Multiresponse model based design parameters optimization: engineering approach

S.A. Popov[1,*], A.M. Abramov[2]

[1] Department Engineering Technology, Yaroslav-the-Wise State Novgorod University (NovSU),
ul. B.S. Peterburgskaya 41, Veliky Novgorod, 173003 Russia
[2] Department Engineering Technology, Yaroslav-the-Wise State Novgorod University (NovSU),
ul. B.S. Peterburgskaya 41, Veliky Novgorod, 173003 Russia

Abstract

A practical algorithm of determination of the set of Pareto optimal output parameters and a solution of the reverse optimization problem have been proposed. Basics of building and usage of multiresponse model are presented, which allows to perform multicriteria optimization optimization for different design and manufacturing situations.

1. Introduction

Well-founded definition of regions of admissible values of the vector of input (independent) constructive or technological parameters $\mathbf{X} = \{x_1, x_2, \ldots x_k\}^T$, allows to exclude non-optimal solutions at the early stages of design or in the technological process. To solve this problem, it is necessary to build a model of the dependence of the vector of output parameters $\mathbf{Y} = \{y_1, y_2, \ldots y_m\}^T$ from the input parameters and on its basis to solve the problem of determining the optimum values of the output parameters taking into account the available technological (constructive) constraints [1].

Then, for the obtained optimal values of the output parameters, it is necessary to determine the corresponding values of the input parameters (the inverse optimization problem).

2. Definition of the regions of admissible values of output parameters

The region $\mathbf{N}$ of changing of the input parameters is determined by the physical factors and is usually given by a system of inequalities in the form of $x_{i1} \leq x_i \leq x_{i2}, i = \overline{1,k}$. .The region $\mathbf{M}$ of admissible values of output parameters (parameters-acceptance criteria) $\mathbf{Y} = \{y_1, y_2, \ldots y_m\}^T$ is given by the normative values of these parameters in the form of a system of inequalities $y_{j1} \leq y_j \leq y_{j2}$. $j = \overline{1,m}$. Vector $\mathbf{X} = \{x_1, x_2, \ldots x_k\}$ in the space of the input technological parameters determines the point in the space of the parameters- acceptance criteria corresponding to the given article.

If the transformation $\mathbf{X} \Rightarrow \mathbf{Y}$ is known in the form of a multiresponse function $\mathbf{Y} = \mathbf{F}(\mathbf{X},\mathbf{B})$, where $\mathbf{F}(\mathbf{X},\mathbf{B})$ – multiresponse function, which is determined experimentally, and $\mathbf{B}$ is a coefficient vector, then for a given point $\mathbf{X}$ in the space of input parameters it is possible to determine the corresponding point $\mathbf{Y}$ in the space of output parameters. Thus, a k-dimensional region $\mathbf{N}$ ($k \geq m$) is represented into the space of output parameters in the form of a m-dimensional region $\mathbf{T}$.

*S.A. Popov. Email:stanislav.popov@novsu.ru

There are, also, functional limitations, determined by the principles of operation of the designed equipment or technological process.

These constraints are represented as a system of inequalities $G(Y) \leq 0$ and they define the region **K** for which these constraints are satisfied [2].

If the point **X** corresponds to a point **Y** inside the region of nondefective items **M** ($Y \in M$), then this product meets the requirements of the technical specifications for these parameters Taking into account functional limitations, the desired region **S** of output parameters is determined by the intersection of the regions **T**, **M**, **K**, i.e. **S** = **T**∪**M**∪**K** (Fig. 1).
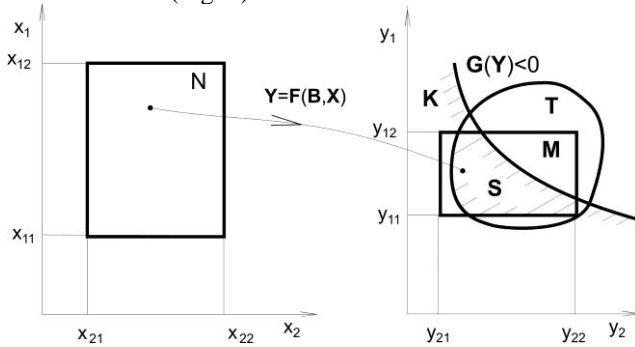

