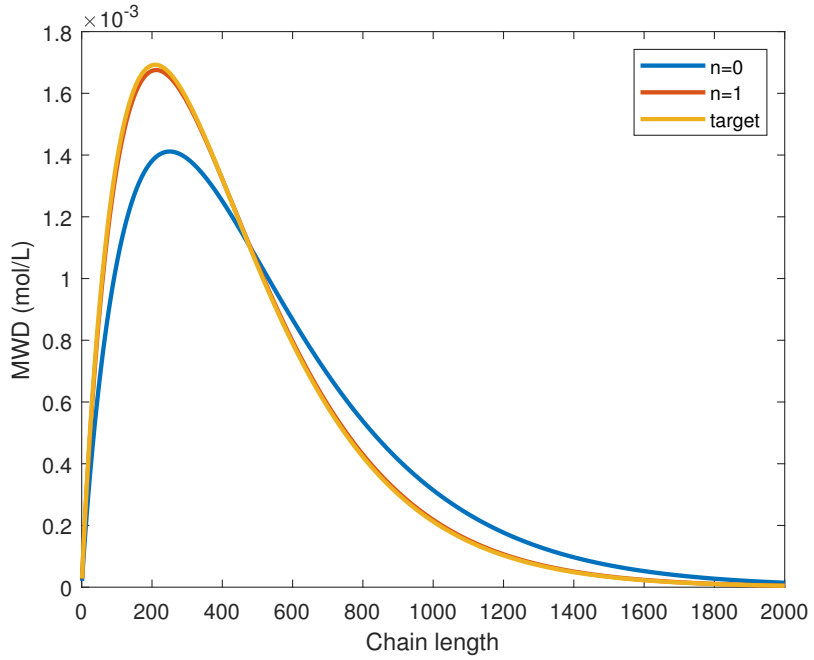

Figure 4: MWD at final time step after SOC-PDF when $\mathbf{d}$ changes by 20 % from the nominal point.

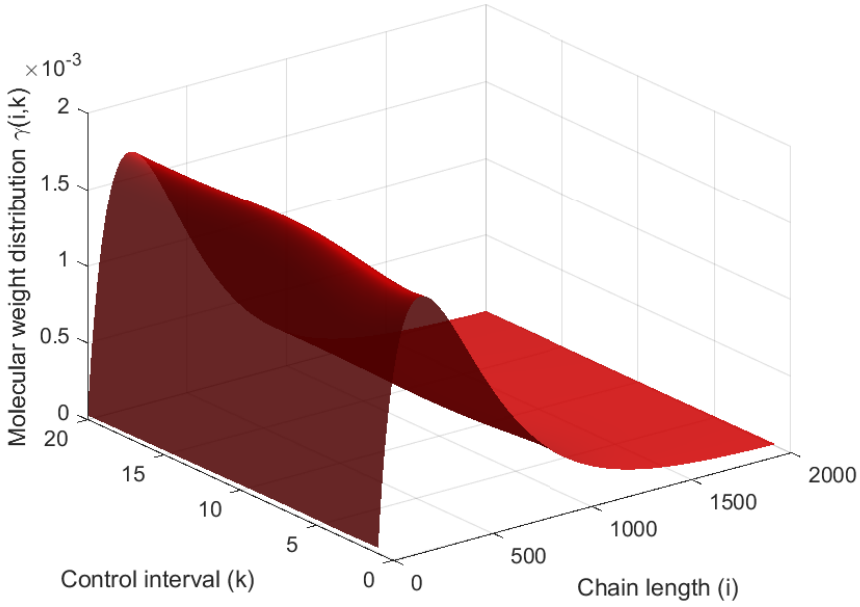

Figure 5: Trajectory of the actual MWD after SOC-PDF using the first order moment when $\mathbf{d}$ changes by 20 % from the nominal point.

SOC-PDF method designs self-optimizing CVs offline based on information of moments and the PDF-shaping control can be achieved through a very simple online control law even in the presence of disturbances and noise.

