Preprints of the
9th International Symposium on Advanced Control of Chemical Processes
The International Federation of Automatic Control
June 7-10, 2015, Whistler, British Columbia, Canada

Process Parameter Optimization based on LW-PLS in Pharmaceutical Granulation Process ⋆

Ryosuke Yoshizaki * **Manabu Kano** * **Shuichi Tanabe** **
Takuya Miyano **

* *Department of Systems Science, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto-shi, Kyoto 606-8501, Japan, (e-mail: manabu@human.sys.i.kyoto-u.ac.jp).*
** *Formulation Technology Research Laboratories, Daiichi Sankyo Co., Ltd., Hiratsuka 254-0014, Japan*

Abstract: A process optimization method based on partial least squares (PLS) has been used in pharmaceutical processes. However, its applicability and performance are limited because PLS cannot cope with nonlinearity and changes in process characteristics. In this research, a new process optimization method based on locally weighted PLS (LW-PLS) is proposed. To solve a nonlinear optimization problem based on LW-PLS, in which any global model is not constructed, self-adaptive differential evolution (jDE) is adopted. The validity of the proposed method is demonstrated through a numerical example and an industrial case study of a pharmaceutical granulation process.

Keywords: Global optimization, Modeling, Nonlinear models, Adaptive algorithms, Locally weighted partial least squares, Quality by design, Self-adaptive differential evolution, Pharmaceutical granulation process.

1. INTRODUCTION

Quality by Design (QbD) has become important in the pharmaceutical industry since a report and guidelines were issued by FDA and ICH (Lawrence (2008)). The guidelines (US FDA et al. (2006)) state that "The aim of pharmaceutical development is to design a product quality and its manufacturing process to consistently deliver the intended performance of the product" and also "Quality cannot be tested into products, i.e., quality should be built in by design". For this purpose, it is crucial to understand a relationship between quality attributes and input variables such as material attributes and process parameters. Furthermore, an optimization method is required to derive process parameters that make final quality attributes reach desired values.

A process optimization method based on partial least squares (PLS) has been used in pharmaceutical processes (Muteki et al. (2011)). In this method, a PLS model was built to relate final tablet attributes with raw material attributes and process parameters. Then the process parameters were optimized by using sequential quadratic programming (SQP). The result was demonstrated that the PLS-based optimization method counteracted the effect of raw material variability by changing the process parameters and achieve the desired final tablet attributes. However, applicability and performance of the PLS-based optimization method are limited because PLS cannot cope with nonlinearity and changes in process characteristics. To overcome such weaknesses, locally weighted PLS (LW-PLS) was proposed and has been applied to various industrial processes (Kim et al. (2011); Nakagawa et al. (2012); Kim et al. (2013a)). LW-PLS is a kind of just-in-time (JIT) modeling methods, in which local PLS models are built on demand by using a database. A local PLS model is developed from past input-output data around a query i.e., a new sample, when an output estimation is required. Since local models built adaptively, LW-PLS can cope with changes in process characteristics as well as process nonlinearity.

This present work proposes a new optimization method based on LW-PLS. To solve a nonlinear optimization problem based on LW-PLS, in which any global model is not constructed, we adopt evolutionary computing. The famous algorithms are particle swarm optimization (PSO) (Kennedy (2010)) and differential evolution (DE) (Storn and Price (1997)) in this field. PSO has been applied to many practical problems since the algorithm is easy to understand; in contrast, it has been demonstrated that DE is superior to PSO in convergence properties through some benchmark functions (Vesterstrom and Thomsen (2004)). DE is a simple yet powerful algorithm for global optimization. However, it is not easy to tune parameters of DE, which are kept fixed throughout the entire evolutionary process. To deal with the problem, self-adaptive DE (jDE) was proposed; its effectiveness was demonstrated through some benchmark functions (Brest et al. (2006)).

⋆ This work was partially supported by Japan Society for the Promotion of Science (JSPS), Grant-in-Aid for Scientific Research (C) 24560940.

The algorithm controls two of three parameters of DE every evolutionary process. In this work, jDE is adopted to solve the global optimization problem.

In the following sections, modeling methods and optimization methods are explained. Then, the optimization method based on LW-PLS and jDE is described; the validity of the proposed method is demonstrated through a numerical example and an industrial case study of a pharmaceutical granulation process.

2. MODELING METHODS

2.1 Partial Least Squares (PLS)

PLS is the most widely used method in chemometrics for multivariate calibration and finds increasing interest also in other areas (Varmuza and Filzmoser (2008)). It is a linear regression method and can be executed through several algorithms. PLS2 for multiple output systems is described here.

Let assume that we have N measurements of M input variables and L output variables. In PLS2, an input variable matrix $\boldsymbol{X} \in \Re^{N\times M}$ and an output variable matrix $\boldsymbol{Y} \in \Re^{N\times L}$ are decomposed by using R latent variables.

