

A cautious approach to robust design with model parameter uncertainty

DANIEL W. APLEY^{1,*} and JEONGBAE KIM²

¹Department of Industrial Engineering & Management Sciences, Northwestern University, Evanston, IL 60208-3119, USA
E-mail: apley@northwestern.edu

²Korea Telecom Headquarters, 206 Jungja-Dong Bundang-Gu, Seongnam, Kyunggi, Korea, 463-711
E-mail: jeongbae@kt.com

Received September 2009 and accepted August 2010

Industrial robust design methods rely on empirical process models that relate an output response variable to a set of controllable input variables and a set of uncontrollable noise variables. However, when determining the input settings that minimize output variability, model uncertainty is typically neglected. Using a Bayesian problem formulation similar to what has been termed *cautious control* in the adaptive feedback control literature, this article develops a cautious robust design approach that takes model parameter uncertainty into account via the posterior (given the experimental data) parameter covariance. A tractable and interpretable expression for the posterior response variance and mean square error is derived that is well suited for numerical optimization and that also provides insight into the impact of parameter uncertainty on the robust design objective. The approach is cautious in the sense that as parameter uncertainty increases, the input settings are often chosen closer to the center of the experimental design region or, more generally, in a manner that mitigates the adverse effects of parameter uncertainty. A brief discussion on an extension of the approach to consider model structure uncertainty is presented.

Keywords: Robust parameter design, cautious control, model uncertainty, Bayesian estimation, quality control, variation reduction, Six Sigma

1. Introduction

In robust parameter design, which has received considerable attention from academia and industry, one optimally selects the levels of a set of controllable variables (a.k.a. inputs) in order to minimize variability in an output response variable, while keeping the mean of the response variable close to a target. The component of the response variability that can be affected by adjusting the inputs is typically assumed to be due to a set of uncontrollable (a.k.a. noise) variables. Hence, minimizing response variability amounts to choosing the inputs so that the output response is robust or insensitive to variations in the noise variables. Two main approaches to this problem are Taguchi's robust parameter design (Taguchi, 1986; Nair, 1992; Wu and Hamada, 2000), which employs signal-to-noise ratios and crossed-array experimental designs, and response surface methodology in conjunction with combined-array designs (Vining and Myers, 1990; Shoemaker *et al.*, 1991; Myers *et al.*, 1992; Lucas, 1994; Khattree, 1996).

This article is focused on the response surface approach, which is often advocated because of its stricter adherence to

well-established techniques for statistical modeling, analysis, and experimental design. Consider the following response surface model, which is widely assumed in robust design studies (Myers and Montgomery, 2002). The output response y is represented as

$$y = \alpha + \boldsymbol{\beta}'\mathbf{g}(\mathbf{x}) + \boldsymbol{\gamma}'\mathbf{w} + \mathbf{w}'\mathbf{B}\mathbf{x} + \varepsilon \quad (1)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_p]'$ is a vector of p controllable input variables, $\mathbf{w} = [w_1, w_2, \dots, w_m]'$ is a vector of m uncontrollable noise variables, and ε is the model residual error. It is assumed that \mathbf{w} is random with mean zero and known covariance matrix $\boldsymbol{\Sigma}_w$ (typically diagonal) and that ε is normally distributed with mean zero and variance σ^2 , independent of \mathbf{w} . Each element of the l -length vector $\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_l(\mathbf{x})]'$ is a known function of the p controllable input variables. The scalar α , the l -length vector $\boldsymbol{\beta}$, the m -length vector $\boldsymbol{\gamma}$, and the $m \times p$ matrix \mathbf{B} comprise the model parameters (excluding σ , which we treat differently), which we denote collectively by the vector $\boldsymbol{\theta} = [\alpha, \boldsymbol{\beta}', \boldsymbol{\gamma}', \mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_p]'$, where \mathbf{b}_i denotes the i th column of \mathbf{B} . If $\mathbf{g}(\mathbf{x}) = \mathbf{x}$, for example, the model includes the main effects of \mathbf{x} and its interactions with \mathbf{w} . A more common choice for $\mathbf{g}(\mathbf{x})$ in robust design studies is $\mathbf{g}(\mathbf{x}) =$

*Corresponding author

$[x_1, x_2, \dots, x_p, x_1^2, x_2^2, \dots, x_p^2, x_1x_2, x_1x_3, \dots, x_1x_p, x_2x_3, \dots, x_{p-1}x_p]'$, in which case the model is full quadratic in \mathbf{x} .

