Research in Engineering Design (2005) 15: 201–215
DOI 10.1007/s00163-004-0054-8

ORIGINAL PAPER

Terrence E. Murphy · Kwok-Leung Tsui
Janet K. Allen

A review of robust design methods for multiple responses

Received: 1 May 2003 / Revised: 28 July 2004 / Accepted: 28 July 2004 / Published online: 25 September 2004
© Springer-Verlag London Limited 2004

Abstract Problems in engineering design often involve determining design variable settings to optimize individual product performance for multiple criteria, which are often in conflict. We review mathematically rigorous techniques from the statistical literature for finding a vector **x** of design variable settings, which produces an optimal compromise solution among a group of prioritized response variables. The best compromise solution is typically gained by optimizing an objective function, which incorporates the prioritized demands of multiple responses. Since most multi-response objective functions are constructed by combining the functions used to optimize univariate responses, a review of the prominent univariate approaches is presented first. A multivariate approach from the engineering literature called the compromise decision support problem (cDSP) is also reviewed. Finally, a table comparing the relative merits of the different multivariate approaches summarizes the article in a concise and user-friendly fashion.

Keywords Optimal design · Robust engineering design · Multiple response · Loss function

1 Frame of reference

One of the most important roles for an engineering designer is to make decisions about products that are being designed. The value of most products hinges upon their ability to satisfy multiple functional criteria. Typically, a designer is asked to determine design variable settings to optimize product performance on these multiple criteria, which are often in conflict. In this article, we call these criteria quality characteristics or responses, and note that these are often of varying priority to the end user. This requires a design decision maker (DDM) to prioritize and/or assign quantitative importance measures to responses in order to make the best compromise choices.

Our approach is founded on the idea of bounded rationality, as proposed by Simon (1976). Unlike the ideal state commonly assumed for economic analyses, which assumes that a designer has perfect and complete information, bounded rationality more closely reflects the actual state of the world in which information is uncertain, incomplete, and complex, and where the number of potential courses of action is nearly infinite. If complete information were available, it would be possible to invoke a single objective design making in all circumstances, as recommended by Hazelrigg (1996) and others. However, in recognition of the fact that the world is often less than ideal, we assert that, in practice, rigorous methods for multi-criteria design decision making are essential.

In this paper, we explore the formulation of engineering design decisions in the context of the more general, mathematically rigorous techniques documented in the statistical literature for finding a vector **x** of variable settings to yield an optimal compromise solution among a group of prioritized response variables. We will examine several attributes of each approach, including how the correlation structure of the multiple responses is utilized in the optimization process.

Some techniques assume that the multiple responses are independent of each other. This implies that variation in one response is not related to variation in any other response. While this assumption brings mathematical simplification to statistical analysis, it is not the reality of most design situations. For example, customer criteria for a car include fuel efficiency, cost, reliability,

T. E. Murphy (✉) · K.-L. Tsui
School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, Georgia, USA
E-mail: tmurphy@isye.gatech.edu
Tel.: +1-404-5082416
Fax: +1-404-8941608

J. K. Allen
The G.W. Woodruff School of Mechanical Engineering,
Georgia Institute of Technology,
Atlanta, Georgia, USA

maneuverability, capacity, vehicle weight, and driving comfort. These are clearly correlated since higher vehicle weight typically implies both higher capacity and lower fuel efficiency. Hence, the assumption of response independence is an impracticality in many design problems.

Other techniques actively exploit the correlations of the responses as a source of information while searching for optimal design parameters. This is a statistical advantage since an additional source of response information, i.e., correlation structure, is being put to use. Lastly, there are techniques which, while not actively harnessing the correlation information, are not hampered by an assumed independence of the responses. Regardless of how the response correlation structure is employed, all techniques examined assume that ordering and weighting of responses are carried out by a single DDM and are transitive.

One of the commonly cited examples in the statistical literature is that of manufacturing a beef stew military field ration, as detailed in Contreras et al. (1995). In this case, there are two important quality characteristics (i.e., responses); namely, heating rate index and the lethality index. The heating rate index is the rate at which the product may be brought to sterilizing temperature and the lethality index is an indicator of microbiological safety. The five design variables include sauce viscosity, residual gas, solid-to-liquid ratio, net weight, and speed of rotation of the food pouch during the heating process.

A DDM wants to choose the settings of these five variables so that the heating rate is as fast as possible, since this expedites manufacturing, and so that the lethality index stays above a certain minimum to guarantee consumer safety. Furthermore, the DDM wants to minimize how far the lethality index rises above the required safety level since flavor deteriorates as this index rises. This last requirement is an example of a constraint within the multi-response optimization.

Even for the simple case of only two responses, the statistically based methods typically employ an objective function incorporating the relative importance of the two responses. The design goal is to identify the specific design variable settings that optimize the objective function. In most statistically based methods, the key to finding these optimal design parameters is choosing the appropriate objective function for the design situation at hand.

Important considerations when choosing the multi-response optimization approach include how many and which type of individual responses are handled, how they are weighted, the type of modeling used to represent individual responses or the objective function, the number of responses reasonably managed by the objective function, and the specific optimization techniques which complement that function.

Since most statistically based multi-response objective functions are formed by combining objective functions used to optimize single responses, in Sect. 2, we review the common statistical techniques for optimizing a single response. In Sect. 3, we compare the different multi-response objective functions formed by additive and multiplicative combination of the univariate objective functions. In Sect. 4, we review the compromise decision support problem (cDSP), a hybrid formulation incorporating concepts from both traditional mathematical programming and goal programming, which is used in engineering sectors for solving the multi-response optimization problem. In Sect. 5, we compare the different multi-response techniques with respect to a number of important metrics, including their ability to manage constraints and how the optimal solutions are affected by shifts in target or specification.

2 Single response optimization from the statistics literature

A common criterion for univariate response optimization is to find the design variable settings which produce a response that is both on-target and of sufficiently low variability. While the ideal univariate response is simultaneously on-target and of minimal variance, reality usually forces the DDM to make trade-offs between on-target performance and low variance. In the first subsection, we examine two methods for focusing on the interplay between the location and variation of the system response:

1. Loss functions, especially the two-step strategies (Taguchi 1986)
2. The dual response method

In the second subsection, we define utility functions, i.e., objective functions measuring the positive value, or worth, of a set of design variable settings, and their use in univariate optimization. The majority of the statistical literature dealing with univariate response optimization consists of techniques built around loss or utility functions.

2.1 The loss function approach

In this subsection, we examine two different ways of focusing on the interplay between the location and variation of the system performance. The first of these is the use of loss functions with emphasis on the two-step strategies (Taguchi 1986), which exploit the decomposition of the expected value of the squared-error loss function. The second method is a strategy called the dual response method (Vining and Myers 1990), which optimizes either location or variance while constraining the other.

2.1.1 Commonly used loss functions

There are a few standard loss functions which are commonly used to evaluate process function. Among these standard losses, the squared-error loss function is the most important and is defined as follows:

$$L(Y,\ t) = (Y - t)^2 \tag{1}$$

where Y represents the actual process response and t is the targeted value. A loss occurs if the response Y deviates from its target t. This loss function originally became popular in estimation problems considering unbiased estimators of unknown parameters (Berger 1995). The expected value of $(Y-t)^2$ can be easily expressed as:

$$\begin{aligned} \mathrm{E}(L) &= A_0 \mathrm{E}(Y - t)^2 \\ &= A_0 \left[\mathrm{VAR}(Y) + (\mathrm{E}(Y) - t)^2 \right] \end{aligned} \tag{2}$$

where VAR(Y) and E(Y) are the mean and variance of the process response, respectively, and A_0 is a proportional constant representing the economic costs of the squared error loss. If E(Y) is on target, then the squared-error loss function reduces to the process variance. Its similarity to the criterion of least squares in estimation problems makes the squared-error loss function easy for statisticians and engineers to grasp. Furthermore, the calculations for most decision analyses based on squared-error loss are straightforward and easily seen as a trade-off between variance and the square of the off-target factor.

To provide decision makers with flexible weighting of the off-target squared and variance components, Box and Jones (1990) introduced the following general class of squared-error loss functions:

$$\mathrm{E}(L) = A_0 \left[\pi(\mathrm{E}(Y) - t)^2 + (1 - \pi)\mathrm{VAR}(Y) \right] \tag{3}$$

where $0 \le \pi \le 1$.

Note that the loss function shown in Eq. 1 is a special case of this loss function where:

$$\pi = (1 - \pi) = 0.5$$

which places equal weight on the squared off-target and variance components. The DDM wishes to identify the vector of design variable settings that minimizes the expected value of the loss function.

2.1.2 Optimization with two-step procedures

Taguchi (1986) introduced robust parameter design, a method for designing processes that are robust (i.e., insensitive) to uncontrollable variation, to major American corporations. The objective of this methodology is to find the settings of design variables that minimize the expected value of squared-error loss, as defined in Eq. 2.