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E} = \sum_{r=1}^{R} \boldsymbol{t}_r \boldsymbol{p}_r^T + \boldsymbol{E} \tag{1}$$

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Q}^T + \boldsymbol{F} = \sum_{r=1}^{R} \boldsymbol{u}_r \boldsymbol{q}_r^T + \boldsymbol{F} \tag{2}$$

where $\boldsymbol{T} = \begin{bmatrix}\boldsymbol{t}_1\, \boldsymbol{t}_2\, \cdots\, \boldsymbol{t}_R\end{bmatrix} \in \Re^{N\times R}$ and $\boldsymbol{U} = \begin{bmatrix}\boldsymbol{u}_1\, \boldsymbol{u}_2\, \cdots\, \boldsymbol{u}_R\end{bmatrix} \in \Re^{N\times R}$ are latent variable matrices (score matrices) of $\boldsymbol{X}$ and $\boldsymbol{Y}$, $\boldsymbol{P} = \begin{bmatrix}\boldsymbol{p}_1\, \boldsymbol{p}_2\, \cdots\, \boldsymbol{p}_R\end{bmatrix} \in \Re^{M\times R}$ and $\boldsymbol{Q} = \begin{bmatrix}\boldsymbol{q}_1\, \boldsymbol{q}_2\, \cdots\, \boldsymbol{q}_R\end{bmatrix} \in \Re^{L\times R}$ are loading matrices of $\boldsymbol{X}$ and $\boldsymbol{Y}$, and $\boldsymbol{E}$ and $\boldsymbol{F}$ are error matrices. The score matrices are expressed as

$$\boldsymbol{T} = \boldsymbol{X}\boldsymbol{W} \tag{3}$$

$$\boldsymbol{U} = \boldsymbol{Y}\boldsymbol{C} \tag{4}$$

where $\boldsymbol{W} \in \Re^{M\times R}$ and $\boldsymbol{C} \in \Re^{L\times R}$ are weighting matrices of $\boldsymbol{X}$ and $\boldsymbol{Y}$. In addition, $\boldsymbol{T}$ and $\boldsymbol{U}$ are related by

$$\boldsymbol{U} = \boldsymbol{T}\boldsymbol{D} + \boldsymbol{H} \tag{5}$$

where $\boldsymbol{D} \in \Re^{R\times R}$ is a regression coefficient diagonal matrix and $\boldsymbol{H}$ is error matrix.

For any new sample $\boldsymbol{x}_q$, a score vector $\boldsymbol{\tau}_q = \begin{bmatrix}\tau_1\, \tau_2\, \ldots\, \tau_R\end{bmatrix}^T$ and an estimated input variable $\hat{\boldsymbol{x}}_q$ can be computed as

$$\boldsymbol{\tau}_q^T = \boldsymbol{x}_q^T \boldsymbol{W} \tag{6}$$

$$\hat{\boldsymbol{x}}_q^T = \boldsymbol{\tau}_q^T \boldsymbol{P}^T. \tag{7}$$

To verify the validity of the model for the new sample, the Hotelling's T^2 and the squared prediction error Q are often used.

$$T_q^2 = \sum_{r=1}^{R} \frac{\tau_r^2}{\sigma_{t_r}^2} \tag{8}$$

$$Q_q = ||\boldsymbol{x}_q - \hat{\boldsymbol{x}}_q||^2 \tag{9}$$

where σ_{t_r} is a standard deviation of $\boldsymbol{t}_r$.

2.2 Locally Weighted PLS (LW-PLS)

LW-PLS is a JIT modeling method, which can cope with changes in process characteristics as well as nonlinearity. In LW-PLS, a local PLS model is built from past data around a query only when an output estimation is required.

Assume that input and output variable matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ are stored in a database. A local PLS model is constructed by weighting samples with a similarity matrix

$$\boldsymbol{\Omega} = \mathrm{diag}(\omega_1, \omega_2, \ldots, \omega_N) \in \Re^{N\times N} \tag{10}$$

where similarity ω_n depends on a weighted euclidean distance d_n between a query $\boldsymbol{x}_q$ and samples $\boldsymbol{x}_n$.

$$\omega_n = \exp\left(-\frac{d_n \varphi}{\sigma_d}\right) \tag{11}$$

$$d_n = \sqrt{(\boldsymbol{x}_n - \boldsymbol{x}_q)^T \boldsymbol{\Theta} (\boldsymbol{x}_n - \boldsymbol{x}_q)} \tag{12}$$

$$\boldsymbol{\Theta} = \mathrm{diag}(\theta_1,\ \theta_2,\ \ldots,\ \theta_M) \in \Re^{M\times M} \tag{13}$$

where σ_d is a standard deviation of $\{d_n\}$, φ is a localization parameter, $\boldsymbol{\Theta}$ is a weight matrix, and θ_m is a weight for the m-th input variable. Since LW-PLS derives a local model from past samples similar to a query, the similarity should be appropriately defined to realize accurate estimation (Nakagawa et al. (2012); Kano and Fujiwara (2013); Kim et al. (2013b)). LW-PLS is equivalent to PLS when the localization parameter is set as $\varphi = 0$; it includes PLS as a special case. Therefore, estimation performance of LW-PLS is better than or at least the same as that of PLS.