Fig. 1. Determination of the region S of admissible values of output parameters

3. Solution of the inverse optimization problem with incontrollable input parameters

The task of determining the admissible input parameters is the determination of such region of input parameters which ensures the belonging of the given article to the class of nondefective ones, which is the content of the problem of reverse optimization. If the image **L** of the region **S** is defined in the space of input parameters, then it is possible to specify the technological parameters $X \in$ **L**, ensuring certainly nondefective articles (Fig. 2). To specify independent intervals of the admissible values of technological parameters in the region **L**, it is necessary to inscribe the *k*-dimensional parallelepiped **R** in accordance with the criterion of optimality.

If the input parameters are uncontrollable and have scatted values (Fig. 2), then the percentage of yield is used as a criterion for the optimality of the intervals of admissible values.
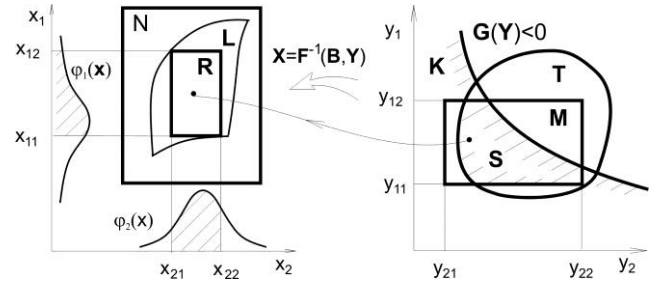

Fig. 2. Obtaining independent intervals of admissible values for input technological parameters

In this case, the optimal region **R** is constructed on the basis of the joint probability density function of the input parameters $\varphi(x_i)$. The criterion of optimality D for determining the intervals of admissible values is the yield of nondefective articles by technological parameters in the form

$$D = \int_{x_{11}}^{x_{12}} \int_{x_{21}}^{x_{22}} \cdots \int_{x_{k1}}^{x_{k2}} \varphi(x_1, x_2, \ldots, x_k) dx_1 dx_2 \ldots dx_k , \qquad (1)$$

where $\varphi(x_1, x_2, \ldots, x_k)$ is the joint probability density function of technological parameters.

The construction of the region **R** in accordance with the criterion of optimality (1) is performed by the method of statistical tests.

4. Construction of a multiresponse model and its use for determining input parameters

The $X \Rightarrow Y$ transformation is represented as a multiresponse model

$$Y = F(X,B), \qquad (2)$$

where $F(X,B) = \{f_1(X,B), f_2(X,B), \ldots, f_m(X,B)\}^T$ is m-dimensional vector of the functions, $Y = \{y_1, y_2, \ldots y_m\}^T$ are admittance/reject criterion (responses), $X = \{x_1, x_2, \ldots x_k\}^T$ are independent variables, $B = \{b_1, b_2, \ldots b_l\}^T$ is l-dimensional vector of coefficients whose exact values are determined from the experimental data.

In the general case, the calculation of the estimates of the coefficients of the model (2) is carried out using an iterative procedure [3]

$$B^{s+1} = B^s + K \sum_{j=1}^{n} P(X_j, B^s) V_E^{-1} [Y_j - F(X_j, B^s)], \qquad (3)$$

where matrix V_E is covariance matrix of observation errors and

$$P(X_j) = \left\{ \frac{\partial f_1(X_j, B)}{\partial B}, \frac{\partial f_2(X_j, B)}{\partial B}, \cdots, \frac{\partial f_m(X_j, B)}{\partial B} \right\}.$$

The covariance matrix for estimating the coefficients (3) is

$$\mathbf{K} = \left[\sum_{j=1}^{n} \mathbf{P}(\mathbf{X}_j) \mathbf{S}_{\mathbf{E}}^{-1} \mathbf{P}^{\mathrm{T}}(\mathbf{X}_j) \right]^{-1}, \qquad (4)$$

where $\mathbf{S}_{\mathbf{E}}$ is covariance matrix estimate $\mathbf{V}_{\mathbf{E}}$.