Robust design (RD) assumes that the appropriate performance measure i.e., Y can be modeled as a transfer function of the fixed control variables and the random noise variables of the process as follows:

$$Y = f(\mathbf{x},\ \mathbf{N},\ \theta) + \varepsilon \tag{4}$$

where $\mathbf{x} = (x_1, \ldots, x_p)^{\mathrm{T}}$ is the vector of control factors, $\mathbf{N} = (N_1, \ldots, N_q)^{\mathrm{T}}$ is the vector of noise factors, θ is the vector of unknown response model parameters, and f is the transfer function for Y. The control factors are assumed to be fixed and represent the fixed design variables. The noise factors $\mathbf{N}$ are assumed to be random and represent the uncontrolled sources of variability in production. The pure error ϵ represents the remaining variability that is not captured by the noise factors, and is assumed to be normally distributed with zero mean and finite variance.

Taguchi (1986) divides the design variables, i.e., control variables, into two subsets, $\mathbf{x} = (\mathbf{x}_{\mathrm{a}}, \mathbf{x}_{\mathrm{d}})$, where $\mathbf{x}_{\mathrm{a}}$ and $\mathbf{x}_{\mathrm{d}}$ are called, respectively, the adjustment and the non-adjustment design factors. An adjustment factor influences process location while remaining effectively independent of process variation. A non-adjustment factor influences process variation.

Taguchi (1986) also introduced a family of performance measures called signal-to-noise ratios (SNR), whose specific form depends on the desired response outcome. The case where the response has a fixed non-zero target is called the nominal-the-best (NTB) case. Likewise, the cases where the response has a smaller-the-better target or a larger-the-better target are, respectively, called the STB and LTB cases. For these three cases, Taguchi defined the SNR as follows:

$$\mathrm{SNR} = \begin{cases} 10 \log_{10} \left[\frac{\bar{Y}^2}{s_{\bar{Y}}^2} \right] & \text{for the NTB case} \\ -10 \log_{10} \left[\frac{1}{n} \sum_{j=1}^{n} Y_j^2 \right] & \text{for the STB case} \\ -10 \log_{10} \left[\frac{1}{n} \sum_{j=1}^{n} \left(\frac{1}{Y_j} \right)^2 \right] & \text{for the LTB case} \end{cases} \tag{5}$$

where $\bar{Y}$ and $s_{\bar{Y}}^2$ are, respectively, the sample mean and variance of the response estimated at each test combination of design variables selected for experimentation. The Y_j are the individual response observations out of a total of n.

The concept of the SNR originated in electrical engineering to quantify a circuit's ability to discern a meaningful signal from the noise invariably transmitted with it. The SNR is best demonstrated by the NTB case of Eq. 5, which is the logarithm of the ratio of the sample response mean (i.e., signal) squared to the sample response variance (i.e., noise). The SNR is the most controversial part of Taguchi's (1986) methods because it combines location and dispersion in a single performance measure. Other methods examine mean and variance as separate performance measures.

To accomplish the objective of minimal expected squared-error loss for the NTB case, Taguchi (1986) proposes the following two-step optimization procedure:

1. Calculate and model the SNRs and find the non-adjustment factor settings, i.e., $\mathbf{x}_{\mathrm{d}}$ which maximize the SNR

2. Shift mean response to the target by changing the adjustment factor(s) $\mathbf{x}_a$

For the STB and LTB cases, Taguchi (1986) recommends directly searching for the values of the design vector $\mathbf{x}$ which maximize the respective SNRs. An alternative to Taguchi's optimization approaches for these cases is proposed by Tsui and Li (1994).

In general, Taguchi (1986) gave no justification for how this use of the SNR would achieve the stated goal of minimal average squared-error loss. Leon et al. (1987) defined a function called the performance measure independent of adjustment (PerMIA), which justified the use of a two-step optimization procedure. They also showed that Taguchi's (1986) SNR for the NTB case is a PerMIA when an adjustment factor exists, and when the process response transfer function is of a specific multiplicative form. When Taguchi's (1986) SNR complies with the properties of a PerMIA, his two-step procedure minimizes the squared-error loss.

Leon et al. (1987) also emphasized two major advantages of the two-step procedure:

- It reduces the dimension of the original optimization problem
- It does not require re-optimization for future changes of the target value

There is an extensive body of statistics literature which examines the methods of Taguchi (1987). The most often cited include Kackar (1985), Nair (1986), Box (1988), Box et al. (1988), Nair (1992), and Tsui (1996a).

2.1.3 Optimization with the dual response approach

The dual response approach individually models mean and variance, optimizing one while constraining the other. Vining and Myers (1990) applied the dual response approach to Taguchi's three static situations (as listed immediately below) with the added constraint $\mathbf{x}'\mathbf{x} = \varrho^2$, which restricts the search area to a spherical region of radius ϱ.

For the STB characteristic:

$$\begin{aligned} &\text{minimize } \hat{\mu} \\ &\text{subject to } \hat{\sigma}^2 \leqslant t \end{aligned} \tag{6}$$

For the LTB characteristic:

$$\begin{aligned} &\text{maximize } \hat{\mu} \\ &\text{subject to } \hat{\sigma}^2 \leqslant t \end{aligned} \tag{7}$$

For the NTB characteristic:

$$\begin{aligned} &\text{minimize } \hat{\sigma}^2 \\ &\text{subject to } \hat{\mu} = t \end{aligned} \tag{8}$$

del Castillo and Montgomery (1993) attacked the same problem by searching for the optimal values of $\mathbf{x}$ using the generalized reduced gradient (GRG), a commonly used non-linear programming primal algorithm. They cite the GRG's ability to consider more general forms of response surfaces and its ability to directly specify the radius of the spherical search region as advantages over the approach of Vining and Myers (1990).

Lin and Tu (1995) propose directly minimizing the weighted combination of the squared off-target factor and variance:

$$\text{minimize } \pi_1(\hat{\mu} - t)^2 + \pi_2\hat{\sigma}^2 \tag{9}$$

This approach allows the user to incorporate his/her preferences regarding the trade-off between the off-target and variance terms. This optimization criterion is easily applied to the NTB case, but finding an appropriate value for t is difficult in the STB and LTB cases. Tsui (1996b) pointed out that the minimization of this criterion is invariant to π_1 and π_2 if $\hat{\mu} = t$ is satisfied.

2.2 The utility function approach

A direct conceptual opposite of the loss function, a utility function maps a specific action (i.e., specific design variable settings) to an expected utility value (value or worth of a process response). Utility theory deals with the development of such functions based on a number of assumptions, such as the possibility of stating ordered preferences for the potential outcomes of a process. Berger (1995) details the four axioms which a utility function must satisfy. He requires that the utility function is bounded. However it is possible to develop a weaker set of axioms for unbounded utility functions. Once the utility function has been formulated, the DDM can employ non-linear direct search methods to find the vector of design variable settings that maximizes the desirability function.

Harrington (1965) introduced a univariate utility function called the desirability function, which gives a quality value between 0 (i.e., unacceptable quality) and 1 (i.e., further improvement would be of no value) of a quality characteristic of a product or process. He defined the two-sided desirability function as follows:

$$d_i = \mathrm{e}^{-|Y_i'|^c} \tag{10}$$

where e is the natural logarithm constant, c is a positive number subjectively chosen for scaling of the curve, and Y_i' is a linear transformation of the univariate response Y_i such that:

$$Y_i' = \frac{2Y_i - (Y_{\mathrm{USL}} + Y_{\mathrm{LSL}})}{(Y_{\mathrm{USL}} - Y_{\mathrm{LSL}})} \quad \text{for } \mathrm{LSL} \leqslant Y_i \leqslant \mathrm{USL} \tag{11}$$

where Y_i, Y_{USL}, and Y_{ISL} are the current value of Y and the upper and lower specification limits of Y, respectively. Y_i is a linear mapping of the within-specification Y_i values to values between -1 and $+1$. These properties of Y_i ensure that the desirability function has the following properties:

– Converges to $d_i = 0$ as $|Y_i|$ exceeds 1.0
– Passes through $d_i = e^{-1} = 0.37$ when $|Y_i| = 1.0$
– Passes through $d_i = e^0 = 1.0$ at the mid-specification pointIt is of special interest to note that, for $c = 2$, a mid-specification target, and response values within the specification limits, this desirability function can be expressed as the natural logarithm constant raised to the squared-error loss function with a specific proportional constant:

$$\begin{aligned} d_i &= e^{-|Y_i'|^2} \\ &= e^{-\left|\frac{2Y_i-(Y_{USL}+Y_{LSL})}{(Y_{USL}-Y_{LSL})}\right|^2} \\ &= e^{-4\left|\frac{Y_i-\left(\frac{Y_{USL}+Y_{LSL}}{2}\right)}{(Y_{USL}-Y_{LSL})}\right|^2} \end{aligned}$$

which, for the definition of $t_i = \frac{Y_{USL}+Y_{LSL}}{2}$, leads to the following expression:

$$d_i = e^{-\frac{4}{(Y_{USL}-Y_{LSL})^2}|Y_i-t_i|^2} \tag{12}$$

3 Multi-response optimization

In Sect. 2, we introduced loss and utility functions and showed how the relations between off-target and variance components underlie the loss function optimization strategies for single responses. Multi-response optimization typically combines the loss or utility functions of individual responses into a multivariate function to evaluate the sets of responses created by a particular set of design variable settings. The two subsections deal with the additive and multiplicative combination of loss and utility functions, respectively.