Given a query $\boldsymbol{x}_q$, the LW-PLS algorithm using singular value decomposition is described as follows.

1) Determine the number of latent variables R and set $r = 1$.
2) Calculate the similarity matrix $\boldsymbol{\Omega}$ based on Eqs. (10)-(13).
3) Calculate $\boldsymbol{X}_r$, $\boldsymbol{Y}_r$, and $\boldsymbol{x}_{q,r}$;

$$\boldsymbol{X}_r = \boldsymbol{X} - \boldsymbol{1}_N \begin{bmatrix}\bar{x}_1\ \bar{x}_2\ \cdots\ \bar{x}_M\end{bmatrix} \tag{14}$$

$$\boldsymbol{Y}_r = \boldsymbol{Y} - \boldsymbol{1}_N \begin{bmatrix}\bar{y}_1\ \bar{y}_2\ \cdots\ \bar{y}_L\end{bmatrix} \tag{15}$$

$$\boldsymbol{x}_{q,r} = \boldsymbol{x}_q - \begin{bmatrix}\bar{x}_1\ \bar{x}_2\ \cdots\ \bar{x}_M\end{bmatrix}^T \tag{16}$$

$$\bar{x}_m = \frac{\sum_{n=1}^{N} \omega_n x_{nm}}{\sum_{n=1}^{N} \omega_n} \tag{17}$$

$$\bar{y}_l = \frac{\sum_{n=1}^{N} \omega_n y_{nl}}{\sum_{n=1}^{N} \omega_n} \tag{18}$$

where $\boldsymbol{1}_N \in \Re^N$ is a column vector with all its entires equal to 1.
4) Set $\hat{\boldsymbol{y}}_q = \begin{bmatrix}\bar{y}_1\ \bar{y}_2\ \cdots\ \bar{y}_L\end{bmatrix}^T$.
5) Set left and right singular vectors of $\boldsymbol{X}_r^T \boldsymbol{\Omega} \boldsymbol{Y}_r$ as $\boldsymbol{w}_r$ and $\boldsymbol{c}_r$ respectively, which correspond to the maximum singular value.
6) Derive the r-th latent variables of $\boldsymbol{X}$ and $\boldsymbol{Y}$, and the r-th regression coefficient;

$$\boldsymbol{t}_r = \boldsymbol{X}_r \boldsymbol{w}_r \tag{19}$$

$$\boldsymbol{u}_r = \boldsymbol{Y}_r \boldsymbol{c}_r \tag{20}$$

$$d_r = \frac{\boldsymbol{t}_r^T \boldsymbol{\Omega} \boldsymbol{u}_r}{\boldsymbol{t}_r^T \boldsymbol{\Omega} \boldsymbol{t}_r}. \tag{21}$$

7) Derive the r-th latent variable of $\boldsymbol{x}_q$;

$$t_{q,r} = \boldsymbol{x}_{q,r}^T \boldsymbol{w}_r \tag{22}$$

$$u_{q,r} = d_r t_{q,r}. \tag{23}$$

8) Derive the r-th loading vectors of $\boldsymbol{X}$ and $\boldsymbol{Y}$;

$$\boldsymbol{p}_r = \frac{\boldsymbol{X}_r^T \boldsymbol{\Omega} \boldsymbol{t}_r}{\boldsymbol{t}_r^T \boldsymbol{\Omega} \boldsymbol{t}_r} \tag{24}$$

$$\boldsymbol{q}_r = \frac{\boldsymbol{Y}_r^T \boldsymbol{\Omega} \boldsymbol{u}_r}{\boldsymbol{u}_r^T \boldsymbol{\Omega} \boldsymbol{u}_r}. \tag{25}$$

9) Update $\hat{\boldsymbol{y}}_q = \hat{\boldsymbol{y}}_q + u_{q,r} \boldsymbol{q}_r$.
10) If $r = R$, end. Otherwise, calculate $\boldsymbol{X}_{r+1}$, $\boldsymbol{Y}_{r+1}$, and $\boldsymbol{x}_{q,r+1}$;

$$\boldsymbol{X}_{r+1} = \boldsymbol{X}_r - \boldsymbol{t}_r \boldsymbol{p}_r^T \tag{26}$$

$$\boldsymbol{Y}_{r+1} = \boldsymbol{Y}_r - \boldsymbol{t}_r d_r \boldsymbol{c}_r^T \tag{27}$$

$$\boldsymbol{x}_{q,r+1} = \boldsymbol{x}_{q,r} - t_{q,r} \boldsymbol{p}_r. \tag{28}$$

Then set $r = r + 1$, and go to step 5.