The constructed model can be used to solve the inverse problem, that is, to estimate the values of the input parameters for given values of the output parameters. When the type of the model (2) is selected and the estimates of its coefficients (3) are calculated, the calculation of the estimates of the input values of the parameters is performed by the formula

$$\mathbf{X} = \mathbf{F}^{-1}(\mathbf{B}, \mathbf{Y}), \qquad (5)$$

where $\mathbf{F}^{-1}(\cdot)$ is the function inverse to the function $\mathbf{F}(\cdot)$, $\mathbf{Y}$ is the vector of the output parameters.

Expanding expression (2) in terms of independent variables in a Taylor series, we obtain

$$\Delta \mathbf{Y} \approx \Omega(\mathbf{B}, \mathbf{X}) \Delta \mathbf{X}, \qquad (6)$$

where $\Omega^{\mathrm{T}} = \left\{ \dfrac{\partial \mathrm{f}_1(\mathbf{B}, \mathbf{X})}{\partial \mathbf{X}}, \dfrac{\partial \mathrm{f}_2(\mathbf{B}, \mathbf{X})}{\partial \mathbf{X}}, \ldots, \dfrac{\partial \mathrm{f}_m(\mathbf{B}, \mathbf{X})}{\partial \mathbf{X}} \right\}.$

Then, in the case of the equality of the number of independent and dependent variables, the solution of the nonlinear problem (6) can be obtained by the following iterative procedure

$$\mathbf{X}^{s+1} = \mathbf{X}^s + (\Omega)^{-1} \Delta^s, \qquad (7)$$

where s is iteration number, $\Delta = \mathbf{Y} - \mathbf{F}(\mathbf{B}, \mathbf{X})$.

The covariance matrix $\mathbf{V}_{\mathbf{X}}$ of the estimates $\mathbf{X}$ is derived from expression (6) and is equal to

$$\mathbf{V}_{\mathbf{X}} \approx \Omega^{-1} \mathbf{P}^{\mathrm{T}}(\mathbf{B}, \mathbf{X}) \mathbf{V}_{\mathbf{B}} \mathbf{P}(\mathbf{B}, \mathbf{X}) (\Omega^{-1})^{\mathrm{T}}. \qquad (8)$$

If the number of input parameters k is greater than the number of output parameters m, then for solving the equation (5) it is necessary to specify (k-m) input parameters, and the remaining m input parameters are determined according to equation (7)

When constructing the model (2), it is necessary to perform:

(i) analysis of the covariance matrix of observation errors;
(ii) the choice and justification of the model;
(iii) checking the significance of coefficient estimates;
(iv) checking the adequacy of the model.

4. Approximation of the admissible set of input parameters

As a result of analyzing the matrix $\mathbf{W}$, the region of admissible values of the input parameters is presented in a discrete form and its boundaries are not defined analytically. Often in practice, this form of the region of permissible values leads to problems when trying to analyze it and use it for decision making. For example, when a vector of input parameters is set on the boundary

of an admissible region, it is not clear whether it enters the region of admissible values.

To solve the approximation problem, the form of the region of admissible values should be approximated by some analytic surface. This convolution of the region of admissible values greatly simplifies its use. For example, it allows to easily set the desired input parameters within this region. To identify the shape of a admissible region, it is possible to construct it in the form of a separating surface that includes all (or at least the majority) values of the input parameters related to the admissible region and does not include parameters that are not within the admissible region.