3.1 Additive combination of univariate loss functions

For univariate responses, the expected squared-error loss is a convenient way to evaluate the loss caused by deviation from target because of its decomposition into squared off-target and variance terms. A natural extension of this loss function to multiply correlated responses is the multivariate quadratic loss (MQL) function of the deviation vector $(\mathbf{Y}-\tau)$, where $\mathbf{Y} = (Y_1,\ldots,Y_r)^T$ and $\tau = (t_1,\ldots,t_r)^T$, i.e.:

$$\mathrm{MQL}(\mathbf{Y},\ \tau) = (\mathbf{Y}-\tau)^T\mathbf{A}(\mathbf{Y}-\tau) \tag{13}$$

where $\mathbf{A}$ is a positive definite constant matrix. The values of the constants in $\mathbf{A}$ are related to the costs of non-optimal design, such as the costs related to repairing and/or scrapping a non-compliant product. In general, the diagonal elements of $\mathbf{A}$ represent the weights of the r characteristics and the off-diagonal elements represent the costs related to pairs of responses being simultaneously off-target.

It can be shown that, if $\mathbf{Y}$ follows a multivariate normal distribution with mean vector $\mathrm{E}(\mathbf{Y})$ and covariance matrix $\Sigma_{\mathbf{Y}}$, the average (expected) loss can be written as:

$$\begin{aligned} \mathrm{E(MQL)} &= \mathrm{E}\left[(\mathbf{Y}-\tau)^T\mathbf{A}(\mathbf{Y}-\tau)\right] \\ &= \mathrm{trace}[\mathbf{A}\Sigma_{\mathbf{Y}}] + [\mathrm{E}(\mathbf{Y})-\tau]^T\mathbf{A}[\mathrm{E}(\mathbf{Y})-\tau] \end{aligned} \tag{14}$$

where the off-target vector product $[\mathrm{E}(\mathbf{Y})-\tau]^T\mathbf{A}[\mathrm{E}(\mathbf{Y})-\tau]$ and $\mathrm{trace}[\mathbf{A}\Sigma_{\mathbf{Y}}]$ are multivariate analogs to the squared off-target component and variance of the univariate squared-error loss function. Moving all response means to target simplifies the expected multivariate loss to the $\mathrm{trace}[\mathbf{A}\Sigma_{\mathbf{Y}}]$ term. The trace-covariance term shows how the values of $\mathbf{A}$ and the covariance matrix $\sum_{\mathbf{Y}}$ weight the individual responses within expected multivariate loss.

The simplest approach to solve the robust design problem is to apply algorithms to directly minimize the average loss function in Eq. 14. Since the mean vector and covariance matrix are usually unknown, Pignatiello (1993) suggests their estimation by the sample mean vector and sample covariance matrix, or a fitted model based on a sample of observations of the multivariate responses. This strategy of optimizing the MQL function directly employs the correlation structure of the responses in the trace component.

To demonstrate how this MQL function additively combines the individual loss functions, we look at the simplest multivariate case; that of two properly ordered responses.

Let $\mathbf{Y} = [Y_1,\ Y_2],\ \tau = [t_1,\ t_2]$, and $\mathbf{A} = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$.

Then:

$$\begin{pmatrix} Y_1 - t_1 \\ Y_2 - t_2 \end{pmatrix}^T \mathbf{A} \begin{pmatrix} Y_1 - t_1 \\ Y_2 - t_2 \end{pmatrix} = a_1(Y_1 - t_1)^2 + a_2(Y_2 - t_2)^2$$

For this simplest of cases, the MQL function is equivalent to adding the individual squared-error loss functions of each response. This MQL function becomes increasingly complex with larger numbers of responses and interactions, indicated by non-zero terms in the off-diagonal elements of $\mathbf{A}$. We repeat the same example with a non-zero off-diagonal element.

Let $\mathbf{Y} = [Y_1,\ Y_2],\ \tau = [t_1,\ t_2]$, and $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$.

Then:

$$\begin{aligned} \begin{pmatrix} Y_1 - t_1 \\ Y_2 - t_2 \end{pmatrix}^T \mathbf{A} \begin{pmatrix} Y_1 - t_1 \\ Y_2 - t_2 \end{pmatrix} &= a_{11}(Y_1 - t_1)^2 \\ &\quad + 2a_{12}(Y_1 - t_1)(Y_2 - t_2) \\ &\quad + a_{22}(Y_2 - t_2)^2 \end{aligned}$$

3.1.1 The Mahalanobis distance

Khuri and Conlon (1981) propose an algorithm for the optimization of a multi-response system which seeks the

group of design settings that minimizes the Euclidean distance from a vector of idealized responses, i.e., the Mahalanobis distance (MD). The MD is a function of the estimated responses and their covariance structure.

Their procedure assumes that all response functions depend on the same set of design variables and can be represented by polynomial regression models of the same degree within the region of interest. They reduce the multiple responses to a linearly independent subset and calculate least squares estimates for these responses from the multi-response data set.

They express the r linearly independent response functions in the following multivariate form:

$$\Xi = \mathbf{X}\Theta + \varepsilon \tag{15}$$

where $\Xi = [\mathbf{Y}_1,..., \mathbf{Y}_r]$ is the $n \times r$ matrix consisting of the r column vectors corresponding to the n observations of each of the responses. For each of the n input vectors, there are r different response values which make up this $n \times r$ multi-response data matrix.

$\mathbf{X}$ is the $n \times p$ full column rank matrix consisting of n rows (i.e., $\mathbf{x}_1,..., \mathbf{x}_n)^T$. Each row contains the p model terms consisting of the union of all polynomial model terms from the r different response models. These same p model terms, however, take on different values depending on the design vector corresponding to each row.

$\Theta = [\theta_1, \theta_2,..., \theta_r]$ is the $p \times r$ matrix consisting of r column vectors, each having p model parameters, and $\epsilon = [\epsilon_1,..., \epsilon_r]$ is the $n \times r$ matrix consisting of the r column vectors corresponding to the error terms of the response models. The usual assumptions are that the rows of ϵ are statistically mutually independent, each having a zero mean vector and a common covariance matrix Σ. An unbiased estimate of the covariance matrix (i.e., $\hat{\Sigma}$) is typically used.

Each individual response value, i.e., $Y_i(\mathbf{x_j})$, where Y_i indicates a specific response variable and $\mathbf{x_j}$ is a specific design vector, can be modeled by the following polynomial equation of degree g:

$$\hat{Y}_i(\mathbf{x_j}) = \mathbf{z}_\mathbf{j}^T(\mathbf{x_j})\hat{\theta}_i \tag{16}$$

where $\mathbf{z}_\mathbf{j}^T(\mathbf{x_j})$ is the single row vector of dimension p from the $\mathbf{X}$ matrix of Eq. 15 corresponding to $\mathbf{x_j}$.

Khuri and Conlon (1981) recommend the following distance measure, i.e., the MD:

$$\mathrm{MD}\left[\hat{\mathbf{Y}}(\mathbf{x_j}),\ \tau\right] = \left[\frac{\left(\hat{\mathbf{Y}}(\mathbf{x_j}) - \tau\right)^T \Sigma^{-1}\left(\hat{\mathbf{Y}}(\mathbf{x_j}) - \tau\right)}{\mathbf{z}_\mathbf{j}^T(\mathbf{x_j})\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{z_j}\mathbf{x_j}}\right]^{1/2} \tag{17}$$

where $\hat{\mathbf{Y}}(\mathbf{x_j}) = \left[\hat{Y}_1(\mathbf{x_j}), \ldots, \hat{Y}_r(\mathbf{x_j})\right]$ is the vector of estimated responses from a particular design vector $\mathbf{x_j}$, $\tau^T = [\tau_1, \tau_2,..., \tau_r]$ is the vector of individual optimal responses and $(\mathbf{X}^T\mathbf{X})^{-1}$ is the inverse of the squared design matrix.

Khuri and Conlon's (1981) optimal solution is the vector of design variables $\mathbf{x_j}$ which minimizes the distance measure $\mathrm{MD}\left[\hat{\mathbf{Y}}(\mathbf{x_j}),\ \tau\right]$. For the case of potential fluctuation around the idealized response values, they propose a procedure for finding control variable settings which produce a min–max solution for the distance metric involving a modified version of the same distance measure. Since this procedure only models the subset of linearly independent responses, it does not extract all the statistical information available.

3.1.2 Additive formation of multivariate loss functions

In this section, we briefly review the literature for examples of multivariate functions formed by the additive combination of univariate loss and utility functions. We list the cases in order of increasing complexity.