3. OPTIMIZATION METHODS

3.1 Differential Evolution (DE)

DE is a simple yet powerful algorithm for global optimization (Storn and Price (1997)). DE has become popular because good convergence properties were demonstrated through some benchmark functions (Vesterstrom and Thomsen (2004)). There are several variants of DE; a DE/best/1/bin strategy, which is more robust than the other strategies (Mezura-Montes et al. (2006)), is described here. Mutation, crossover, and selection are the important processes to generate next generation vectors. By iterating these processes, DE searches for an optimum solution. It is noted that the DE algorithm cannot guarantee optimality.

Figure 1 illustrates the mutation process in $A = 2$ dimensions. In the DE algorithm, a vector operated in the following processes is called a target vector. For the target vector $\boldsymbol{x}_{i,g} \in \Re^A$, which is the i-th and g-generation vector, a mutant vector $\boldsymbol{v}_{i,g+1}$ is generated as

$$\boldsymbol{v}_{i,g+1} = \boldsymbol{x}_{\mathrm{best},g} + F(\boldsymbol{x}_{r_1,g} - \boldsymbol{x}_{r_2,g}) \tag{29}$$

where r_1, r_2 $(r_1 \neq r_2 \neq i) \in [1, NP]$ are index numbers chosen randomly, NP denotes the number of the target vectors, $F \in [0, 2]$ is an amplification factor, and $\boldsymbol{x}_{\mathrm{best},g}$ is the best vector, which minimizes an objective function value in g-generation.

Figure 2 illustrates the crossover process in $A = 6$ dimensions. The target vector $\boldsymbol{x}_{i,g}$ is mixed with the mutant vector $\boldsymbol{v}_{i,g+1}$ to yield a trial vector $\boldsymbol{u}_{i,g+1} = [u_{i1,g+1}\, u_{i2,g+1}\, \ldots\, u_{iA,g+1}]^T$ given by

$$u_{ij,g+1} = \begin{cases} v_{ij,g+1} & \mathrm{rand}_{ij} \leq CR \text{ or } j = j_r \\ x_{ij,g} & \text{otherwise} \end{cases} \tag{30}$$

where $CR \in [0, 1]$ is a crossover control parameter, $\mathrm{rand}_{ij} \in [0, 1]$ is a uniform random number, and $j_r \in [1, A]$ is an index number generated randomly. By introducing j_r, it ensures that the trial vector takes over from the mutant vector at least one element.

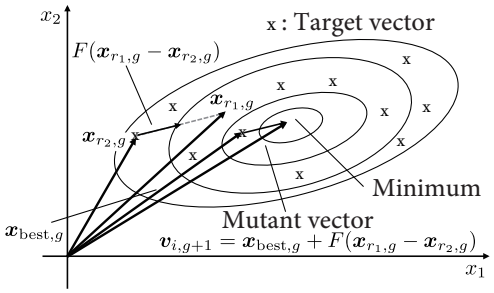


Fig. 1. Mutation process in $A = 2$

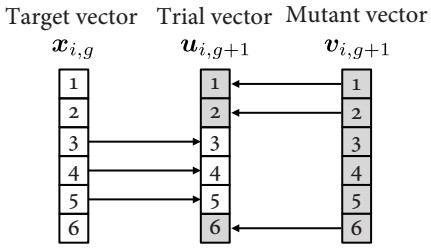


Fig. 2. Crossover process in $A = 6$

The selection process is used to generate the next target vector.

$$\boldsymbol{x}_{i,g+1} = \begin{cases} \boldsymbol{u}_{i,g+1} & J(\boldsymbol{u}_{i,g+1}) \leq J(\boldsymbol{x}_{i,g}) \\ \boldsymbol{x}_{i,g} & \text{otherwise} \end{cases} \tag{31}$$

where $J(\cdot)$ is an objective function to minimize.

3.2 Self-Adaptive DE (jDE)

Tuning three control parameters of DE, i.e., F, CR, and NP, is not easy and usually conducted by trial and error; therefore it is time-consuming. To improve the efficiency of DE, jDE controls F and CR every evolutionary process.

In jDE, the amplification factor F and the crossover control parameter CR are controlled (Brest et al. (2006)).

$$F_{i,g+1} = \begin{cases} F_l + \mathrm{rand}_1 \times F_u & \mathrm{rand}_2 < \tau_1 \\ F_{i,g} & \text{otherwise} \end{cases} \tag{32}$$

$$CR_{i,g+1} = \begin{cases} \mathrm{rand}_3 & \mathrm{rand}_4 < \tau_2 \\ CR_{i,g} & \text{otherwise} \end{cases} \tag{33}$$

where F_l, F_u, τ_1, τ_2 are control parameters of F and CR, and $\mathrm{rand}_i \in [0, 1]$ $(i = 1, 2, 3, 4)$ is a uniform random number. In general, these parameters are set as $F_l = 0.1$, $F_u = 0.9$, and $\tau_1 = \tau_2 = 0.1$ (Brest et al. (2006)). The mutant vector $\boldsymbol{v}_{i,g+1}$ and the trial vector $\boldsymbol{u}_{i,g+1}$ is yielded by using $F_{i,g+1}$ and $CR_{i,g+1}$.