In these circumstances, it is proposed to use the method of least squares to construct a separating surface [5].]. In accordance with this method, estimates of the coefficients $\mathbf{A}$ of the separating surface $\mathrm{g}(\mathbf{X}, \mathbf{A}) = 0$ are calculated so that they provide a minimum of the sum of squares of deviations in the form

$$\sum_{i=1}^{n} [\mathrm{g}(\mathbf{X}_i, \mathbf{A}) - y_i]^2 = \min_{\mathbf{B}} \left\{ \sum_{i=1}^{n} [\mathrm{g}(\mathbf{X}_i, \mathbf{A}) - y_i]^2 \right\}, \qquad (9)$$

where $\mathrm{g}(\mathbf{X}, \mathbf{A})$ is separating function, $\mathbf{X} = \{x_1, x_2, \ldots, x_r\}^{\mathrm{T}}$ is vector of parameters, $\mathbf{A} = \{a_1, a_2, \ldots, a_l\}^{\mathrm{T}}$ is vector of coefficients, y is the attribute of the point of the admissible region (for example, $y = 1$ if the parameter enters the admissible region, and $y = -1$ otherwise), n is the sample size.

However, with this approach, the contribution to the sum of the deviations (9) will be given by parameters with large values of the function $\mathrm{g}(\mathbf{X}, \mathbf{A})$, which leads to significant recognition errors, especially in the complex form of the separating function, which is typical for technical applications.

We assume that for each parameter there is a probability of belonging to an admissible region q_1 and the probability does not belong to this region q_2 ($\mathrm{q}_1 + \mathrm{q}_2 = 1$). Experimentally belonging to the admissible region ($y = 1$) is determined from the matrix $\mathbf{W}$ (the parameters enter the matrix $\mathbf{W}$).

This gives an estimate y of the value of the membership function $\mathrm{q}(\mathbf{X}) = \mathrm{q}_1 - \mathrm{q}_2$ as a function of the magnitude of the parameter vector.

As a result of the construction of an admissible region for n items, the estimation of the membership vector $\mathbf{Y} = \{y_1, y_2, \ldots, y_n\}^{\mathrm{T}}$ is determined. Approximation of the dependence of the magnitude q on the values of the parameters can be presented in the form [6]:

$$\mathrm{q}(\mathbf{X}, \mathbf{B}) = \mathrm{sign}[\mathrm{f}(\mathbf{X}, \mathbf{A})]\{1 - \exp[-|\mathrm{f}(\mathbf{X}, \mathbf{A})|]\}, \qquad (10)$$

where $\mathrm{f}(\mathbf{X}, \mathbf{A})$ is some approximation of function $\mathrm{g}(\mathbf{X}, \mathbf{A})$.

The function (10) is equal to zero at the interface between two classes (belonging to the range of admissible values or non-belonging) and tends to 1 (or -1) when moving

away from the separating surface. The separating surface is determined by the equation

$$g(\mathbf{X}, \mathbf{A}) = 0 . \qquad (11)$$

In this case, the calculation of the coefficient estimates in expression (11) by the least squares method is performed using the following iterative procedure [6]:

$$\mathbf{A}^{s+1} = \mathbf{A}^s + \mathbf{V_A} \sum_{i=1}^{n} \mathbf{P}(\mathbf{X}_i, \mathbf{A}^s)[y_i - \mathbf{q}(\mathbf{X}_i, \mathbf{A})], \qquad (12)$$

where s is iteration number and

$$\mathbf{P}(\mathbf{X}, \mathbf{A}) = \text{sign}[f(\mathbf{X}, \mathbf{A})]\exp[-|f(\mathbf{X}, \mathbf{A})|]\frac{\partial f(\mathbf{X}, \mathbf{A})}{\partial \mathbf{A}} .$$

For a complete quadratic function with three parameters, for example, one can obtain $\frac{\partial f(\mathbf{X}, \mathbf{A})}{\partial \mathbf{A}} = \left\{ 1, x_1, x_2, x_3, x_1 x_2, x_1 x_3, \ldots, x_3^2 \right\}^{\mathrm{T}} .$

Covariance matrix for estimating coefficients $\mathbf{A}$ is calculated by the formula

$$\mathbf{V_A} = \sigma_e^2 \left[\sum_{i=1}^{n} \mathbf{P}(\mathbf{X}_i, \mathbf{A})\mathbf{P}^{\mathrm{T}}(\mathbf{X}_i, \mathbf{A}) \right]^{-1}, \qquad (13)$$

where σ_e^2 is variance of class recognition error.