Table 4: Closed-loop global loss evaluated from the original nonlinear dynamic model

	Average	Maximum	Standard deviation
SOC-PDF (this paper)	7.4245×10^{-7}	4.1918×10^{-6}	7.6368×10^{-7}
SDC	2.1025×10^{-6}	1.2616×10^{-4}	1.2654×10^{-5}

5 Conclusion

This paper for the first time introduces self-optimizing control into the field of PDF-shaping problems and establishes its basic framework. Since it is generally hard to control PDF directly, the finite moments of the PDF are involved as available measurement variables to obtain the globally optimal controlled variables offline. When they are controlled constant online, the PDF is close to the target. As demonstrated in the styrene polymerization case, the proposed dynamic global self-optimizing control method achieves better performance than the existing stochastic distribution control method.

However, this work is limited to a simulated environment at present and has not yet been extended to a real-world polymerization process. For industrial application, several critical considerations are listed as follows:

- Key process variables, including reactor temperature, pressure, liquid level, feed flow rates and etc., should be measured online. The real-time data can be acquired using the OPC (OLE for Process Control) protocol from sensors.

- In cases where online measurement of the molecular weight distribution (MWD) is not feasible, soft sensors[36] can serve to estimate the dynamic MWD online based on the equation-oriented (EO) approach.[37]

- The collected data (under various operational conditions) is used to build and calibrate the dynamic model of the polymerization process. Based on the developed model and soft sensors to estimate dynamic MWD, the optimization problem of minimizing the global average loss is solved offline to determine the optimal controlled variables (CVs). Then the SOC methodology involves maintaining the CVs at zero setpoints during online operation, thereby achieving self-optimizing control of the MWD.

Recently, our research group has made significant progress in setting up a pilot polymerization plant. The pilot plant is fully equipped with data communication capabilities, laying the groundwork for implementing and testing the SOC strategy in a real-world setting in the future.

Acknowledgement

The authors gratefully acknowledge the financial support from the Institute of Zhejiang University-Quzhou Science and Technology Project (IZQ2022KYZX12 and IZQ2022KJ3001).

References

(1) Gentekos, D. T.; Sifri, R. J.; Fors, B. P. Controlling polymer properties through the shape of the molecular-weight distribution. *Nature Reviews Materials* **2019**, *4*, 761–774.

(2) Kiparissides, C. Challenges in particulate polymerization reactor modeling and optimization: A population balance perspective. *Journal of Process Control* **2006**, *16*, 205–224.

(3) Crowley, T. J.; Choi, K. Y. Experimental studies on optimal molecular weight distribution control in a batch-free radical polymerization process. *Chemical Engineering Science* **1998**, *53*, 2769–2790.

(4) Takamatsu, T.; Shioya, S.; Okada, Y. Molecular weight distribution control in a batch polymerization reactor. *Industrial & engineering chemistry research* **1988**, *27*, 93–99.

(5) Chang, J.-S.; Liao, P.-H. Molecular weight control of a batch polymerization reactor: experimental study. *Industrial & engineering chemistry research* **1999**, *38*, 144–153.

(6) Kiparissides, C.; Seferlis, P.; Mourikas, G.; Morris, A. J. Online optimizing control of molecular weight properties in batch free-radical polymerization reactors. *Industrial & engineering chemistry research* **2002**, *41*, 6120–6131.

(7) Alhamad, B.; Romagnoli, J.; Gomes, V. Advanced modelling and optimal operating strategy in emulsion copolymerization: Application to styrene/MMA system. *Chemical Engineering Science* **2005**, *60*, 2795–2813.

(8) Alhamad, B.; Romagnoli, J. A.; Gomes, V. G. On-line multi-variable predictive control of molar mass and particle size distributions in free-radical emulsion copolymerization. *Chemical Engineering Science* **2005**, *60*, 6596–6606.

(9) Vicente, M.; BenAmor, S.; Gugliotta, L. M.; Leiza, J. R.; Asua, J. M. Control of molecular weight distribution in emulsion polymerization using on-line reaction calorimetry. *Industrial & engineering chemistry research* **2001**, *40*, 218–227.

