European Symposium on Computer Aided Process Engineering – 15
L. Puigjaner and A. Espuña (Editors)
© 2005 Elsevier Science B.V. All rights reserved.

The Statistical Simplex Method for Experimental Optimization with Process Data

Ernesto C. Martínez*
INGAR-Instituto de Desarrollo y Diseño (CONICET-UTN)
Avellaneda 3657, S3002 GJC, Santa Fe, Argentina

Abstract

Experimental optimization with scarce and noisy process data is a key issue in laboratory automation for faster chemical process research and development, real-time process optimization, extremum-seeking control systems and self-calibrating instruments. To deal successfully with noise and uncontrollable factors in experimental design for process optimization a statistical characterization of a local optimum is proposed. The Kendall's tau statistic is used for characterizing a local optimum as a cluster center of strongly correlated points. A statistical simplex algorithm that resorts to correlation-based ranking of simplex vertices for reflection, expansion, contraction and shrinking steps is proposed. Results obtained in run-to-run optimization of the operating policy of a semi-batch reactor are presented.

Keywords: Process development, Experimental optimization, Simplex search method, Non-parametric statistics, Batch-to-batch optimization.

1. Introduction

Finding optimal operating conditions fast for process systems is a key competitive factor in fine chemicals and pharmaceutical industrial sectors to increase the value added in many stages of a product/process lifecycle, including recipe development, scale-up and run-to-run optimization (Öberg and Deming, 2000; Matsumoto, Du and Lindsey, 2002). Experimental optimization is also relevant for on-a-chip implementation of extremum-seeking control systems and self-calibrating instruments. Three factors: noise, time and cost, constitute major stumbling blocks to the design of algorithms for evolutionary optimization with process data. The Simplex search method originally proposed by Nelder and Mead (Lagarias *et al*., 1998) has been successfully used for mathematical functions but has severe difficulties to deal with noise, bias and process discontinuities that are ubiquitous in experimental optimization problems. As it is the case for all optimization algorithms, the simplex search method has been designed to handle outputs from a mathematical function. When the method is used with process data which may incorporate noise or outliers stagnation problems are often encountered (Kelley, 1999; McKinnon, 1998).

* *Author email:* ecmarti@ceride.gov.ar

2. Correlation and optimality

The performance index $g(\boldsymbol{x})$ to be optimized is typically unknown except for the scarce, noisy and biased information $y(\boldsymbol{x}_i)$, i=1,2,…, provided by a sequence of experimental trials. As a result, efficient detection of a local minimum or maximum in the face of output variability demands developing a statistical model for local optima. A robust and quite general assumption about sampled information in the vicinity of a minimum (maximum) is the *local monotonicity* assumption*:*

"*Sampled values of the objective function for inputs* $\boldsymbol{x}_i$ *closer to a local minimum (maximum)* $\boldsymbol{x}^*$ *should exhibit a greater degree of positive (negative) correlation than those of inputs that are farther away.*"

In mathematical terms, this assumption requires that, for a given candidate optimum $\boldsymbol{x}^*$, there locally exists a monotonic relationship φ such that:

$$y(\boldsymbol{x}) = \varphi(\left\|\boldsymbol{x} - \boldsymbol{x}^*\right\|) \tag{1}$$

It is worth noting that this guiding model for a local optimum does not impose any constraint on the shape of φ in the vicinity of the minimum (maximum) as long as the function is monotonically increasing (decreasing) with respect to the distance from the optimum. Furthermore, beyond the assumption of symmetry with respect to the distance metric, local optimum characterization is independent of any continuousness assumption for the $g(\boldsymbol{x})$, or its derivatives. Also, there is no assumption about a given noise distribution for $y(\boldsymbol{x}_i)$, i=1,2,…,. The only constraint is that the expected value of y, $E(y)$, for a given $\boldsymbol{x}$ is equal to $g(\boldsymbol{x})$.