The standard dual response approach is to select \mathbf{x} to minimize:

$$\text{Var}_{\varepsilon, \mathbf{w}}(y | \boldsymbol{\theta}, \sigma) = (\boldsymbol{\gamma} + \mathbf{B}\mathbf{x})' \boldsymbol{\Sigma}_{\mathbf{w}}(\boldsymbol{\gamma} + \mathbf{B}\mathbf{x}) + \sigma^2, \quad (2)$$

subject to the constraint that the mean

$$E_{\varepsilon, \mathbf{w}}(y | \boldsymbol{\theta}, \sigma) = \alpha + \boldsymbol{\beta}'\mathbf{g}(\mathbf{x}) \quad (3)$$

equals some specified target T . Alternatively, one may select \mathbf{x} to minimize a Mean Square Error (MSE) objective function $E_{\varepsilon, \mathbf{w}}[(y - T)^2 | \boldsymbol{\theta}, \sigma]$. The subscripts on the variance and expectation operators indicate which random variables the operations are with respect to, and we have written them as conditioned on the model parameters $\boldsymbol{\theta}$ and σ . The reason for the latter is that when optimizing \mathbf{x} , one generally estimates the parameters using experimental design and analysis techniques and then views the estimates as if they were the true parameters. Hence, parameter uncertainty due to estimation error is generally neglected, which, as noted in Shoemaker *et al.* (1991), could actually result in an increase in the response variance.

This article develops a method for taking parameter uncertainty into consideration when choosing the input settings. The objective is to find robust design input settings for which the response is robust to parameter estimation errors, as well as to the noise \mathbf{w} . The Bayesian MSE, which is referred to as the *Cautious Robust Design* (CRD) objective function in this article, is minimized. Here, the CRD objective function is defined as

$$J_{\text{CRD}}(\mathbf{x}) = E_{\varepsilon, \mathbf{w}, \boldsymbol{\theta}, \sigma}[(y - T)^2 | \mathbf{Y}], \quad (4)$$

where \mathbf{Y} denotes the observed response values over the experiment from which the parameters are estimated. The subscripts $\boldsymbol{\theta}$ and σ are added on the expectation operator to indicate that it is with respect to the posterior distribution of the parameters, given the data \mathbf{Y} , in addition to the distributions of \mathbf{w} and ε . It will be shown that $J_{\text{CRD}}(\mathbf{x})$ can be expressed as a quite tractable function of \mathbf{x} , $\hat{\boldsymbol{\theta}}$, $\hat{\sigma}^2$, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, $\boldsymbol{\Sigma}_{\mathbf{w}}$, and T , where $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ denote the posterior means (point estimates) of $\boldsymbol{\theta}$ and σ^2 , and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ denotes the posterior covariance matrix of $\boldsymbol{\theta}$. Thus, minimizing $J_{\text{CRD}}(\mathbf{x})$ will yield optimal \mathbf{x} settings that are a function of the posterior covariance $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, thereby taking into account parameter uncertainty. The Bayesian strategy of minimizing an objective function of the form of Equation (4) bears close resemblance to what has been referred to as cautious control in the adaptive control literature (Åström and Wittenmark, 1995). Hence the use of the CRD terminology.

The CRD objective function (or the posterior variance of the response in an analogous dual response CRD formulation, which will be considered later in this article) is a natural extension of the standard robust design objective function and, hence, should have familiar conceptual appeal to practitioners. The resulting CRD approach has a number of attractive characteristics. It leads to a relatively

simple, closed-form expression for the objective function. Other Bayesian approaches for considering parameter uncertainty in robust design (reviewed in Section 2) require Monte Carlo simulation to calculate the objective function. A different Monte Carlo simulation must be conducted for each \mathbf{x} of interest, which prohibits analytical optimization of the objective function and complicates numerical optimization. The analytical expressions provide a convenient smooth function that can be easily evaluated within an optimization routine. More generally, these analytical expressions provide insight into the mechanisms behind robustness to parameter uncertainty or the lack thereof.