As an estimate of the variance in the recognition of classes, we used the estimate of the variance of the model's adequacy error, calculated by the formula:

$$s_e^2 = \frac{1}{n-l}\sum_{i=1}^{n}[q(\mathbf{X}_i, \mathbf{A}) - y_i]^2 , \qquad (14)$$

where l is number of coefficients in the model (11).

The diagonal elements of the matrix $\mathbf{V_B}$ (13) represent the variances of the estimates of the corresponding coefficients, which makes it possible to calculate their t-statistics and, therefore, to verify the significance of the coefficient estimates.

The lower bound of the one-sided confidence interval for the separating function that ensures the probability of recognizing the first class of at least P% is defined as:

$$q(\mathbf{X}, \mathbf{A}) = q(\mathbf{X}, \mathbf{A}) + t_{P,f} s_e , \qquad (15)$$

where t_P is Student's quantile for confidence probability P and number of degrees of freedom f, $\mathbf{A}$ is coefficient estimator vector.

Then the separating surface is defined by the expression

$$q(\mathbf{X}, \mathbf{A}) + t_{P,f} s_e = 0 . \qquad (16)$$

By changing confidence probability $\mathbf{P}$, one can achieve such a "compression" of region $g(\mathbf{X}, \mathbf{A})$ so that it does not include points that do not belong to the admissible region $\mathbf{D}$. In this case, the resulting surface will be inscribed into the region $\mathbf{D}$ with some "reserve".

To calculate vector of estimates of the coefficients of separating function $\mathbf{A}$ in accordance with iteration procedure (12), a computer program has been developed. A two-dimensional illustration of the construction of approximating surface for three parameters is shown in Fig.3.
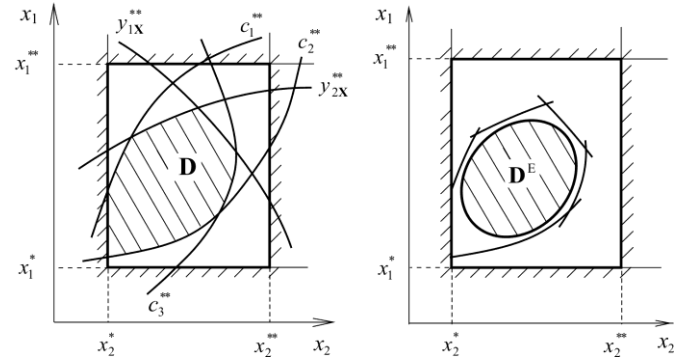

Fig. 3. Approximation of a two-parameter region of admissible values **D** by an inscribed ellipse **DE** with a given confidence probability

5. Construction of a discrete Pareto-optimal set

The Pareto-optimal set will be constructed on the basis of the matrix **E**, which determines the admissible set of input parameters **D**. The search for Pareto-optimal points is performed among the rows of the matrix **W** in accordance with the following conditions

The point $\mathbf{X}_i$ ($\mathbf{X}_i \notin \mathbf{D}$) is Pareto optimal if there is no such point $\mathbf{X} \in \mathbf{D}$ that $f_v(\mathbf{X}) \le f_v(\mathbf{X}_i)$ for all $v = \overline{1, m+k}$ and at least for one v exists $f_v(\mathbf{X}) < f_v(\mathbf{X}_i)$. A set $\mathbf{P} \subset \mathbf{D}$ is called Pareto-optimal if it consists of all Pareto optimal points. Based on the analysis of the Pareto-optimal set, the decision-maker determines the most preferable option $\mathbf{X}^0$.

Pareto set is important for multicriterion optimization tasks, because, firstly, it is easier for designer to analyze it than the entire admissible set, secondly, whatever system of preferences designer used when comparing different vectors from the admissible region, the optimal vector always belongs to the Pareto set. It is known [1] that if admissible set **D** is closed and criteria $f_v(\mathbf{X})$ are continuous, then Pareto set is not empty. This means that in any project task, we must determine the set of Pareto-optimal solutions.