$$\boldsymbol{v}_{i,g+1} = \boldsymbol{x}_{\mathrm{best},g} + F_{i,g+1}(\boldsymbol{x}_{r_1,g} - \boldsymbol{x}_{r_2,g}) \tag{34}$$

$$u_{ij,g+1} = \begin{cases} v_{ij,g+1} & \mathrm{rand}_{ij} \leq CR_{i,g+1} \text{ or } j = j_r \\ x_{ij,g} & \text{otherwise.} \end{cases} \tag{35}$$

4. OPTIMIZATION BASED ON LW-PLS AND JDE

LW-PLS can cope with changes in process characteristics as well as nonlinearity, and a problem of model maintenance can be solved. jDE has an important advantage

over standard optimization methods; it is not necessary to derive a gradient of a global objective function $J(\cdot)$ in evolutionary computation.

In this research, a new process optimization method based on LW-PLS and jDE is proposed. Given raw material attributes $\boldsymbol{x}_{raw}$, the optimization problem to find the best process parameters $\boldsymbol{z} = \left[z_1\, z_2\, \cdots\, z_I\right]^T$ is formulated as

$$\min_{\boldsymbol{z}} J(\boldsymbol{z}) = \underbrace{(\boldsymbol{y}_{\mathrm{des}} - \hat{\boldsymbol{y}})^T \boldsymbol{\Lambda} (\boldsymbol{y}_{\mathrm{des}} - \hat{\boldsymbol{y}})}_{J_q(\boldsymbol{z})} + \underbrace{\boldsymbol{\lambda}^T \boldsymbol{z}}_{J_o(\boldsymbol{z})} \tag{36}$$

subject to

$$\begin{array}{ll} \hat{\boldsymbol{y}} = f(\boldsymbol{x}) & \boldsymbol{x} = \left[\boldsymbol{x}_{raw}^T\ \boldsymbol{z}^T\right]^T \\ y_{l,\min} \le \hat{y}_l \le y_{l,\max} & z_{i,\min} \le z_i \le z_{i,\max} \\ T^2 \le T^2_{\max} & Q \le Q_{\max} \end{array} \tag{37}$$

where J_q denotes a sum of weighted squared errors between desired product quality attributes $\boldsymbol{y}_{\mathrm{des}}$ and estimated values $\hat{\boldsymbol{y}}$, J_o denotes an operational cost, $f(\boldsymbol{x})$ is a LW-PLS model that is built every time when $\boldsymbol{x}$ is given, $\boldsymbol{\Lambda}$ is a weight matrix, $\boldsymbol{\lambda}$ is a weight vector, and the subscripts min and max denote the lower limit and the upper limit, respectively. $T^2_{\max}$ and $Q_{\max}$ of each LW-PLS model are derived in the following procedure.

(1) Calculate the similarity matrix $\boldsymbol{\Omega}$ for a query $\boldsymbol{x}_q$.
(2) Compose a new input variable matrix $\boldsymbol{X}_s \in \Re^{S\times M}$ using S samples from a database in descending order of similarity, where S $(S \le N)$ is a parameter for determining $T^2_{\max}$ and $Q_{\max}$.
(3) Set $T^2_{\max}$ and $Q_{\max}$ as the maximum values of T^2 and Q of $\boldsymbol{X}_s$.

The optimization procedure based on LW-PLS and jDE is as follows.

1) Set $g = 1$ and generate initial target vectors $\boldsymbol{x}_{1,1}$, …, $\boldsymbol{x}_{NP,1} \in \mathcal{S}$, where $\mathcal{S}$ is a subset satisfying the constraints given by Eq. (37).
2) Evaluate the initial target vectors based on Eq. (36) and determine the best vector $\boldsymbol{x}_{\mathrm{best},1}$.
3) Generate mutant vectors $\boldsymbol{v}_{1,g+1}$, …, $\boldsymbol{v}_{NP,g+1}$ based on Eqs. (32) and (34).
4) Generate trial vectors $\boldsymbol{u}_{1,g+1}, \ldots, \boldsymbol{u}_{NP,g+1}$ based on Eqs. (33) and (35).
5) Generate next generation target vectors $\boldsymbol{x}_{1,g+1}$, …, $\boldsymbol{x}_{NP,g+1}$ based on Eq. (31).
6) Evaluate the next generation vectors based on Eq. (36) and determine the best vector $\boldsymbol{x}_{\mathrm{best},g+1}$.
7) Set $g = g + 1$. If $g = g_{\max}$, where $g_{\max}$ denotes a maximum generation value, set a solution as $\boldsymbol{x}_{\mathrm{best},g_{\max}}$ and end. Otherwise, go to step 3.