Kumar et al. (2000) suggest creating a multi-response utility function as the additive combination of utility functions from the individual responses. If Y_i is the value of response i, each response has utility function $P_i(Yi)$ and the overall utility function is defined as:

$$P(Y_1, \ldots, Y_r) = \sum_{i=1}^{r} \omega_i P_i(Y_i) \tag{18}$$

where ω_i is the weight of each response and $\sum_{i=1}^{r} \omega_i = 1$. Here, the goal is to find the set of design variable settings that maximizes the overall utility function.

For cases where the target is the mid-specification point, Artiles-Leon (1996) proposes standardizing the squared-error loss function with the following proportionality constant:

$$A_0 = \left[\frac{2}{\mathrm{USL}_i - \mathrm{LSL}_i}\right]^2 \tag{19}$$

where LSL_i and USL_i are the upper and lower specification limits for each Y_i, respectively. With this constant, the standardized squared-error loss (SLOSS) function for a single response can be written as:

$$\begin{aligned}\mathrm{SLOSS}(Y_i) &= \left[\tfrac{2}{\mathrm{USL}_i - \mathrm{LSL}_i}\right]^2 (Y_i - t_i)^2 \\ &= 4\left[\tfrac{Y_i - t_i}{\mathrm{USL}_i - \mathrm{LSL}_i}\right]^2\end{aligned} \tag{20}$$

This standardized loss takes the value of 0 at the target and the value of 1 at the specification limits. A multivariate loss function is constructed simply as the sum of these dimensionless standardized loss functions.

The total standard loss (TSLOSS) function corresponds to the vector of responses ($Y_1,..., Y_r$) and is defined as:

$$\mathrm{TSLOSS}(Y_1, \ldots, Y_r) = 4\sum_{i=1}^{r}\left[\frac{Y_i - t_i}{\mathrm{USL}_i - \mathrm{LSL}_i}\right]^2 \tag{21}$$

where t_i is the target value for each Y_i. Assuming that all responses are uncorrelated and equally weighted, the individual standardized loss functions are simply added.

Ames et al. (1997) proposed a multivariate loss function, the global quality loss (GQL) function as:

$$\mathrm{GQL}(Y_1, \ldots, Y_r) = \sum_{i=1}^{r} \omega_i (Y_i - t_i)^2 \tag{22}$$

where the squared-error loss function of uncorrelated responses (Y_1,..., Y_r) and target values (t_1,..., t_r) are weighted by the constants (ω_1,..., ω_r). The GQL is a simple addition of the squared error losses of the individual responses with the scaling constants representing response priority.

For the subset of responses for which quadratic response surface models exist (i.e., Y_1,..., Y_m), they define the global quality loss function of process (GQLP) as:

$$\mathrm{GQLP}(Y_1, \ldots, Y_m) = \sum_{i=1}^{m} \omega_i \left[Y_i(x_1, \ldots, x_p) - t_i + \varepsilon_i \right]^2 \tag{23}$$

where the $Y_i(\mathrm{X}_1,..., X_p)$ are the responses approximated by quadratic response surface models and the ϵ_i are the residuals of the models. When the ϵ_i are small compared to the off-target factors, the GQLP can be defined as the quality loss function of process (QLP):

$$\mathrm{QLP}(Y_1, \ldots, Y_m) = \sum_{i=1}^{m} \omega_i \left[Y_i(x_1, \ldots, x_p) - t_i \right]^2 \tag{24}$$

Minimizing the QLP also minimizes the GQLP for the following two cases:

- ϵ_i are independent and of equal variance
- When off-target contributions to the loss function are significantly larger than the random error contributions

The first condition is a common assumption of response surface modeling and the second is a common characteristic in product development. In their photographic application, weighting is assigned by subjectively ranking the effect of a particular response being off-target.

3.1.3 Optimization of multivariate loss functions

For the expected MQL of Eq. 14, Pignatiello (1993) introduces a two-step procedure for finding the design variable settings that minimize this composite cost of poor quality. His procedure assumes that the responses follow a multivariate normal distribution, are NTB, and follow an additive model. His two-step procedure involves minimizing trace[$\mathbf{A}\Sigma_{\mathbf{Y}}$] and potentially adjusting the means of E($\mathbf{Y}$) to target.

Tsui (1999) extended Pignatiello's (1993) two-step procedure to situations where responses may be NTB, STB, or LTB. He divides the r responses into two subsets:

- $\mathbf{Y}_1 = (Y_1, \ldots, Y_{r_1})^{\mathrm{T}}$, i.e., those responses whose means can be adjusted to target
- $\mathbf{Y}_2 = (Y_{r_1+1}, \ldots, Y_r)^{\mathrm{T}}$, i.e., those responses whose means can't be adjusted to target

He defines a corresponding division of the mean and target vectors as:

$$\mathbf{E}(\mathbf{Y})^{\mathrm{T}} = (\mathbf{E}(\mathbf{Y}_2),\ \mathbf{E}(\mathbf{Y}_2)) \text{ and } \tau^{\mathrm{T}} = \left(\tau_1^{\mathrm{T}},\ \tau_2^{\mathrm{T}}\right)$$

and the corresponding partitioned components of the $\mathbf{A}$ matrix in Eq. 14 as:

$$\mathbf{A}_{11},\ \mathbf{A}_{12},\ \mathbf{A}_{21}, \text{ and } \mathbf{A}_{22}$$

Under the assumption that $\mathbf{A}$ is symmetric, the average loss can be written as:

$$\begin{aligned} \mathrm{E(MQL)} &= \mathrm{trace}[A\Sigma_{\mathbf{Y}}] + [\mathrm{E}(\mathbf{Y}_1) - \tau_1]^{\mathrm{T}} \mathbf{A}_{11} [\mathrm{E}(\mathbf{Y}_1) - \tau_1] \\ &\quad + 2[\mathrm{E}(\mathbf{Y}_1) - \tau_1]^{\mathrm{T}} \mathbf{A}_{12} [\mathrm{E}(\mathbf{Y}_2) - \tau_2] \\ &\quad + [\mathrm{E}(\mathbf{Y}_2) - \tau_2]^{\mathrm{T}} \mathbf{A}_{22} [\mathrm{E}(\mathbf{Y}_2) - \tau_2] \\ &= \mathrm{trace}[A\Sigma_{\mathbf{Y}}] + \mathbf{OT}_1 + \mathbf{OT}_{12} + \mathbf{OT}_2 \end{aligned} \tag{25}$$

where $\mathbf{OT}_i$ refers to the respective off-target vector.

He assumes the covariance matrix of $\mathbf{Y}$ and the third off-target component are functions of the non-adjustment factors $\mathbf{x}_2$ only, and that the adjustment factors $\mathbf{x}_1$ can be used to shift the mean vector E($\mathbf{Y}_1$) to its target τ_1. It follows that $\Sigma_{\mathbf{Y}} = f(\mathbf{x}_2)$ and $\mathbf{OT}_2 = f(\mathbf{x}_2)$, since the terms $\mathbf{OT}_1$ and $\mathbf{OT}_{12}$ drop to zero when $\mathrm{E}(\mathbf{Y}_1) = t_1$. For this set of assumptions, Eq. 25 can be minimized by the following two-step procedure:

1. Find values of $\mathbf{x}_2$ that minimize trace[$\mathbf{A}\Sigma_{\mathbf{Y}}$] + $\mathbf{OT}_2$, say $\mathbf{x}_2^*$
2. At the values of $\mathbf{x}_2^*$, find values of $\mathbf{x}_1$ that shift the mean vector E($\mathbf{Y}_1$) to its target t_1

Since the stated assumptions are those of the single characteristic problem under an additive model, this procedure is only appropriate when the responses follow an additive model. Additional two-step procedures for constrained and unconstrained minimization of the MQL function for non-additive models are derived.