The construction of Pareto-optimal set begins after obtaining admissible set **D** in the space of input parameters in the form of a matrix W. The first r columns of this matrix form discrete admissible set **D** in the space of input parameters containing s points.

The following m columns form the image of set **D** in space of parameters- acceptance criteria $\mathbf{F}(\mathbf{D})$. At the request of designer, it is possible to include in optimization process also k parameters that are not

parameters- acceptance criteria and for which technical limitations are not specified, but which are desirable to be optimized.

For each point $\mathbf{X}_i$ all local criteria are calculated $f_v(\mathbf{X}_i)$, $v = \overline{1, m+k}$. Thus, for each test point, a line is obtained in test table $\mathbf{W}$, which has dimensions $[n \times (m+k)]$ in the form of:

$$\mathbf{W} = \left\{ \begin{array}{ccccccc} x_{11} & \cdots & x_{1r} & f_{1,m+1}(\mathbf{X}_1) & \cdots & f_{1,m+k}(\mathbf{X}_1) \\ x_{21} & \cdots & x_{2r} & f_{2,m+1}(\mathbf{X}_2) & \cdots & f_{2,m+k}(\mathbf{X}_2) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & \cdots & x_{nr} & f_{n,m+1}(\mathbf{X}_n) & \cdots & f_{n,m+k}(\mathbf{X}_n) \end{array} \right\}, \quad (17)$$

where n is number of test points, m is number of parameters-acceptance criteria, k is number of controlled output parameters that are not constrained.

By checking the conditions for a point to belong to Pareto-optimal set for all points of the matrix $\mathbf{W}$ (17), only the Pareto-optimal points remain in this table.

Consider simplified illustration of constructing Pareto-optimal solution set (the criteria should be minimized). In Fig. 4 Pareto-optimal set in the space of criteria is represented by solid line confined by points $\mathbf{A}$ and $\mathbf{B}$.
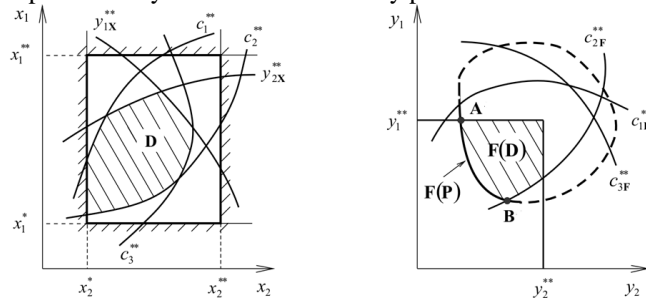

Fig. 4. The Pareto set $\mathbf{F(P)}$ obtained with functional and criterial constraints

Conclusion

The proposed engineering method for constructing an admissible set of parameters allows one to obtain an admissible set in an analytical form, which makes it possible to provide an admissible probability of an error of its construction. The Pareto-optimal set is constructed in discrete form with a specified accuracy. The optimization procedure is based on a multiresponse model, the construction of which is the first stage of multi-criteria optimization.

References

[1] Sobol, I. and , Statnikov, R. (2006) The choice of optimal parameters in problems with many criteria. M.: Drofa.
[2] Fyodorov, V. (1971) Theory of optimal experiment. M.: Nauka.
[3] Podinovsky, V. and Nogin, V. (2007) Pareto-optimal solutions of multicriteria problems. M.: Fizmatlit.
[4] Fletcher, R. (1980) Practical Methods of Optimization N.Y.: John Wiley and Sons.
[5] Popov,S (2010) Construction and use of a separating surface for determining the composition of a multicomponent mixture. Factory Laboratory. Diagnosis of materials vol.: 4, p. 40-42
[6] Popov,S. and Galaktionov, N. (2010) Optimization of design parameters based on the multi-response model Vestnik of NovSU Vol. 61: p. 43-46.