The format of the remainder of this article is as follows. Section 2 reviews prior Bayesian and frequentist approaches for considering model structure and parameter uncertainty in experimental-based process optimization and robust design. Section 3 discusses the prior and posterior distributions for the parameters and provides expressions for $\hat{\boldsymbol{\theta}}$, $\hat{\sigma}^2$, and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$. These in turn are used in Section 4 to develop a tractable expression for $J_{\text{CRD}}(\mathbf{x})$. For a special case of the model (1) that is linear in \mathbf{x} (i.e., $\mathbf{g}(\mathbf{x}) = \mathbf{x}$), a closed-form expression exists for the \mathbf{x} that minimizes the CRD objective function. This provides insight into how CRD ensures robustness to parameter uncertainty, which is discussed in Section 5. In Section 6, a leaf spring manufacturing example from the literature is used to illustrate CRD and compare it to standard robust design in which parameter uncertainty is neglected. Although the CRD objective function considers variability due to parameter uncertainty on par with variability due to the noise variables, it has a natural decomposition into (i) the standard robust design expression for the response variability that results when the parameters are treated as known; and (ii) the additional response variability due to parameter uncertainty. Section 7 continues the leaf spring example to illustrate how this decomposition provides insight into whether the current experiment yields sufficient information for optimizing the process versus whether additional experimentation is needed to reduce parameter uncertainty. Section 8 discusses the distinctions between parameter uncertainty and noise variability. Section 9 considers the implications of having constraints on the input variables, which are common in robust design optimization problems. Section 10 briefly discusses an extension of the CRD approach that considers uncertainty in the model structure. Section 11 discusses an extension of the CRD concepts to dual-response robust design, in which the objective is to minimize $\text{Var}_{\varepsilon, \mathbf{w}, \boldsymbol{\theta}, \sigma}[y | \mathbf{Y}]$, subject to the constraint $E_{\varepsilon, \mathbf{w}, \boldsymbol{\theta}, \sigma}[y | \mathbf{Y}] = T$. Section 12 concludes the article.

2. Review of prior work on model and parameter uncertainty in experimental-based process optimization and robust design

Cautious adaptive feedback control strategies that involve a Bayesian MSE objective function have been widely

investigated (Åström and Wittenmark, 1995). Recently, Apley (2004) and Apley and Kim (2004) have investigated non-adaptive versions of cautious control. Their work was in the context of automatic feedback control, in which the input settings \mathbf{x} are actively adjusted online as each new response observation is obtained. In the context of robust design, in which the online input settings are held fixed at some optimized values based on an offline experiment, the use of the Bayesian MSE objective function (4) follows Apley and Kim (2002) and Kim (2002). A number of other approaches have also been proposed for taking into account model uncertainty in robust design and, more generally, in experimental-based process optimization. Box and Hunter (1954) developed a confidence region for the values of \mathbf{x} that constitute a stationary point for (e.g., that maximize or minimize) a response surface. The confidence region is with respect to uncertainty in the parameters, which is taken into account from a frequentist perspective via the dependence of the confidence region on the distribution of the parameter estimates. Peterson *et al.* (2002) developed an alternative confidence region that distinguishes saddle points from minima/maxima and that can handle constraints on the inputs. In another frequentist approach, Myers and Montgomery (2002, p. 576) derived an unbiased estimate of $\text{Var}_{\varepsilon, \mathbf{w}}(y | \boldsymbol{\theta}, \sigma)$ in Equation (2), which involves parameter uncertainty via the dependence of their bias correction term on the covariance matrix of $\hat{\boldsymbol{\theta}}$. They recommended minimizing the unbiased variance estimate and also discussed a graphical approach in which one plots unbiased estimates of the mean and variance in Equations (2) and (3) as functions of \mathbf{x} , while simultaneously displaying a confidence region for the true \mathbf{x} settings that minimize $\text{Var}_{\varepsilon, \mathbf{w}}(y | \boldsymbol{\theta}, \sigma)$. Miró-Quesada and Del Castillo (2004) used the same bias correction term as Myers and Montgomery (2002), but their objective was to minimize an unbiased estimate of $\text{Var}_{\hat{\boldsymbol{\theta}}, \mathbf{w}}(\hat{y}(\mathbf{w}) | \boldsymbol{\theta}, \sigma)$, where $\hat{y}(\mathbf{w})$ is Equation (1) with $\hat{\boldsymbol{\theta}}$ substituted for $\boldsymbol{\theta}$. Parameter uncertainty was taken into account from a frequentist perspective by virtue of the variance operation being with respect to $\hat{\boldsymbol{\theta}}$, as well as \mathbf{w} , which resulted in an expression that was a function of the covariance matrix of $\hat{\boldsymbol{\theta}}$. In a sense that will be discussed in Section 11, the CRD objective function that is adopted results in a more complete accounting of parameter uncertainty. Sahni *et al.* (2009) considered model uncertainty in the context of mixture-process optimization. Monroe *et al.* (2010) considered the effects of model uncertainty on the selection of optimal designs for accelerated life tests.