(10) Vicente, M.; Sayer, C.; Leiza, J.; Arzamendi, G.; Lima, E.; Pinto, J.; Asua, J. Dynamic optimization of non-linear emulsion copolymerization systems: Open-loop control of composition and molecular weight distribution. *Chemical Engineering Journal* **2002**, *85*, 339–349.

(11) Yoo, K.-Y.; Jeong, B.-G.; Rhee, H.-K. Molecular weight distribution control in a batch polymerization reactor using the on-line two-step method. *Industrial & engineering chemistry research* **1999**, *38*, 4805–4814.

(12) Koolivand, A.; Shahrokhi, M.; Farahzadi, H. Optimal control of molecular weight and particle size distributions in a batch suspension polymerization reactor. *Iranian Polymer Journal* **2019**, *28*, 735–745.

(13) Wang, H. *Bounded dynamic stochastic systems: modelling and control*; UK: Springer-Verlag, 2000.

(14) Zhang, J.; Yue, H.; Zhou, J. Predictive PDF control in shaping of molecular weight distribution based on a new modeling algorithm. *Journal of process control* **2015**, *30*, 80–89.

(15) Ren, M.; Zhang, Q.; Zhang, J. An introductory survey of probability density function control. *Systems Science & Control Engineering* **2019**, *7*, 158–170.

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

(17) Jäschke, J.; Cao, Y.; Kariwala, V. Self-optimizing control - A survey. *Annual Reviews in Control* **2017**, *43*, 199–223.

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

(19) Alstad, V.; Skogestad, S. Null Space Method for Selecting Optimal Measurement Combinations as Controlled Variables. *Industrial & Engineering Chemistry Research* **2007**, *46*, 846–853.

(20) Halvorsen, I. J.; Skogestad, S.; Morud, J. C.; Alstad, V. Optimal Selection of Controlled Variables. *Industrial & Engineering Chemistry Research* **2003**, *42*, 3273–3284.

(21) Ye, L.; Cao, Y.; Li, Y.; Song, Z. A Data-driven Approach for Selecting Controlled Variables. *IFAC Proceedings Volumes* **2012**, *45*, 904–909, 8th IFAC Symposium on Advanced Control of Chemical Processes.

(22) Ye, L.; Cao, Y.; Skogestad, S. Global Self-Optimizing Control for Uncertain Constrained Process Systems. *IFAC-PapersOnLine* **2017**, *50*, 4672–4677, 20th IFAC World Congress.

(23) Cao, Y.; Ye, L. Enhanced global self-optimizing control. *Computer Aided Chemical Engineering* **2017**, *40*, 1651–1656.

(24) Grema, A. S.; Cao, Y.; Grema, M. B. Data-driven self-optimizing control: constrained optimization problem. *Ife Journal of Science* **2018**, *20*, 273–278.

(25) Su, H.; Zhou, C.; Cao, Y.; Yang, S. Machine learning for constrained self-optimizing control. *Computer Aided Chemical Engineering* **2021**, *50*, 1203–1208.

(26) Su, H.; Zhou, C.; Cao, Y.; Yang, S.-H.; Ji, Z. An intelligent approach of controlled variable selection for constrained process self-optimizing control. *Systems Science & Control Engineering* **2022**, *10*, 65–72.

(27) Su, H.; Cao, Y.; Yang, S.-H.; Ye, L. Lagrange-Based Global Self-Optimizing Control for Constraint Activeness Varying Processes. *IEEE Access* **2023**, *11*, 44600–44613.

(28) Ye, L.; Cao, Y.; Yuan, X. Global Approximation of Self-Optimizing Controlled Variables with Average Loss Minimization. *Industrial & Engineering Chemistry Research* **2015**, *54*, 12040–12053.

(29) Ye, L.; Skogestad, S. Dynamic self-optimizing control for unconstrained batch processes. *Computers & Chemical Engineering* **2018**, *117*, 451–468.

(30) Ye, L.; Cao, Y.; Li, Y.; Song, Z. Approximating Necessary Conditions of Optimality as Controlled Variables. *Industrial & Engineering Chemistry Research* **2013**, *52*, 798–808.