Up to this point, we have examined squared-error loss functions whose expected value is decomposed into off-target and variance components. Ribeiro and Elsayed (1995) introduced a multivariate loss function which considers, in addition to off-target and variance components, a factor accounting for fluctuation in the supposedly fixed design variable settings. Use of this gradient loss function assumes models of each response (Y_1,..., Y_r) as a function of the process design variables ($\mathbf{x}_1$,..., $\boldsymbol{x}_p$), and estimates the variability induced on Y_i due to the variability of the process parameters using the following terms:

$$\hat{\sigma}_{Y_i}^2 = \sum_{k=1}^{p} \hat{\sigma}_{x_k}^2 \left(\frac{\partial Y_i}{\partial x_k} \right)^2 \tag{26}$$

for when the fluctuations in ($\mathbf{x}_1$,... $\mathbf{x}_p$) are independent of each other, where $\hat{\sigma}^2_{Y_i}$ are the estimated variance of each Y_i, $\hat{\sigma}^2_{x_k}$ are the estimated fluctuations in the design parameters, and $\frac{\partial Y_i}{\partial x_k}$ are the model-predicted shifts in Y_i from the random variation of x_k. When the fluctuations in ($\mathbf{x}_1$,..., $\mathbf{x}_p$) are correlated, this variance term is defined as:

$$\hat{\sigma}^2_{Y_i} = \sum_{k=1}^{p} \hat{\sigma}^2_{x_k} \left(\frac{\partial Y_i}{\partial x_k}\right) + \sum_{k \neq l} \hat{\rho}_{kl} \hat{\sigma}^2_{x_k} \hat{\sigma}^2_{x_l} \left(\frac{\partial Y_i}{\partial x_k}\right)\left(\frac{\partial Y_i}{\partial x_l}\right) \tag{27}$$

where $\hat{\rho}_{kl}$ is the estimated correlation between each pair of design variables x_k and x_l. The authors' multivariate gradient loss (MGL) function is, then, the weighted sum of the individual gradient functions, which, for the case of independent variation in the design settings, can be expressed as:

$$\mathrm{MGL}(\mathbf{x}_j) = \sum_{i=1}^{r} \omega_i \left[(Y_i - t_i)^2 + \hat{\sigma}^2_{Y_i} + \sum_{k=1}^{p} \hat{\sigma}^2_{x_k} \left(\frac{\partial Y_i}{\partial x_k}\right)^2 \right] \tag{28}$$

where MGL($\mathbf{x}_j$) is the MGL function for a particular design vector $\mathbf{x}_j$, and ω_i and t_i are the weights and targets of the individual responses, respectively. The authors allow for a very explicit, quantitative ranking of the responses through the ω_i term. They find the optimal process parameters (i.e., design factor settings) through standard non-linear search techniques.

Ribeiro et al. (2000) extend the gradient loss function (Ribeiro and Elsayed 1995) by adding a term for manufacturing costs. They first convert the dimensionless loss function values (Ribeiro and Elsayed 1995) into dollars by defining the proportionality constant K as:

$$\kappa = \frac{\Delta \mathrm{Value}}{\Delta \mathrm{MGL}} = \frac{A_{\mathrm{mp}} - B_{\mathrm{mp}}}{\mathrm{MGL}_A - \mathrm{MGL}_B} \tag{29}$$

where A_{mp} and B_{mp} are the market prices of class A products, i.e., those with all responses close to target, and of class B products, i.e., those with at least one response out of specification, and where MGL_A and MGL_B are the values of the loss function (Ribeiro and Elsayed 1995) corresponding to the class A and B products. This proportionality constant is then multiplied by the loss function value (Ribeiro and Elsayed 1995) to yield the equivalent lost dollar value resulting from a particular group of design settings $C_{\mathrm{Q}}(\mathbf{x})$:

$$C_{\mathrm{Q}}(\mathbf{x}) = \kappa \mathrm{MGL}(\mathbf{x}) \tag{30}$$

where $\mathbf{x}$ is a vector of design factor settings. They introduce manufacturing costs by starting with a multi-response experiment with r responses ($i = 1,..., r$). They model manufacturing costs $C_{\mathrm{M}}(\mathbf{x})$ as:

$$C_{\mathrm{M}}(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\theta + \varepsilon \tag{31}$$

where $\mathbf{x}$ is the design vector of p regressors, θ is a p-dimensional vector of regression coefficients, and ϵ is the residual. Finally, an extended multivariate loss function which includes costs of poor quality and manufacturing C($\mathbf{x}$) is defined as:

$$C(\mathbf{x}) = C_{\mathrm{Q}}(\mathbf{x}) + C_{\mathrm{M}}(\mathbf{x}) \tag{32}$$

The weighting of responses is accomplished directly through the weighting factor defined for the MGL function (Ribeiro and Elsayed 1995). The authors employ optimization techniques to find the vector $\mathbf{x}$ which minimizes this overall cost function.

3.2 Multivariate utility functions from multiplicative combination

In this section, a multivariate desirability function is constructed from the geometric average of the individual desirability functions of each response. The geometric average (GA) of r components (d_1,..., d_r) is the rth root of their products:

$$\mathrm{GA}(d_1, \ldots, d_r) = \left[\prod_{i=1}^{r} d_i\right]^{\frac{1}{r}} \tag{33}$$

The GA is then a multiplicative combination of the individuals. When combining individual utility functions whose values are scaled between 0 and 1, the GA yields a value less than or equal to the lowest individual utility value. For rating the composite quality of a product, this prevents any single response from reaching an unacceptable value, since a very low value of any crucial characteristic (e.g., safety feature or cost) will render the entire product worthless to the end user.

To demonstrate the simplest case of a geometric average, we show the case of combining two squared-error loss functions. For responses (Y_1, Y_2) with respective loss functions $L(Y_1) = a_1(Y_1 - t_1)^2$ and $L(Y_2) = a_2(Y_2 - t_2)^2$, where ($t_1$, t_2) are the respective targets of the responses, the geometric average of the two loss functions is:

$$\begin{aligned} \mathrm{GA}[(L_1),\ (L_2)] &= \sqrt{a_1(Y_1 - t_1)^2 a_2(Y_2 - t_2)^2} \\ &= \sqrt{a_1 a_2}|(Y_1 - t_1)||(Y_2 - t_2)| \end{aligned} \tag{34}$$

which we recognize as the cross product of the roots of the squared-error loss functions of each response. Since the multiplicative combination of loss functions becomes complicated very quickly, consider the geometric average of two of individual response desirability functions (Harrington 1965). In Eq. 12, we expressed this desirability function as:

$$d_i = \mathrm{e}^{-\frac{4}{(Y_{\mathrm{USL}} - Y_{\mathrm{LSL}})^2}|Y_i - t_i|^2} \tag{35}$$

Therefore, the GA of two such univariate desirability functions results in the following:

$$\mathrm{GA}[(d_1),\ (d_2)] = \sqrt{\mathrm{e}^{-\frac{4}{\left(Y_{\mathrm{USL}(1)} - Y_{\mathrm{LSL}(1)}\right)^2}(Y_1 - t_1)^2}} \cdot \sqrt{\mathrm{e}^{-\frac{4}{\left(Y_{\mathrm{USL}(2)} - Y_{\mathrm{LSL}(2)}\right)^2}(Y_2 - t_2)^2}}$$

which can be expressed as:

$$\mathrm{GA}[(d_1),\ (d_2)] = \sqrt{\mathrm{e}^{-\left[a_1(Y_1-t_1)^2+a_2(Y_2-t_2)^2\right]}} \tag{36}$$

which we recognize as the square root of the natural logarithm constant raised to the sum of the respective loss functions. This example demonstrates the link between the desirability function (Harrington 1965) and the squared-error loss function.

3.2.1 Modifications of the desirability function

In order to allow the DDM to place the ideal target value anywhere within the specifications, Derringer and Suich (1980) introduced a modified version of the desirability function (Harrington 1965). Their desirability function for a one-sided specification is:

$$d_i = \begin{cases} 0 & Y_i \leqslant Y_{i*} \\ \left[\frac{Y_i - Y_{i*}}{Y_i^* - Y_{i*}}\right]^{\phi_i} & Y_{i*} \leqslant Y_i \leqslant Y_i^* \\ 1 & Y_i \geqslant Y_{i*} \end{cases} \tag{37}$$

where Y_{i*} and Y_i^* are the minimal and maximal acceptable levels of Y_i, respectively, Y_i is the response predicted by a certain set of design variable settings, and ϕ_i is a positive constant whose increasing magnitude creates a correspondingly more convex desirability curve. Values of $\phi_i < 1$ create concave curves allowing higher desirability values with Y_i values relatively close to the minimal acceptable level, while values of $\phi_i > 1$ only allow high desirability values when Y_i is very close to the maximal acceptable level. When $\phi_i = 1$, the d_i value is a linear scale between the minimal and maximal acceptable values of the response.

Their desirability function for two-sided specifications is:

$$d_i = \begin{cases} \left[\frac{Y_i - Y_{i*}}{Y_i^* - Y_{i*}}\right]^{\varphi_i} & Y_{i*} \leqslant Y_i \leqslant t_i \\ \left[\frac{Y_i - Y_{i*}}{Y_i^* - Y_{i*}}\right]^{\psi_i} & t_i \leqslant Y_i \leqslant Y_i^* \\ 0 & Y_i < Y_{i*} \text{ or } Y_i < Y_i^* \end{cases} \tag{60}$$

where t_i is the target value of Y_i, and ϕ_i and ψ_i are positive constants chosen by the DDM to indicate the importance of an individual response being close to its target. Larger values of ϕ_i and ψ_i create a desirability curve with a sharper peak at the target value, but with more rapid drop-off as the response moves off-target. Lower values of ϕ_i and ψ_i create a flatter desirability curve that is much less sensitive to a response being off-target.

The desirability function (Harrington 1965) is a special case of the Derringer and Suich (1980) desirability functions, which permit a target value t_i anywhere within the specification limits. Like Harrington (1965), they do not provide for explicit weighting of the individual responses in the overall desirability function. Ranking is implied by the relative steepness of the gradients of the desirability curves, which are in turn the results of specific choices of ϕ_i, ϕ_i, and ψ_i.