A number of Bayesian approaches have also been proposed for taking model uncertainty into account. Chipman (1998) considered the posterior distribution of $\{\boldsymbol{\theta}, \sigma\} | \mathbf{Y}$ and then recommended a Monte Carlo simulation in which values of $\{\boldsymbol{\theta}, \sigma\}$ are drawn from their posterior distribution and substituted into Equations (2) and (3) (or any other robust design criterion). A separate Monte Carlo simulation is conducted for each value of \mathbf{x} of interest. The average and/or sample variance of Equations (2) and

(3) over the Monte Carlo simulation can guide a designer in choosing \mathbf{x} settings for which the response mean and variance are robust to uncertainty in $\boldsymbol{\theta}$. Chipman (1998) also considered uncertainty in the model structure. Using the approach of Box and Meyer (1993), Chipman calculated the posterior probabilities that each model within some class (e.g., all models consisting of subsets of the individual terms in Equation (1)) is the true one, and then within the Monte Carlo simulation they drew the model structures, as well as the parameters, from their posterior distributions.

Peterson (2004) and Miró-Quesada *et al.* (2004) proposed a Bayesian approach in which one calculates the posterior (given \mathbf{Y}) probability that y falls within some specified tolerance interval. The posterior distribution of $y | \mathbf{Y}$ considers uncertainty/randomness in \mathbf{w} , ε , $\boldsymbol{\theta}$, and σ . Peterson (2004) considered the noiseless case (i.e., terms involving \mathbf{w} absent from Equation (1)), and Miró-Quesada *et al.* (2004) extended the approach to include noise. Rajagopal and Del Castillo (2005) and Rajagopal *et al.* (2005) further extended the approach to incorporate uncertainty in the model structure, the former treating the noiseless case and the latter including noise. They used the approach of Box and Meyer (1993), also used by Chipman (1998), to calculate the posterior model probabilities. For the analyses with no noise, they utilized analytical expressions for the posterior distribution of $y | \mathbf{Y}$ (a t -distribution under a certain choice of priors). For the analyses with noise terms in the model, they relied heavily on Monte Carlo simulation to calculate the objective function for each \mathbf{x} of interest, as in Chipman (1998).

Relative to the aforementioned Bayesian approaches for robust design with noise, a primary advantage of the proposed approach is that it is possible to derive a relatively simple closed-form analytical expression for the objective function. As mentioned in the Introduction, the analytical expression provides insight into robustness issues; facilitates optimization; and offers a natural decomposition of the response variability into the standard robust design component (i.e., assuming $\boldsymbol{\theta}$ coincides with $\hat{\boldsymbol{\theta}}$) and the additional component due to parameter uncertainty. This allows one to conveniently plot the individual components versus \mathbf{x} . A plot of the additional variability due to parameter uncertainty is informative when deciding whether further experimentation is necessary to reduce parameter uncertainty, which is illustrated with examples later. Moreover, graphical exploration of each component plot is quite useful if there are other design considerations (qualitative or quantitative) that are difficult to incorporate into a formal mathematical optimization criterion, as is often the case in practical robust design problems.

Another obvious difference between the approach proposed in this article and the approaches of Peterson (2004) and Miró-Quesada *et al.* (2004) is that the underlying design criteria are quite different. When deciding which method is more appropriate, one should also consider

which criterion is more physically meaningful for the problem at hand. For some problems, minimizing the posterior MSE may be more meaningful than maximizing the posterior probability that y falls within some specified tolerance interval, and *vice versa* for other problems.