(31) Mnatsakanov, R. M. Hausdorff moment problem: Reconstruction of probability density functions. *Statistics & probability letters* **2008**, *78*, 1869–1877.

(32) Hai-yan, W.; Yu, C.; Jing, W. Control of Molecular Weight Distribution Based on Finite Order Moments. *Control Engineering of China* **2019**, *26*, 2171.

(33) Stacy, C. J. Molecular weight distribution of polyphenylene sulfide by high temperature gel permeation chromatography. *Journal of applied polymer science* **1986**, *32*, 3959–3969.

(34) Kosky, P.; McDonald, R.; Guggenheim, E. Determination of end-group concentrations and molecular weight of poly (butylene terephthalate) by solid-state fourier transform infrared spectroscopy. *Polymer Engineering & Science* **1985**, *25*, 389–394.

(35) Matson, J. B.; Steele, A. Q.; Mase, J. D.; Schulz, M. D. Polymer characterization by size-exclusion chromatography with multi-angle light scattering (SEC-MALS): a tutorial review. *Polymer Chemistry* **2024**, *15*, 127–142.

(36) Kadlec, P.; Gabrys, B.; Strandt, S. Data-driven soft sensors in the process industry. *Computers & chemical engineering* **2009**, *33*, 795–814.

(37) Kang, J.; Shao, Z.; Chen, X.; Gu, X.; Feng, L. Fast and reliable computational strategy for developing a rigorous model-driven soft sensor of dynamic molecular weight distribution. *Journal of Process Control* **2017**, *56*, 79–99.

(38) Soroush, M.; Kravaris, C. Nonlinear control of a batch polymerization reactor: an experimental study. *AIChE journal* **1992**, *38*, 1429–1448.

(39) Crowley, T. J.; Choi, K. Y. Calculation of molecular weight distribution from molecular weight moments in free radical polymerization. *Industrial & engineering chemistry research* **1997**, *36*, 1419–1423.

(40) Storti, G.; Polotti, G.; Cociani, M.; Morbidelli, M. Molecular weight distribution in emulsion polymerization. I. the homopolymer case. *Journal of Polymer Science Part A: Polymer Chemistry* **1992**, *30*, 731–750.

(41) Storti, G.; Polotti, G.; Canu, P.; Morbidelli, M. Molecular weight distribution in emulsion polymerization. II. The copolymer case. *Journal of Polymer Science Part A: Polymer Chemistry* **1992**, *30*, 751–777.

(42) Christov, L.; Georgiev, G. An algorithm for determination of the copolymer molecular weight distribution by Markov chain simulation. *Macromolecular theory and simulations* **1995**, *4*, 177–193.

(43) Yue, H.; Cao, L. Reaction extent modeling of a butadiene polymerization process. *IFAC Proceedings Volumes* **2000**, *33*, 1007–1012.

(44) Yue, H.; Zhang, J.; Wang, H.; Cao, L. Shaping of molecular weight distribution using B-spline based predictive probability density function control. *Proceedings of the 2004 American control conference* **2004**, *4*, 3587–3592.

(45) Salazar, A.; Gugliotta, L. M.; Vega, J. R.; Meira, G. R. Molecular weight control in a starved emulsion polymerization of styrene. *Industrial & engineering chemistry research* **1998**, *37*, 3582–3591.

(46) Nele, M.; Pinto, J. C. Molecular-Weight Multimodality of Multiple Flory Distributions. *Macromolecular theory and simulations* **2002**, *11*, 293–307.

(47) Dodson, C.; Wang, H. Iterative approximation of statistical distributions and relation to information geometry. *Statistical inference for stochastic processes* **2001**, *4*, 307–318.

(48) Jones, D.; Wang, H. A new non-linear optimal control strategy for paper formation. *Measurement and Control* **1999**, *32*, 241–245.

(49) Sun, X.; Yue, H.; Wang, H. Modelling and control of the flame temperature distribution using probability density function shaping. *Transactions of the Institute of Measurement and Control* **2006**, *28*, 401–428.