5. CASE STUDIES

In this section, the proposed optimization method based on LW-PLS is applied to a numerical example, which is a well-known benchmark function, and a pharmaceutical granulation process. It is noted that all input and output variables were normalized so that each variable has zero mean and unit variance. To evaluate estimation performance, root mean squared error of prediction (RMSEP) is used.

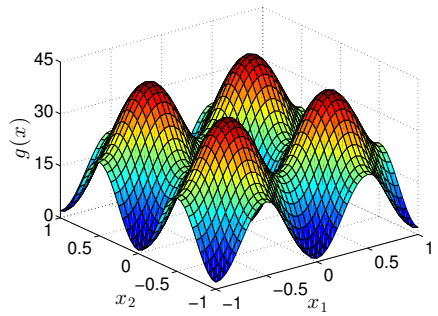


Fig. 3. Rastrigin function $g(\boldsymbol{x})$

$$\mathrm{RMSEP} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_{nl} - \hat{y}_{nl})^2} \tag{38}$$

where y_{nl} is a measured value of the n-th sample of the l-th output variable and $\hat{y}_{nl}$ is an estimated value of y_{nl}.

5.1 Numerical Example

In this example, Rastrigin function

$$g(\boldsymbol{x}) = 20 + \sum_{i=1}^{2} x_i^2 - 10\cos(2\pi x_i) \tag{39}$$
$$x_i \in \left[-1, 1\right]\ (i = 1, 2)$$

is used to test optimization performance. Figure 3 shows Rastrigin function $g(\boldsymbol{x})$, which has the minimum:

$$g_{\min}(\boldsymbol{x}^*) = 0, \quad \boldsymbol{x}^* = \boldsymbol{0}. \tag{40}$$

Modeling PLS and LW-PLS models were built from 10000 samples, which were generated randomly in the domain of the Rastrigin function. The similarity matrix $\boldsymbol{\Theta}$ was set as a unit matrix. The parameters of PLS and LW-PLS were tuned by using leave-one-out cross validation (LOOCV); the numbers of latent variables of PLS and LW-PLS were set as 1 and 2, and localization parameter of LW-PLS was set as $\varphi = 130$.

Optimization Muteki et al. (2011) formulated the optimization problem based on PLS and its problem was solved by SQP; the approach was adopted when PLS was used. On the other hand, jDE was adopted as the optimization method when LW-PLS was used. In this numerical example, the optimization problem was formulated as

$$\min_{\boldsymbol{x}} J_1(\boldsymbol{x}) = (y_{\mathrm{des}} - \hat{y})^2 \tag{41}$$

subject to

$$\begin{array}{ll} \hat{y} = f(\boldsymbol{x}) & \\ y_{\min} \le \hat{y} \le y_{\max} & x_{m,\min} \le x_m \le x_{m,\max} \\ T^2 \le T^2_{\max} & Q \le Q_{\max} \end{array}$$

where $f(\boldsymbol{x})$ was a PLS or LW-PLS model.

The desired value was set as $y_{\mathrm{des}} = 0$. The parameters of jDE are shown in Table 1. The number of samples for the constraints of LW-PLS was set as $S = 100$. The optimization results are shown in Table 2, where J_{1t} is a true objective function value calculated using the true

Table 1. Parameters of jDE in numerical example

NP	$g_{\max}$	$F_{i,1}$	$CR_{i,1}$	F_l	F_u	τ_1	τ_2
100	1000	0.7	0.95	0.1	0.9	0.1	0.1

Table 2. Optimization results in numerical example

Model	$J_1(\boldsymbol{x})$	$J_{1t}(\boldsymbol{x})$	x_1	x_2
PLS	20.269	2.025	-0.983	1.000
LW-PLS	0.000	0.021	-0.009	0.005

Table 3. Input variables x and output variables y in granulation process

Variable	Name
x_1	Inlet air temperature [G]
x_2	Inlet air volume 1 [G]
x_3	Inlet air volume 2 [G]
x_4	Spray rate [G]
x_5	Spray air volume [G]
x_6	Inlet air temperature [D]
x_7	Inlet air volume [D]
x_8	Product temperature [D]
y_1	Maximum moisture content
y_2	Specific volume (Static)
y_3	Specific volume (Dynamic)
y_4	Particle size 1
y_5	Particle size 2
y_6	Particle size 3

function $y = g(\boldsymbol{x})$ instead of $f(\boldsymbol{x})$. The optimization performance was significantly improved by using LW-PLS in comparison with PLS, because $g(\boldsymbol{x})$ was nonlinear. We can conclude that optimization based on LW-PLS is superior to the method based on PLS.

5.2 Pharmaceutical Granulation Process

The target of this case study is a commercial scale pharmaceutical granulation process. The objective is to derive process parameters, which make granule attributes achieve nearly their target values and reduce the operational cost simultaneously.