Derringer (1994) added explicit weighting terms to the geometric average of the individual desirability functions as follows:

$$\mathrm{D} = \mathrm{GA}\left(d_1^{\omega_1}, \ldots,\ d_r^{\omega_r}\right) = \left(\prod_{i=1}^{r} d_i^{\omega_i}\right)^{\frac{1}{\sum^{\omega_i}}} \tag{39}$$

where setting all the $\omega_i = 1$ yields the geometric average of Eq. 33.

del Castillo et al. (1996) note that, since the desirability functions (Harrington 1965; Derringer and Suich 1980) are non-differentiable at the target points, only direct search optimization methods are applicable. Since the much more efficient gradient-based methods require first-order differentials at all points, they propose using a piecewise continuous desirability function in which the non-differentiable points are corrected using a local polynomial approximation.

Kim and Lin (2000) propose finding the vector of design variable settings $\mathbf{x}$ which maximizes the minimum level that the geometric average of the individual desirability functions may obtain. They state the multi-response optimization problem as:

$$\max_{\mathbf{x}} \iota \tag{40}$$

subject to $d_i(\mathbf{x}) \geq \iota$ for $i = 1, 2, \ldots, r$ and for $\mathbf{x} \in \Omega$, where $d_i(\mathbf{x})$ are the desirability functions of the individual estimated responses $Y_i(\mathbf{x})$. The goal is to identify the $\mathbf{x}$ which maximizes the minimum degree of satisfaction (ι) with respect to all the responses within the experimental region, i.e.:

$$\max_{\mathbf{x} \in \Omega}(\min\,[d_1(\mathbf{x}), \ldots, d_r(\mathbf{x})]) \tag{41}$$

The advantage of this approach is that it does not assume any form or degree of the estimated response models, and is insensitive to the potential dependence between responses. This is shown by contrasting with the method of Khuri and Conlon (1981), which uses only the subset of independent responses and requires that all the independent responses have same order polynomial models of the same subset of design variables. Furthermore, the ι term embodies the overall degree of satisfaction and allows for a quantitative way to compare the results induced by different $\mathbf{x}$ vectors.

They suggest a desirability function of the form:

$$d(z) = \begin{cases} \frac{\exp(\varsigma) - \exp(\varsigma|z|)}{\exp(\varsigma) - 1} & \text{if } \varsigma \neq 0 \\ 1 - |z| & \text{if } \varsigma = 0 \end{cases} \tag{42}$$

where ς is a constant ($-\infty \leq \varsigma \leq \infty$) called the exponential constant and z is a standardized parameter representing distance of the estimated response from its target in units of maximum allowable deviation.

z is calculated differently depending on whether the response is NTB, STB, or LTB. For the NTB case with a symmetric desirability function, z is defined as:

$$z = \frac{Y_i(\mathbf{x}) - t_i}{Y_i^{\max} - t_i} = \frac{Y_i(\mathbf{x}) - t_i}{t_i - Y_i^{\min}} \tag{43}$$

for $Y_i^{\min} \leqslant Y_i(\mathbf{x}) \leqslant Y_i^{\max}$ and where t_i is the target of response i and $Y_i^{\min}$ and $Y_i^{\max}$ are the minimum and maximum values of the individual response, respectively. This z function for the NTB case is readily modified for an asymmetric desirability function.

For the STB case:

$$z = \frac{Y_i(\mathbf{x}) - Y_i^{\min}}{Y_i^{\max} - Y_i^{\min}} \quad \text{for } Y_i^{\min} \leqslant Y_i(\mathbf{x}) \leqslant Y_i^{\max} \tag{44}$$

For the LTB case:

$$z = \frac{Y_i^{\max} - Y_i(\mathbf{x})}{Y_i^{\max} - Y_i^{\min}} \quad \text{for } Y_i^{\min} \leqslant Y_i(\mathbf{x}) \leqslant Y_i^{\max} \tag{45}$$

It is easily verified that $(-1 \leq z \leq +1)$ for an NTB response and $(0 \leq z \leq 1)$ for the STB and LTB responses. In all cases, $d(z)$ is maximized when $z = 0$, which happens when $Y_i(\mathbf{x})$ is equal to the target value.

The choice of ς determines the relative concavity of the individual desirability curves. Increasingly negative values of ς produce decreasingly concave curves (i.e., more convex) and increasingly positive values of ς produce desirability curves of growing concavity. They define relative concavity and convexity, respectively, as insensitivity and sensitivity to the off-target distance.

For example, a response with $\varsigma = -5$ has a desirability curve whose values drop sharply with increasing off-target distance, while a response with $\varsigma = +5$ has desirability values that change slowly as the response moves further off-target. The relative values of the ς variable approximate the weighting of the individual responses on the geometric average. Hence, weighting is accomplished indirectly by choosing lower ς values for the responses of higher priority.

Kim and Lin (2000), furthermore, incorporate a technique that accounts for the predictive ability of the individual response models. They do this by transforming the original ς values to ς' indicative of the predictive ability of the response models. For example:

$$\varsigma' = \varsigma + \left(1 - R^2\right)\left(\varsigma^{\max} - \varsigma\right) \tag{46}$$

will decrease each ς value inversely with rising R^2, where R^2 is the standard coefficient of determination in linear regression, and $\varsigma^{\max}$ is a sufficiently large value of ς such that $d(z)$ with $\varsigma^{\max}$ is extremely concave, hence, having negligible effect on the optimization.

This makes the resulting desirability curve more convex, i.e., of higher priority in the geometric average, as the R^2 values increase. This effectively adjusts the relative weighting of each individual desirability function according to the predictive ability of the corresponding response model so that better predictive models get higher weighting than poorly predictive models. They demonstrate the attainment of a design vector yielding a higher overall desirability using the ς' transformation than that obtained with the original ς value. Although the authors use R^2 in their example, the DDM can use any preferred metric of predictive ability. In general, the desirability function approach neither assumes response independence nor exploits the response correlation information.

3.2.2 *The non-domination search technique*

Loy et al. (2000) extend the dual response approach (Vining and Myers 1990) to the multi-response case by searching for groupings of design variable vectors which yield responses that are non-dominated with respect to each other for each of the r separate responses. This approach assumes a finite sample space with models for both the mean and variance of each response so that they can be predicted for all responses from all possible design vectors.

This technique's relation to the dual response approach (Vining and Myers 1990) is evident in the following formulations for the NTB, LTB, and STB cases:

$$\text{NTB}: \quad \begin{aligned} &\min f_1(\mathbf{x}) = |Y(\mathbf{x}) - \tau| \\ &\min f_2(\mathbf{x}) = \hat{\sigma}_Y^2 \end{aligned} \tag{47}$$

$$\text{LTB}: \quad \begin{aligned} &\max f_1(\mathbf{x}) = Y(\mathbf{x}) \\ &\min f_2(\mathbf{x}) = \hat{\sigma}_Y^2 \end{aligned} \tag{48}$$

$$\text{STB}: \quad \begin{aligned} &\min f_1(\mathbf{x}) = Y(\mathbf{x}) \\ &\min f_2(\mathbf{x}) = \hat{\sigma}_Y^2 \end{aligned} \tag{49}$$

wherein each response $f_1(\mathbf{x})$ is optimized with the constraint of minimizing variance, i.e., $f_2(\mathbf{x}) = \hat{\sigma}_Y^2$.

We now demonstrate the meaning of non-domination using the NTB case of Eq. 47 as an example. A vector of design variable settings (i.e., $\mathbf{x1}$) is said to dominate another design vector $\mathbf{x2}$ when no value of $[f_1(\mathbf{x2}), f_2(\mathbf{x2})]$ is less than the corresponding element of $[f_1(\mathbf{x1}), f_2(\mathbf{x1})]$, and at least one value of $f_1(\mathbf{x2})$ or $f_2(\mathbf{x2})$ is strictly greater than the corresponding element of $[f_1(\mathbf{x1}), f_2(\mathbf{x1})$.

When the first grouping of non-dominated design vectors (i.e., front) is formulated for each of the r responses, the DDM searches for the intersection of these r first fronts. This intersection set can range from the empty set to a large number of design vectors. For the empty set, the DDM proceeds to examine the intersections of the successive fronts of the r responses. The authors apply this procedure to a military field ration study (Wurl and Albin 1999) and identify design vectors yielding comparable results to those identified using the expected loss approach (Pignatiello 1993) and the desirability approach (Derringer and Suich 1980).

This technique differs from the expected loss and desirability approaches in that it allows the DDM to proceed without subjectively prioritizing or combining the multiple response loss or utility functions into a single overall objective function. This means that the DDM does not have to struggle with different engi-

neering units or the conceptual challenge of combining different types of responses into a single objective function. Likewise, the DDM does not have to consider the trade-off between off-target and variance components of the same overall objective function. This effectively shifts the required engineering judgment from the front end of the optimization process to the end of the process. With respect to response correlation structure, this technique neither assumes response independence nor exploits this source of statistical information.