Sampled data will locally fit the model of optima in Equation (1) to different extents. The existence of a local optimum $\boldsymbol{x}^*$ requires a monotonic association between $x_i = \left\|\boldsymbol{x}_i - \boldsymbol{x}^*\right\|$ and $g(\boldsymbol{x}_i)$. Under the influence of noise, we are interested in finding enough evidence to accept or reject the hypothesis of independence, i.e. no correlation between x_i and $y(\boldsymbol{x}_i), i = 1,2,....$, against the alternative that may correspond to a positive or negative correlation, depending if $\boldsymbol{x}^*$ is a local minimum or maximum, respectively. To test for the strength of this relationship between x_i and $y(\boldsymbol{x}_i), i = 1,2,....$, it is tackled here using nonparametric statistical methods that can be applied more broadly since much fewer assumptions about the data set need to be made. A typical non-parametric measure of correlation that can be used for this purpose is the *Kendall's correlation coefficient* τ (Gibbons, 1993).

Denoting the ranks of x_i, y_i by r_i, s_i respectively, the Kendall's tau coefficient is defined to measure the strength of the association (correlation) between the ranks of x and y. If such dependency between the ranks exists, then if we arrange the x-ranks in ascending order, i.e. so that $r_i = i$, then y-ranks should show, despite noise and bias, an increasing trend when there is positive association (local minimum) or negative association (local maximum). Accordingly, Kendall proposed that after arranging observations in the increasing order of x-ranks, we score each paired difference $s_j - s_i$ for $i = 1,2,...,n-1$ with $j > i$ as +1 if this difference is positive and –1 if this difference is negative. Kendall called these differences *concordances* and *discordances*, respectively, with respect to an expected monotonic trend between the ranks. Denoting

the sums of observed concordances and discordances by n_c and n_d respectively, the equation for the sample Kendall's coefficient τ is:

$$\tau = (n_c - n_d) \Big/ \frac{1}{2} n(n-1) \tag{2}$$

Since there exist $1/2\ n(n-1)$ pairs $s_j - s_i$, if all are concordances $n_c = 1/2\ n(n-1)$ and $n_d = 0$, then $t_K = 1$ and monotonic association is perfectly positive. Similarly, if all are discordances, monotonic association is perfectly negative with $\tau = 1$. If the rankings of x and y are independent we do expect a fair mix of concordances and discordances, whence $\tau \approx 0$.

2.1 Example

Consider the chemical reaction system conducted in an isothermal semi-batch reactor, which behaves according to the 'unknown' mechanism:

$$A + B \underset{k_2}{\overset{k_1}{\Leftrightarrow}} I \xrightarrow[k_3]{} C; \qquad B + B \xrightarrow[k_4]{} D \ \text{(order 2.5)}$$

At a given operating temperature, the 'true' kinetic parameters (which are also assumed unknown) have the following *mean* values: $\boldsymbol{k}_1 = 1.1355, \boldsymbol{k}_2 = 5.0, \boldsymbol{k}_3 = 4.5239$ and $\boldsymbol{k}_4 = 3.5880$, which correspond to 'perfect' temperature control. In industrial-size batch reactors, temperature control is often far from perfect because of severe nonlinear behaviour and poor modelling. As a result, the final state of each run will show unsystematic variations (see Table 1 below). The reaction system is operated in a semi-batch mode where a stream of reactant *B* ($[\mathrm{B}]_{feed} = 0.2$ moles per liter) is added to a 1000-litres vessel which initially contains 0.2 moles per litre of *A* and no *B,* and is filled to 50%. The objective is to maximize a productivity index (see Table 1) for the process defined as follows:

$$J(x_f) = (1000 \times ([C]_f \times V_f) / t_f \tag{3}$$

For *safety* reasons in downstream processing, the final concentration of unreacted *B* cannot be greater than 0.0032 moles per litre. Hence, if the remaining *B* has a concentration greater that this maximum the batch must be allowed to continue further until a final time t_f where this threshold is achieved; as a result, the reactor productivity is decreased. Due to a heat extraction constraint the feed addition rate is limited to a maximum value of 12.0 litre min^{-1}, whereas because of production scheduling it is expected that the overall reactor cycle does not exceed 180 min.