3. Prior and posterior distributions for the parameters

Let \mathbf{Y} denote the $n \times 1$ vector of observations of y obtained from an experiment, over which the \mathbf{x} and \mathbf{w} settings were varied according to some design matrix \mathbf{Z} . By this it is meant that for the n observations, the model (1) can be written as

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ is an $n \times 1$ Gaussian random vector with zero mean and covariance matrix $\sigma^2 \mathbf{I}_n$, and \mathbf{I}_n denotes the $n \times n$ identity matrix. If k denotes the dimension of the parameter vector $\boldsymbol{\theta}$, each column of the $n \times k$ matrix \mathbf{Z} corresponds to a single term in Equation (1). Each row of \mathbf{Z} consists of a "1" (corresponding to the intercept term α) and the values of $\{g_i(\mathbf{x}): i = 1, 2, \dots, l\}$, $\{w_j: j = 1, 2, \dots, m\}$ and $\{x_i w_j, i = 1, 2, \dots, p; j = 1, 2, \dots, m\}$ for a single experimental run.

For the linear Gaussian model a common choice of prior distributions, which will be assumed in this work, is an uninformative (locally flat) prior for $\log(\sigma)$ and $\boldsymbol{\theta} | \sigma \sim N_k(\boldsymbol{\mu}, \sigma^2 \boldsymbol{\Phi})$. In other words, the prior distribution of σ is $\propto 1/\sigma$, and the prior distribution of $\boldsymbol{\theta}$ given σ is multivariate normal with mean $\boldsymbol{\mu}$ (some specified $k \times 1$ vector) and covariance matrix $\boldsymbol{\Phi}$ (some specified $k \times k$ matrix). One often selects $\boldsymbol{\Phi}$ to be diagonal with diagonal entries $\{\phi_i: i = 1, 2, \dots, k\}$, in which case $\sigma^2 \phi_i$ is the prior variance of θ_i , the i th element of $\boldsymbol{\theta}$. Most of the prior Bayesian treatments of parameter uncertainty reviewed in Section 2 have assumed these priors. Notice that letting $\phi_i \rightarrow \infty$ represents minimal prior knowledge of θ_i .

Under these priors, it can be shown (Bunke and Bunke, 1986, p. 439) that the posterior distribution of $\boldsymbol{\theta}$ given \mathbf{Y} and σ is

$$\boldsymbol{\theta} | \mathbf{Y}, \sigma \sim N_k(\boldsymbol{\mu}_Y, \sigma^2 \boldsymbol{\Phi}_Y),$$

where $\boldsymbol{\mu}_Y = [\boldsymbol{\Phi}^{-1} + \mathbf{Z}'\mathbf{Z}]^{-1}[\boldsymbol{\Phi}^{-1}\boldsymbol{\mu} + \mathbf{Z}'\mathbf{Y}]$, and $\boldsymbol{\Phi}_Y = [\boldsymbol{\Phi}^{-1} + \mathbf{Z}'\mathbf{Z}]^{-1}$. It can also be shown that the posterior distribution of $\sigma^{-2} | \mathbf{Y}$ is gamma. The full posterior distributions will not be needed to derive an expression for $J_{\text{CRD}}(\mathbf{x})$, as will be seen in the following section. All that are needed are the posterior mean and covariance of $\boldsymbol{\theta} | \mathbf{Y}$ and the posterior mean of $\sigma^2 | \mathbf{Y}$. Because $\sigma^{-2} | \mathbf{Y}$ is gamma, it follows that the posterior mean of $\sigma^2 | \mathbf{Y}$ is (Bunke and Bunke, 1986, A 2.22)

$$\hat{\sigma}^2 = E_\sigma[\sigma^2 | \mathbf{Y}] = \frac{[\hat{\boldsymbol{\theta}} - \boldsymbol{\mu}]' \boldsymbol{\Phi}^{-1} [\hat{\boldsymbol{\theta}} - \boldsymbol{\mu}] + [\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\theta}}]' [\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\theta}}]}{n - 2}, \quad (5)$$