4 The compromise decision support problem

Up to this point, all the multi-response optimization techniques have been culled from the statistics literature. To contrast these approaches with an example from the engineering literature, we briefly review an important multi-response technique, which evolved from experimentation in the ship building and design industry. The compromise decision support problem (cDSP) is a mathematical construct with which the conflicting goals in product design are resolved. The high-level description of the baseline, deterministic cDSP presented in Fig. 1 and the following paragraphs are adapted from Mistree et al. (1993).

The cDSP is a multi-objective decision model based on mathematical programming and goal programming. In the cDSP, values of design variables are determined to achieve a set of conflicting goals to the best extent possible while satisfying a set of constraints. The cDSP simultaneously considers system variables, i.e., design variables, system constraints, system goals, and deviation variables. The system variables $\mathbf{x}' = (x_1, \ldots, x_p)$ usually describe design variables of the system and each cDSP must have at least two system variables which may be continuous, discrete, or Boolean. System variables are bounded to help the designer use experience-based judgment in formulating the problem.

System constraints and bounds define the feasible design space and are functions of the system variables only. The responses are assumed to be functions of the system variables, i.e., $\Psi_i(\mathbf{x})$. The designer's aspiration for each response is represented by a system goal (G_i) and the deviation variables $\left(\delta_i^-, \delta_i^+\right)$ are the level of under-achievement or over-achievement of a goal, respectively. The goals are modeled as constraints in the following form:

$$\psi_i(\mathbf{x}) + \delta_i^- - \delta_i^+ = G_i \tag{50}$$

The high-level aim of the cDSP is to minimize the difference between the goal (G_i) of each objective and its actual performance $\Psi_i(\mathbf{x})$. This is accomplished by finding settings of the system variables that minimize the deviation function (DF), the overall objective function

Fig. 1 Baseline cDSP

Given	p,	the number of system variables
	ζ,	the number of equality constraints
	η,	the number of inequality constraints
	ϑ,	the number of objectives
	$\Psi_i(\mathbf{x})$,	the goal achievement functions
	$\gamma_h(\mathbf{x})$,	the constraint functions
	G_i,	the target values for the goals
	$f_K(\delta_i)$,	function to be minimized at priority level K for preemptive form
	ω_i,	weight for the Archimedean form
Find	System Variables	$\mathbf{x} = (x_1, \ldots, x_p)$
	Deviation Variables	$\delta_i^-, \delta_i^+ \; for \; i = 1, \ldots, \vartheta$
Satisfy	Goals	$\Psi_i(\mathbf{x}) + \delta_i^- - \delta_i^+ = G_i$
	Constraints	$\gamma_h(\mathbf{x}) = 0 \; for \; h = 1, \ldots, \zeta$
		$\gamma_h(\mathbf{x}) \geq 0 \; for \; h = \zeta + 1, \ldots, \zeta + \eta$
	Bounds	$x_{l,min} \leq x_l \leq x_{l,max} \; for \; l = 1, \ldots, p$
	Other	$\delta_i^- \cdot \delta_i^+ = 0 \; and \; \delta_i^-, \delta_i^+ \geq 0$
Minimize	Deviation Function (DF)	Archimedean: $DF = \sum_{i=1}^{2\vartheta} (\omega_i^- \delta_i^- + \omega_i^+ \delta_i^+)$ where $\sum_{i=1}^{2\vartheta} \omega_i = 1$
		Preemptive: $DF = \left[f_1(\delta_i^-, \delta_i^+), \ldots, f_K(\delta_i^-, \delta_i^+)\right]$

of the cDSP, which is a function of the deviation variables of the system goals.

The deviation function DF is defined for two cases which differ in how they prioritize the system goals. They are called the Archimedean and Lexicographic weighting schemes. In the Archimedean approach, the DDM explicitly assigns the weights ω_i^- and ω_i^+ to reflect the importance of the individual goals. Lexicographic weighting (Ignizio 1982) does not require the DDM to assign specific weights to the objectives. Rather, goals are rank ordered in terms of their priority and deviation variables in the highest priority goal are minimized first, followed by the deviation variables in the second priority goal, and so on.

The apparent strength of the cDSP is its handling of highly constrained environments. It provides flexible decision support for achieving a compromise among multiple goals while satisfying constraints and bounds. Although the baseline cDSP is deterministic, there exist a number of extensions which reflect the cDSP's nature as a living construct, capable of being strengthened and/or specialized through augmentation.

Bayesian and fuzzy logic versions of the cDSP (Vadde et al 1994; Zhou et al. 1992) allow the DDM to use the cDSP while accommodating uncertainty regarding constraints, values for goal targets, or weighting of goals in the problem formulation. There are many other extensions of the cDSP (Srinivasan et al. 1991; Karandikar and Mistree 1991; Vadde et al. 1994; Seepersad 1997; Seepersad et al. 2002). Of most relevance to this paper is the RD formulation of the cDSP by Chen et al. (1996).

The approach of Chen et al (1996)—a cDSP-based procedure called the robust concept exploration method (RCEM)—is a domain-independent method for generating robust, multidisciplinary design solutions. The RCEM employs experiment-based metamodels to alleviate some of the computational difficulties associated with probability-based robust design. The RCEM defines a design process as either Type I or Type II, depending on where variability arises in the design process.

Type I is a traditional RD with fixed design settings and variability hailing from uncontrollable noise factors. Type II is common to many engineering design simulations where variability is injected around the nominal values of the design variables, rather than emanating from noise variables. Type II is particularly important in the early stages of design since the design will likely evolve over time and the desirable values of the system variables will change, leading to uncertainty.

Both types use response surface models to approximate the expected value of each response. In Type I, the mean response is estimated as a function of system

Fig. 2 cDSP with RD Goal Formulations from Chen et al. (1996)

Given	ω_i,	weight in the Archimedean form
	Tmu_{power},	target for mean of power
	$Tsigma_{power}$,	target for variance of power
	Tmu_{eff},	target for mean of efficiency
	$Tsigma_{eff}$,	target for variance of efficiency
	Tmu_{mass},	target for mean of mass
	$Tsigma_{mass}$,	target for variance of mass
Find	System Variables	$\mathbf{x} = (x_1, \ldots, x_p)$
	Deviation Variables	$\delta_i^-, \delta_i^+ \; for \; i = 1, \ldots, 6$
Satisfy	NTB Goals	
	Bring mean of power to target	$\frac{\hat{\mu}_{power}(\mathbf{x})}{Tmu_{power}} + \delta_1^- - \delta_1^+ = 1$
	Minimize variance of power	$\frac{\hat{\sigma}_{power}(\mathbf{x})}{Tsigma_{power}} + \delta_2^- - \delta_2^+ = 1$
	LTB Goals	
	Maximize mean of efficiency	$\frac{\hat{\mu}_{eff}(\mathbf{x})}{Tmu_{eff}} + \delta_3^- - \delta_3^+ = 1$
	Minimize variance of efficiency	$\frac{\hat{\sigma}_{eff}(\mathbf{x})}{Tsigma_{eff}} + \delta_4^- - \delta_4^+ = 1$
	STB Goals	
	Minimize mean of mass	$\frac{\hat{\mu}_{mass}(\mathbf{x})}{Tmu_{mass}} + \delta_5^- - \delta_5^+ = 1$
	Minimize variance of mass	$\frac{\hat{\sigma}_{mass}(\mathbf{x})}{Tsigma_{mass}} + \delta_6^- - \delta_6^+ = 1$
	Bounds	$x_{l,min} \leq x_l \leq x_{l,max} \; for \; l = 1, \ldots, p$
	Other	$\delta_i^- \cdot \delta_i^+ = 0 \;\; and \;\; \delta_i^-, \delta_i^+ \geq 0$
Minimize	Deviation Function (DF)	Archimedean: $DF = \sum_{i=1}^{12} (\omega_i^- \delta_i^- + \omega_i^+ \delta_i^+)$ where $\sum_{i=1}^{12} \omega_i = 1$

Table 1 Comparing the multi-response optimization techniques

Metric	Technique			
	Expected Loss	Desirability	Non-domination search	cDSP
Maximum number of input variables	Limited only by experimental and computational resources	Limited only by experimental and computational resources	Maximum of 6 recommended	Limited only by experimental and computational resources
Use of response correlations	Directly used	Not used	lndirectly used	Only if variance measure is a stated goal
Maximum number of responses	Limited only by experimental and computational resources	Limited only by experimental and computational resources	Maximum of 3 recommended	Limited only by experimental and computational resources
Experimental replication required?	Yes	No	No	Only if variance measure is a stated goal
Modeling required	Overall exponential loss or mean and variance	Overall desirability or mean and variance	Mean and variance	Continuous or Boolean variable response models
Formulation of mean	lncluded	lncluded	lncluded	Only if mean measure is stated goal
Formulation of variance	Explicit	lmplicit	lncluded	Only if variance measure is a stated goal
Working units of objective function	Monetary loss units	Dimensionless	Original engineering units	Dimensionless
Bounds on response values	Not required	Required	Not required	Required
Weighting of responces	Specified thru loss function coefficients	Indirectly specifeid thru individual desirability curves	Implicit	Lexicographic and archimedean weights
Handling of constraints	May not be possible	May not be possible	Relatively easy	Designhed for many, differentiable constraints
Penalty for deviation from target	Qadratic functions assumed, coefficients required	Explicit functions required	Implicit	Not applicable
Trade-off between off-target and variance	Equal	Not applicable	Implicit	Not applicable
Resulting vector of design variable settings	One x vector or none	One x vector or none	Alternative x vectors likely	One x vector or alternative x vectors depending on formulation
Applicable optimization settings	Nonlinear search	Nonlinear search	Exhaustive non-domination search	Sequential and adaptive linear programming
Consequence of new target	Direct search: need to reoptimize; two-step: no need to reoptimize	Need to reoptimize	Need to reoptimize	Need to reoptimize
Consequence of new specification	Need to reoptimize	Need to reoptimize	Need to reoptimize	Need to reoptimize

variables and an average noise component from each individual noise variable. In Type II, only the system variables are considered. Hence, the noise information is not exploited to the same extent as in the corresponding loss function approaches, whose model terms for each significant control–noise interaction allow a higher degree of resolution.