Suppose data in Table 1 has been obtained and the optimal solution $\boldsymbol{x}^*$ is hypothesized to be one with a semi-batch period (x_1) of 40 min. and a feed rate (x_2) of 9.5 liter/min. In order to assess for this optimum's degree of fit, let's test for a positive monotonic association between *J* and $d = \left\| \boldsymbol{x} - \boldsymbol{x}^* \right\|$. To test for the hypothesis that the optimum is $\boldsymbol{x} = (40, 9.5)^{\mathrm{T}}$, the value of τ and its significance need to be calculated. Table 2 provides ranked data from Table 1; the paired ranks are used to calculate the statistic τ for the hypothetical optimum yielding $\tau = -0.4545$ with $n_c = 17$ concordances and $n_d = 49$ discordances.

Testing for the significance of the hypothesis $H_0 : \tau = 0$ against the alternative $H_1 : \tau < 0$ requires a pre-calculated value of the Kendall's τ critical values for $n = 12$

and a chosen significance level $(1-\alpha)\%$ (Gibbons, 1993). Tables give nominal 5 and 1 per cent critical values for significance when $n=12$ in one-tail test as -0.3929 and –0.5455. Corresponding values for $n_d - n_c$ are 26 and 36, whereas an approximated Monte Carlo estimate of the exact one-tail probability *P* of rejecting H_0 when in fact it is true is 0.0396. Since $-0.4545 < -0.3929$, we reject H_0 and accept the postulated optimum with a confidence of 95%, on the other hand the available data does not provide strong statistical support to accept the postulated optimum at $\boldsymbol{x} = (40, 9.5)^{\mathrm{T}}$ with a confidence level of 99%.

Table 1. Sampled data for the example

Run #	$x1$	$x2$	Tfinal[min]	Index *J*
1	45.0	7.25	102	0.5200
2	50.0	10.00	152	0.4732
3	48.0	8.00	118	0.5099
4	55.0	9.00	156	0.4711
5	38.0	12.00	125	0.5480
6	40.0	9.5	109	0.5221
7	44.0	11.00	139	0.4816
8	54.0	9.20	156	0.4660
9	58.0	8.60	162	0.4435
10	50.0	7.00	112	0.5299
11	52.0	9.60	155	0.4795
12	55.5	9.00	160	0.4400

Table 2. Ranked data for the example

d-ranks	1	2	3	4	5	6	7	8	9	10	11	12
J-ranks	10	7	8	5	12	9	6	3	11	4	1	2

3. Statistical simplex search method

Simplex search methods (Lagarias et al., 1998)) resort to an effective device for parsimoniously sampling the input space in the search for an optimizer. A simplex is a set of *n*+1 points in $\Re^n$. The simplex search algorithm mainly resorts to four operations: *reflection*, *expansion*, *contraction* and *shrinking*. The reflection operation is aimed at determining a local direction for improving the objective function. The logic of the algorithm seeks to take advantage of such direction when the new vertex is better that all the vertices in the current simplex by an expansion operation. On the other hand, when the reflected vertex does not represent an improvement over any of the current vertices, a contraction operation is carried out. Shrinking towards the best vertex is done when contraction fails to produce a vertex that is at least better that the worst one in the current simplex. In the statistical simplex method, these operations are based on the correlation coefficient τ for hypothetical optima located in the simplex vertices and trial points resulting from simplex operations. Additionally, the statistical simplex algorithm resorts to the *replication* of the best vertex in the current simplex.