Modeling Input variables and output variables are listed in Table 3, where [G] and [D] denote a granulation step and a dry step in the granulation process. In this case study, six output variables were estimated by using eight input variables. There were 40 samples; 30 samples were used for modeling and the other 10 samples were used for validation.

The similarity matrix $\boldsymbol{\Theta}$ was set as a unit matrix. The parameters of PLS and LW-PLS were determined through LOOCV; the numbers of latent variables of PLS and LW-PLS were set as 4 and 3, and localization parameter of LW-PLS was set as $\varphi = 1.2$. The estimation results are shown in Table 4, where R denotes the correlation coefficient between the measured values and the estimated values. The scatter plots between the measurements and the estimates of PLS and LW-PLS are shown in Fig. 4. Due to the limit of the available pages, we show only y_1 and y_2. The mean of RMSEP of LW-PLS (0.51) is smaller than that of PLS (0.53); LW-PLS outperforms PLS in terms of the estimation performance.

Table 4. Estimation performance of PLS and LW-PLS in granulation process

		y_1	y_2	y_3	y_4	y_5	y_6
PLS	RMSEP	0.58	0.60	0.70	0.41	0.35	0.55
	R	0.82	0.92	0.84	0.92	0.87	0.71
LW-PLS	RMSEP	0.45	0.62	0.66	0.54	0.35	0.42
	R	0.90	0.90	0.84	0.83	0.88	0.93

Table 5. Parameters of jDE in granulation process

NP	$g_{\max}$	$F_{i,1}$	$CR_{i,1}$	F_l	F_u	τ_1	τ_2
2000	200	0.7	0.95	0.1	0.9	0.1	0.1

Optimization The eight process parameters are optimized to keep the six granule attributes closer to their target values and to achieve the lower operational cost by the proposed method. The objective function in this problem is as follows.

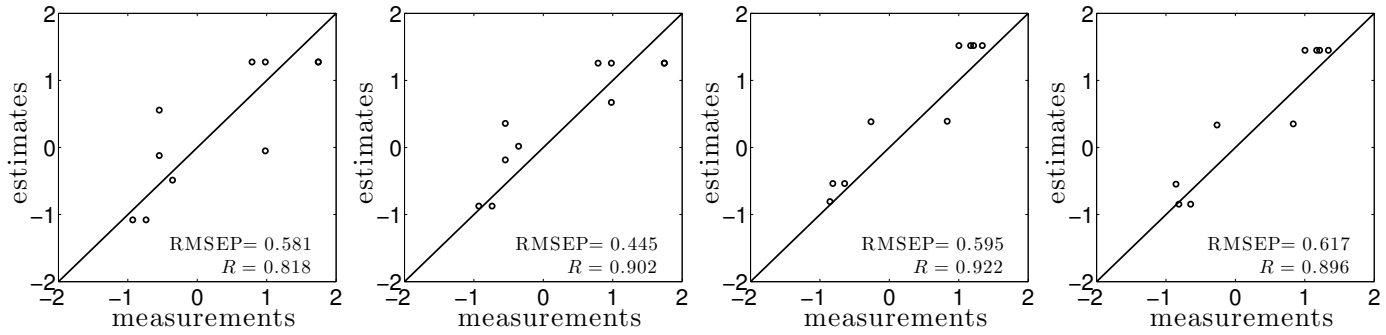
$$\begin{aligned}
&\min_{\boldsymbol{x}} J_2(\boldsymbol{x}) = \alpha_1 J_{2q}(\boldsymbol{x}) + \alpha_2 J_{2o}(\boldsymbol{x}) \qquad (42)\\
&\quad J_{2q}(\boldsymbol{x}) = (\boldsymbol{y}_{\mathrm{des}} - \hat{\boldsymbol{y}})^T \boldsymbol{\Lambda} (\boldsymbol{y}_{\mathrm{des}} - \hat{\boldsymbol{y}})\\
&\quad J_{2o}(\boldsymbol{x}) = \lambda x_4\\
&\qquad \boldsymbol{\Lambda} = \mathrm{diag}(60, 10, 10, 30, 30, 30) \quad \lambda = 21.78\\
&\quad \text{subject to}\\
&\qquad \hat{\boldsymbol{y}} = f(\boldsymbol{x})\\
&\qquad y_{l,\min} \le y_l \le y_{l,\max} \qquad x_{m,\min} \le x_m \le x_{m,\max}\\
&\qquad T^2 \le T^2_{\max} \qquad Q \le Q_{\max}
\end{aligned}$$

where J_{2q} denotes a sum of weighted squared errors, J_{2q} denotes an operational cost, and α_1 and α_2 are weights for J_{2q} and J_{2o}.