Both Type I and Type II typically employ first-order Taylor expansions to approximate variability. Taylor expansions in Type I are a function of the variability of individual noise factors while Type II's are associated with the design factors. Depending on how well these first-order Taylor expansions emulate the true variability, there may be a loss of accuracy regarding the variability contributions within the RCEM.

Like the desirability function and the non-domination search techniques, the RCEM makes no assumptions with respect to response correlation and does not actively use that correlation structure in its optimization algorithm. The RCEM mimics the dual response approach (Vining and Myers 1990) in treating each response's location and dispersion as a pair of criteria where one is optimized and the other is constrained. Chen et al. (1996) demonstrate the RCEM in the design of a solar powered irrigation system and mention other successful applications.

As an example of goal formulation in the RCEM, imagine a simple system with the three response characteristics of power, efficiency, and weight. We wish to design so that power is at a specific level, efficiency is maximized, and weight minimized. These are, respectively, nominal-the-best (NTB), larger-the-better (LTB), and smaller-the-better (STB) goals per the Taguchi terminology. The NTB goals seek design variable settings bringing the response mean as close as possible to its target while minimizing variance. The LTB goals seek to maximize mean response while minimizing variance and the STB goals minimize both mean and variance. Assuming that there are no non-goal-related constraints, Fig. 2 gives the cDSP formulation for the six goals making up this design problem.

5 Comparing the multi-response optimization techniques

In this section, we compare the attributes of expected loss, desirability, the non-domination search, and the cDSP.

To compare the four methods discussed in this paper, Table 1 adds the cDSP to the table in Loy et al. (2000). Because all four approaches can handle the NTB, STB, and LTB response types, this criterion is not included in the table.

Since we have focused on the statistical and optimizing properties of the approaches, the software and computational resources necessary for carrying out the optimization algorithms are not addressed in this paper.

Acknowledgments The authors would like to thank F. Mistree and the referees for their helpful comments. We gratefully acknowledge support from the National Science Foundation, grant DMI-0100123. T.E. Murphy's work was also supported by the School of Industrial and Systems Engineering at the Georgia Institute of Technology. K.-L. Tsui's work was also supported by The Logistics Institute—Asia Pacific in Singapore.

References

Ames A, Mattucci N, MacDonald S (1997) Quality loss functions for optimization across multiple response surfaces. Technometrics 31:401–409

Artiles-Leon N (1996) A pragmatic approach to multiple-response problems using loss functions. Qual Eng 9:213–220

Berger J (1995) Statistical decision theory. Springer, Berlin Heidelberg New York

Box GEP (1988) Signal-to-noise ratios, performance criteria, and transformations. Technometrics 30:1–31

Box GEP, Bisgaard S, Fung C (1988) An explanation and critique of Taguchi's contributions to quality engineering. Qual Reliab Eng Int 4:114–123

Box GEP, Jones S (1990) Designs for minimizing the effects of environmental variables. Technical report, University of Wisconsin

Chen W, Allen J, Tsui K, Mistree F (1996) A procedure for robust design: minimizing variation caused by noise factors and control factors. J Mech Design 118:478–485

Contreras L, Bruins H, Trout J (1995) Heating characteristics of beef stew in MRE pouches as influenced by critical process and product parameters. Technical report 100, Rutgers University, New Jersey

del Castillo E, Montgomery D (1993) A nonlinear programming solution to the dual response problem. J Qual Technol 25:199–204

del Castillo E, Montgomery D, McCarville D (1996) Modified desirability functions for multiple response optimization. J Qual Technol 28:337–345

Derringer GC (1994) A balancing act: optimizing a product's properties. Qual Prog 27:51–58

Derringer GC, Suich R (1980) Simultaneous optimization of several response variables. J Qual Technol 12:214–219

Harrington E (1965) The desirability function. Ind Qual Control 21:494–498

Hazelrigg GA (1996) Systems engineering: an approach to information-based design. Prentice-Hall, Upper Saddle River, New Jersey

Ignizio J (1982) Linear programming in single and multi-objective systems. Prentice-Hall, Englewood Cliffs, New Jersey

Kackar R (1985) Off-Line quality control, parameter design, and the Taguchi method. J Qual Technol 17:176–209

Karandikar HM, Mistree F (1991) The integration of information from design and manufacture through decision support problems. Appl Mech Rev 44:150–159

Khuri A, Conlon M (1981) Simultaneous optimization of multiple responses represented by polynomial regression functions. Technometrics 23:363–375

Kim K, Lin D (2000) Simultaneous optimization of mechanical properties of steel by maximizing exponential desirability functions. Appl Stat 49:311–325

Kumar P, Barua P, Gaindhar JL (2000) Quality optimization (multi-characteristics) through Taguchi's technique and utility concept. Qual Reliab Eng Int 16:475–485

Leon R, Shoemaker A, Kackar R (1987) Performance measures independent of adjustment: an explanation and extension of Taguchi's signal-to-noise ratios. Technometrics 29:253–265

Lin D, Tu W (1995) Dual response surface optimization. J Qual Technol 28:496–499

Loy C, Goh T, Xie M, Tsui K (2000) Design of experiments considering multiple engineering characteristics. In: Proceedings of the IEEE international conference on management of information and technology (ICMIT 2000), Singapore, November 2000, 2:874–879

Mistree F, Hughes OF, Bras BA (1993) The compromise decision support problem and the adaptive linear programming algorithm. Struct Optimization 247–286

Nair VN (1986) Testing in industrial experiments with ordered categorical data. Technometrics 28:283–291

Nair VN (1992) Taguchi's parameter design: a panel discussion. Technometrics 34:127–161

Pignatiello JJ (1993) Strategies for robust multiresponse quality engineering. IIE Trans 25:5–15

Ribeiro JL, Elsayed E (1995) A case study on process optimization using the gradient loss function. Int J Prod Res 33:3233–3248

Ribeiro JL, Fogliatto FS, Caten CST (2000) Minimizing manufacturing and quality costs in multiresponse optimization. Qual Eng 13:191–201

Seepersad C (1997) The utility-based compromise decision support problem with applications in product platform design. Master's thesis, G.W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology

Seepersad C, Mistree F, Allen J (2002) A quantitative approach for designing multiple platforms for and evolving portfolio of products. In: Proceedings of the 39th design automation conference (DAC 2002), New Orleans, Louisiana, June 2002

Simon HA (1976) Administrative behavior: a study of decision-making processes in administrative organization. Free Press, New York

Srinivasan R, Karandikar H, Mistree F (1991) Understanding design–manufacture interaction using compromise decision support problems. Part I: a formulation for composite pressure vessels. Part II: preliminary synthesis of composite pressure vessels. Part III: design for manufacture of composite pressure vessels. Comput Struct 40:679–717

Taguchi G (1986) Introduction to quality engineering: designing quality into products and processes. Kraus International Publications, White Plains, New York

Tsui K-L (1996a) A critical look at Taguchi's modelling approach. J Appl Stat 23:81–95

Tsui K-L (1996b) A multi-step analysis procedure for robust design. Stat Sinica 6:631–648

Tsui K-L (1999) Robust design optimization for multiple characteristic problems. Int J Prod Res 37:433–445

Tsui K-L, Li A (1994) Analysis of smaller and larger the better robust design experiments. Int J Ind Eng 1:193–202

Vadde S, Krishnamachari RS, Allen JK, Mistree F (1994) The Bayesian compromise decision support problem for hierarchical design involving uncertainty. J Mech Des 116:388–395

Vining G, Myers R (1990) Combining Taguchi and response surface philosophies: a dual response approach. J Qual Technol 22:38–45

Wurl R, Albin S (1999) A comparison of multiresponse optimization: sensitivity to parameter selection. Qual Eng 11:405–415

Zhou Q, Allen J, Mistree F (1992) Decisions under uncertainty: the fuzzy compromise decision support problem. Eng Optimization 20:21–43