3.1. Algorithm

0. Initialize. Start with non-degenerated simplex for $\Re^n$.

1. **Ranking.** To label the best ($\boldsymbol{x}_1$), the worst ($\boldsymbol{x}_{n+1}$) and the next-to-worst ($\boldsymbol{x}_n$) vertices in the current simplex, hypothetical reflections of each vertex at a time and their corresponding Kendall are calculated. Assuming that we are seeking to minimize the process output, the worst vertex $\boldsymbol{x}_{n+1}$ is the one that provides a reflected point with the highest Kendall´s tau compared to alternative reflections, whereas the best vertex $\boldsymbol{x}_1$ corresponds to the hypothetical reflection with lowest Kendall´s tau. In case of ties, the corresponding function values for the involved vertices are used for ranking.

2. Reflection. Following vertex ordering, the next experiment is made for the operating conditions corresponding to the reflected point $\boldsymbol{x}_r$ defined as follows:

$$\boldsymbol{x}_r = 2\ \bar{\boldsymbol{x}} - \boldsymbol{x}_{n+1} \tag{4}$$

where $\bar{\boldsymbol{x}} = \sum_{i=1}^{n} \boldsymbol{x}_i / n$ is the centroid of the n best points. Using the full dataset, the Kendall´s tau $\tau_1, \tau_2, \ldots, \tau_n, \tau_{n+1}, \tau_r$ are obtained assuming the location an hypothetical local optimum at each vertex and the reflected point, respectively. If $\tau_n < \tau_r \leq \tau_1$ accept $\boldsymbol{x}_r$ in place of $\boldsymbol{x}_{n+1}$ to form a new simplex, do a **replication** experiment of the best vertex $\boldsymbol{x}_1$ and terminate iteration.

3. Expansion. If $\tau_r > \tau_1$, do an expansion experiment at

$$\boldsymbol{x}_e = 3\ \bar{\boldsymbol{x}} - 2\boldsymbol{x}_{n+1} \tag{5}$$

If $\tau_e > \tau_r$, accept $\boldsymbol{x}_e$ in place of $\boldsymbol{x}_{n+1}$ to form a new simplex and terminate iteration; otherwise, accept $\boldsymbol{x}_r$ in place of $\boldsymbol{x}_{n+1}$ and terminate iteration.

4. Contraction. Whenever $\tau_{n+1} < \tau_r \leq \tau_n$, an *outer contraction* experiment is carried out at

$$\boldsymbol{x}_{oc} = \frac{3}{2}\ \bar{\boldsymbol{x}} - \frac{1}{2}\boldsymbol{x}_{n+1} \tag{6}$$

If $\tau_{oc} > \tau_{n+1}$, accept $\boldsymbol{x}_{oc}$ in place of $\boldsymbol{x}_{n+1}$ to form a new simplex, do a replication experiment and terminate iteration; otherwise, do a **shrinking** step.

If $\tau_r \leq \tau_{n+1}$ an *inner contraction* experiment is done at

$$\boldsymbol{x}_{ic} = \frac{1}{2}\ \bar{\boldsymbol{x}} - \frac{1}{2}\boldsymbol{x}_{n+1} \tag{7}$$

If $\tau_{ic} > \tau_{n+1}$, accept $\boldsymbol{x}_{ic}$ in place of $\boldsymbol{x}_{n+1}$ to form a new simplex, do a replication and terminate iteration; otherwise, do a **shrinking** step.

5. Shrinking. This operation generates a smaller simplex by only retaining the best vertex $\boldsymbol{x}_1$. The coordinates of newly generated n vertices are calculated as:

$$\boldsymbol{x}_i^{'} = \boldsymbol{x}_1 + \vartheta.(\boldsymbol{x}_i^{'} - \boldsymbol{x}_1), 0 < \vartheta < 1, i = 2,\ldots,n,n+1 \tag{8}$$