(50) Campbell, G. M.; Bunn, P.; Webb, C.; Hook, S. On predicting roller milling performance: Part II. The breakage function. *Powder Technology* **2001**, *115*, 243–255.

(51) Sun, X.; Ding, J.; Chai, T.; Wang, H. In *Recent Advances in Intelligent Control Systems*; Yu, W., Ed.; Springer London: London, 2009; pp 125–148.

(52) Wang, H.; Patil, S. V.; Aziz, H. A.; Young, S. Modeling and control using stochastic distribution control theory for intersection traffic flow. *IEEE Transactions on Intelligent Transportation Systems* **2020**, *23*, 1885–1898.

(53) Jin, C.; Wang, Y.; Sun, F.; Liu, T.; Cao, Y. Augmented Quadrature Method of Moments for Solving Population Balance Equations of Industrial Crystallization Processes. *2019 Chinese Automation Congress (CAC)* **2019**, 5724–5729.

(54) Hausdorff, F. Summationsmethoden und momentfolgen. I. *Mathematische Zeitschrift* **1921**, *9*, 74–109.

(55) Athanassoulis, G. A.; Gavriliadis, P. The truncated Hausdorff moment problem solved by using kernel density functions. *Probabilistic Engineering Mechanics* **2002**, *17*, 273–291.

(56) Helton, J. W.; Nie, J. A semidefinite approach for truncated K-moment problems. *Foundations of Computational Mathematics* **2012**, *12*, 851–881.

(57) Huang, L.; Nie, J.; Yuan, Y.-X. Generalized truncated moment problems with unbounded sets. *Journal of Scientific Computing* **2023**, *95*, 15.

(58) Zhou, C.; Su, H.; Tang, X.; Cao, Y.; Yang, S.-h. Global self-optimizing control of batch processes. *Journal of Process Control* **2024**, *135*, 103163.

(59) Yelchuru, R.; Skogestad, S. Convex formulations for optimal selection of controlled variables and measurements using mixed integer quadratic programming. *Journal of Process Control* **2012**, *22*, 995–1007.

(60) Angerman, H. J. The phase behavior of polydisperse multiblock copolymer melts:(a theoretical study). Ph.D. thesis, University of Groningen, 1998.

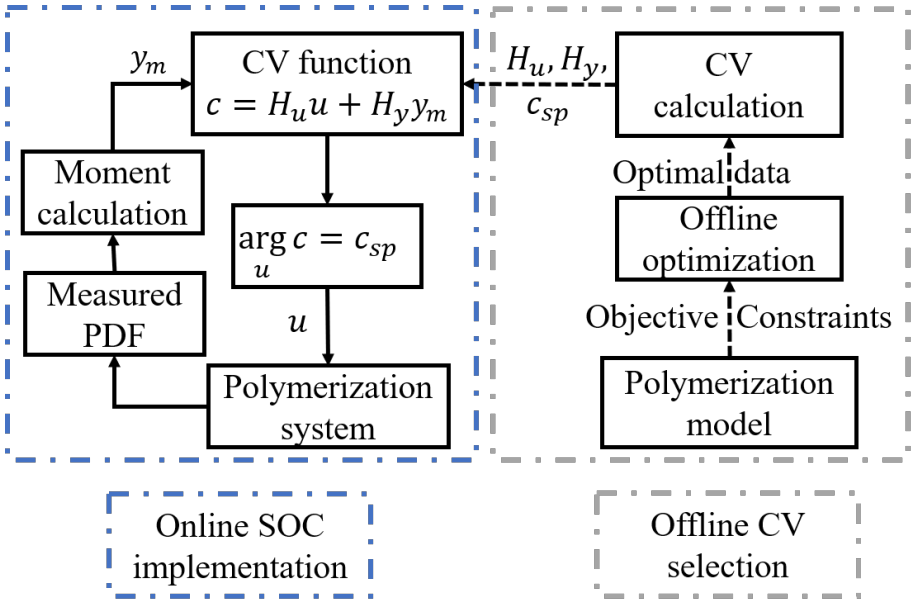

Figure 6: Table of Contents (TOC) graphic.