The parameters of jDE are shown in Table 5. The number of samples for the constraints of LW-PLS was set as $S = 10$. Two solutions changing α_1 are shown in Table 6 (Ex1: $\begin{bmatrix}\alpha_1 & \alpha_2\end{bmatrix} = \begin{bmatrix}100 & 1\end{bmatrix}$, Ex2: $\begin{bmatrix}\alpha_1 & \alpha_2\end{bmatrix} = \begin{bmatrix}10000 & 1\end{bmatrix}$). When $\alpha_1 = 100$, the objective focused on the operational cost J_{2q}; it was demonstrated that the cost was smaller that the cost of another solution. On the other hand, the objective focused on the quality attributes when $\alpha_1 = 10000$; the obtained parameters made the estimated values reach the desired values closely.

6. CONCLUSION

In the present work, a new optimization method based on LW-PLS and jDE was proposed. It was demonstrate that the proposed method is superior to the conventional PLS-based method through a numerical example. In addition, its optimization method was applied to the commercial scale pharmaceutical granulation process, in which the process parameters were optimized. Then the granule attributes reached the desired values and the operational cost was minimized simultaneously.



(a-1) y_1 (PLS) (a-2) y_1 (LW-PLS) (b-1) y_2 (PLS) (b-2) y_2 (LW-PLS)

Fig. 4. Scatter plots between measurements and estimates of PLS and LW-PLS in granulation process

Table 6. Optimization results based on LW-PLS and jDE in granulation process (Ex1: $[\alpha_1\ \alpha_2] = [100\ 1]$, Ex2: $[\alpha_1\ \alpha_2] = [10000\ 1]$)

	min	max	$\boldsymbol{y}_{\mathrm{des}}$	Ex1	Ex2
$J(\boldsymbol{x})$	-	-	-	2405	2818
$J_{2q}(\boldsymbol{x})$	-	-	-	0.092	0.037
$J_{2o}(\boldsymbol{x})$	-	-	-	2396	2445
x_1	90	90	-	90	90
x_2	6	8	-	6.85	6.81
x_3	8	10	-	8.34	8.51
x_4	110	180	-	110	112
x_5	120	180	-	165	168
x_6	90	90	-	90	90
x_7	6	8	-	6.76	6.73
x_8	50	50	-	50	50
y_1	3.10	4.80	4.00	4.00	4.00
y_2	3.12	3.93	3.50	3.50	3.50
y_3	2.40	3.03	2.75	2.75	2.75
y_4	62.0	145	100	101	100
y_5	152	300	205	204	204
y_6	287	654	365	363	365

REFERENCES

Brest, J., Greiner, S., Boskovic, B., Mernik, M., and Zumer, V. (2006). Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems. *Evolutionary Computation, IEEE Transactions on*, 10(6), 646–657.

Kano, M. and Fujiwara, K. (2013). Virtual sensing technology in process industries: trends and challenges revealed by recent industrial applications. *JCEJ*, 46(1), 1–17.

Kennedy, J. (2010). Particle swarm optimization. In *Encyclopedia of Machine Learning*, 760–766. Springer.

Kim, S., Kano, M., Hasebe, S., Takinami, A., and Seki, T. (2013a). Long-term industrial applications of inferential control based on just-in-time soft-sensors: Economical impact and challenges. *Ind. Eng. Chem. Research*, 52(35), 12346–12356.

Kim, S., Kano, M., Nakagawa, H., and Hasebe, S. (2011). Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection. *Int. J. Pharm.*, 421(2), 269–274.

Kim, S., Okajima, R., Kano, M., and Hasebe, S. (2013b). Development of soft-sensor using locally weighted pls with adaptive similarity measure. *Chemom. Intell. Lab. Syst.*, 124, 43–49.

Lawrence, X.Y. (2008). Pharmaceutical quality by design: product and process development, understanding, and control. *Pharm. Research*, 25(4), 781–791.

Mezura-Montes, E., Velázquez-Reyes, J., and Coello Coello, C.A. (2006). A comparative study of differential evolution variants for global optimization. In *Proc. of the 8th Annu. Conf. on Genetic and Evol. Computation*, 485–492. ACM.

Muteki, K., Swaminathan, V., Sekulic, S.S., and Reid, G.L. (2011). De-risking pharmaceutical tablet manufacture through process understanding, latent variable modeling, and optimization technologies. *AAPS PharmSciTech*, 12(4), 1324–1334.

Nakagawa, H., Tajima, T., Kano, M., Kim, S., Hasebe, S., Suzuki, T., and Nakagami, H. (2012). Evaluation of infrared-reflection absorption spectroscopy measurement and locally weighted partial least-squares for rapid analysis of residual drug substances in cleaning processes. *Anal. Chem.*, 84(8), 3820–3826.

Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *J. of Global Optimization*, 11(4), 341–359.

US FDA et al. (2006). Guidance for industry: Q8 pharmaceutical development, us department of health and human service. *FDA, Rockville, MD*.

Varmuza, K. and Filzmoser, P. (2008). *Introduction to multivariate statistical analysis in chemometrics*. CRC press.

Vesterstrom, J. and Thomsen, R. (2004). A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. In *Evol. Computation Congr.*, volume 2, 1980–1987. IEEE.