The Kendall´s tau $\tau_2^{'},\ldots,\tau_n^{'},\tau_{n+1}^{'}$ for $\boldsymbol{x}_i^{'}, i = 2,\ldots,n,n+1$ are calculated using the current data set. If $\tau_i^{'} > \tau_i, \forall\ i = 2,\ldots,n,n+1$, n new experiments are done at $\boldsymbol{x}_i^{'}, i = 2,\ldots,n,n+1$ and the corresponding $\tau_2^{'},\ldots,\tau_n^{'},\tau_{n+1}^{'}$ are recalculated; if still $\tau_i^{'} > \tau_i, \forall\ i = 2,\ldots,n,n+1$ apply, then accept the shrinked simplex, do a replication experiment and terminate iteration. Otherwise, do a **replication experiment** and terminate iteration. The logic of the statistical simplex search method is in Fig. 1.

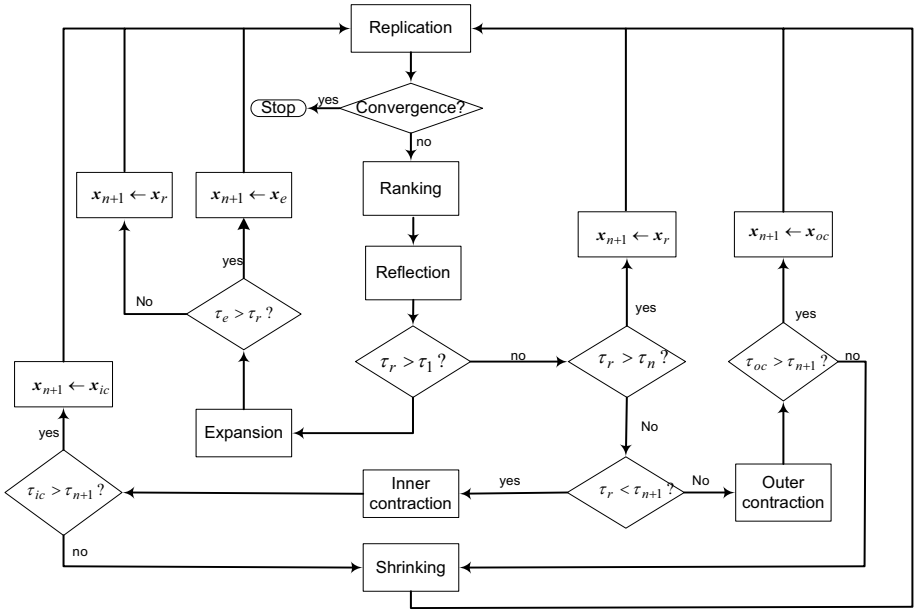

Fig. 1. The statistical simplex search method

3.2. Example (continued)

The statistical simplex method has been applied to the example in the presence of variability (σ^2=0.5) in the kinetic parameters. The results obtained are compared to the standard simplex method for an increasing number of experiments as shown in Table 3.

Table 3. Results for the statistical simplex method compared; $\boldsymbol{x_{inicial}}$=(56, 9)

Statistical Simplex	n=10 J_f=**0.49**	n=20 J_f=**0.52**	n=30 J_f=**0.52**	n=40 J_f=***0.53***
Nelder-Mead Simplex	n=10 J_f=**0.46**	n=20 J_f=**0.47**	n=30 J_f=**0.50**	n=40 J_f=**0.46**

4. Final comments

Based on a novel statistical characterization of optimality a new direct search algorithm using simplex designs for experimental optimization has been proposed. An important advantage of the statistical simplex is that any previous data can be used to advantage.

References

Gibbons J. D., 1993. *Nonparametric statistical methods*, SAGE Publications, London.
Kelley, C. T.,1999. SIAM J. Opt. 10, 43.
Lagarias, J.C., *et al*., 1998. SIAM J. Opt. 9, 112.
McKinnon, K.I.M.,1998. SIAM J. Opt. 9, 148
Matsumoto, T., Du, H. and Lindsey, J. S, 2002. Chemom. Intell. Lab. Syst. 62, 129.
Öberg, T. G. and Deming, S.N. 2000. Chem. Engng. Prog. 96, 53.