where

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= E_{\boldsymbol{\theta}, \sigma}[\boldsymbol{\theta} | \mathbf{Y}] = E_\sigma[E_{\boldsymbol{\theta}}[\boldsymbol{\theta} | \sigma, \mathbf{Y}] | \mathbf{Y}] = E_\sigma[\boldsymbol{\mu}_Y | \mathbf{Y}] = \boldsymbol{\mu}_Y \\ &= [\boldsymbol{\Phi}^{-1} + \mathbf{Z}'\mathbf{Z}]^{-1}[\boldsymbol{\Phi}^{-1}\boldsymbol{\mu} + \mathbf{Z}'\mathbf{Y}], \end{aligned} \quad (6)$$

and

$$\begin{aligned} \boldsymbol{\Sigma}_\theta &= E_{\boldsymbol{\theta}, \sigma}[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' | \mathbf{Y}] \\ &= E_\sigma[E_{\boldsymbol{\theta}}[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' | \sigma, \mathbf{Y}] | \mathbf{Y}] \\ &= E_\sigma[\sigma^2 \boldsymbol{\Phi}_Y | \mathbf{Y}] = \hat{\sigma}^2 \boldsymbol{\Phi}_Y = \hat{\sigma}^2 [\boldsymbol{\Phi}^{-1} + \mathbf{Z}'\mathbf{Z}]^{-1}, \end{aligned} \quad (7)$$

are the posterior mean and covariance of $\boldsymbol{\theta} | \mathbf{Y}$. Notice that with minimal prior knowledge of $\boldsymbol{\theta}$ (i.e., $\boldsymbol{\Phi}^{-1} \rightarrow \mathbf{0}_k$), Equations (6) and (7) reduce to the standard least squares parameter estimates and covariance matrix, respectively, albeit using a different estimate of σ^2 . For minimal prior knowledge of $\boldsymbol{\theta}$, Equation (5) reduces to the residual sum of squares $[\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\theta}}]' [\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\theta}}]$, divided by $n - 2$.

Joseph (2006) and Joseph and Delaney (2007) investigated an alternative choice of prior covariance for $\boldsymbol{\theta}$ that consisted of choosing $\boldsymbol{\Phi}$ so that the resulting prior distribution of the response is in agreement (over some full factorial grid in the design space) with a specified Gaussian random process model for the response. This may be useful in highly fractionated designs in which one prefers to retain many high-order terms in the model.

4. A closed-form expression for the CRD objective function

The results of the previous section yield a rather tractable expression for $J_{\text{CRD}}(\mathbf{x})$. Toward this end, let $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ denote the vector of parameter estimation errors, and let $\tilde{\alpha}$, $\tilde{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\gamma}}$, and $\tilde{\mathbf{B}}$ be defined similarly. Substituting $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + \tilde{\boldsymbol{\theta}}$ for the parameters in Equation (1) gives:

$$y - T = \{\hat{\alpha} + \hat{\boldsymbol{\beta}}' \mathbf{g}(\mathbf{x}) + \hat{\boldsymbol{\gamma}}' \mathbf{w} + \mathbf{w}' \hat{\mathbf{B}} \mathbf{x} - T\} + \{\tilde{\alpha} + \tilde{\boldsymbol{\beta}}' \mathbf{g}(\mathbf{x}) + \tilde{\boldsymbol{\gamma}}' \mathbf{w} + \mathbf{w}' \tilde{\mathbf{B}} \mathbf{x}\} + \varepsilon. \quad (8)$$

Since the posterior distribution of $\boldsymbol{\theta} | \mathbf{Y}$ is multivariate normal with mean $\hat{\boldsymbol{\theta}}$ and covariance $\boldsymbol{\Sigma}_\theta$, the posterior distribution of $\tilde{\boldsymbol{\theta}} | \mathbf{Y}$ is multivariate normal with mean zero and the same covariance. Moreover, $\tilde{\boldsymbol{\theta}} | \mathbf{Y}$ is independent of \mathbf{w} , which denotes some future noise that is independent of the experimental data \mathbf{Y} . Although $\tilde{\boldsymbol{\theta}} | \mathbf{Y}$ is not independent of $\boldsymbol{\varepsilon} | \mathbf{Y}$ (because their distributions both depend on σ), it is straightforward to show that they are uncorrelated, which is the property that is needed in the following.

Substituting Equation (8) into Equation (4) gives:

$$\begin{aligned} J_{\text{CRD}}(\mathbf{x}) &= (\hat{\alpha} + \hat{\boldsymbol{\beta}}' \mathbf{g}(\mathbf{x}) - T)^2 + (\hat{\boldsymbol{\gamma}} + \hat{\mathbf{B}} \mathbf{x})' \boldsymbol{\Sigma}_\alpha (\hat{\boldsymbol{\gamma}} + \hat{\mathbf{B}} \mathbf{x}) \\ &\quad + \boldsymbol{\Sigma}_\alpha + \mathbf{g}'(\mathbf{x}) \boldsymbol{\Sigma}_\beta \mathbf{g}(\mathbf{x}) + \mathbf{x}' \mathbf{A} \mathbf{x} + 2\mathbf{g}'(\mathbf{x}) \boldsymbol{\Sigma}_{\beta\alpha} \\ &\quad + 2\mathbf{x}' \mathbf{a} + d + \hat{\sigma}^2, \end{aligned} \quad (9)$$

where $\boldsymbol{\Sigma}_\alpha = E_{\boldsymbol{\theta}, \sigma}(\tilde{\alpha}^2 | \mathbf{Y})$, $\boldsymbol{\Sigma}_\beta = E_{\boldsymbol{\theta}, \sigma}(\tilde{\boldsymbol{\beta}} \tilde{\boldsymbol{\beta}}' | \mathbf{Y})$ and $\boldsymbol{\Sigma}_{\beta\alpha} = E_{\boldsymbol{\theta}, \sigma}(\tilde{\boldsymbol{\beta}} \tilde{\alpha} | \mathbf{Y})$ denote the posterior variances/covariances

of α and β , and we define $\mathbf{A} = E_{\theta, \sigma}(\tilde{\mathbf{B}}' \Sigma_w \tilde{\mathbf{B}} | \mathbf{Y})$, $d = E_{\theta, \sigma}(\tilde{\gamma}' \Sigma_w \tilde{\gamma} | \mathbf{Y})$, and $\mathbf{a} = E_{\theta, \sigma}(\tilde{\mathbf{B}}' \Sigma_w \tilde{\gamma} | \mathbf{Y})$. In arriving at terms like $\mathbf{x}' \mathbf{A} \mathbf{x}$ in Equation (9), we have used the relationship $E_{\mathbf{w}, \theta, \sigma}[(\mathbf{w}' \tilde{\mathbf{B}} \mathbf{x})^2 | \mathbf{Y}] = E_{\mathbf{w}, \theta, \sigma}(\mathbf{x}' \tilde{\mathbf{B}}' \mathbf{w} \mathbf{w}' \tilde{\mathbf{B}} \mathbf{x} | \mathbf{Y}) = E_{\theta, \sigma}(E_{\mathbf{w}}(\mathbf{x}' \tilde{\mathbf{B}}' \mathbf{w} \mathbf{w}' \tilde{\mathbf{B}} \mathbf{x} | \theta, \sigma, \mathbf{Y}) | \mathbf{Y}) = \mathbf{x}' E_{\theta, \sigma}(\tilde{\mathbf{B}}' \Sigma_w \tilde{\mathbf{B}} | \mathbf{Y}) \mathbf{x} = \mathbf{x}' \mathbf{A} \mathbf{x}$. The scalar d , the $p \times p$ matrix \mathbf{A} , and the $p \times 1$ vector \mathbf{a} can be readily constructed from the covariance matrices Σ_w and Σ_θ using the relationships $d = E_{\theta, \sigma}(\tilde{\gamma}' \Sigma_w \tilde{\gamma} | \mathbf{Y}) = \text{trace}(\Sigma_\gamma \Sigma_w)$, $A_{i,j} = E_{\theta, \sigma}(\tilde{\mathbf{b}}'_i \Sigma_w \tilde{\mathbf{b}}_j | \mathbf{Y}) = \text{trace}(\Sigma_{\mathbf{b}_i \mathbf{b}_j} \Sigma_w)$, and $a_i = E_{\theta, \sigma}(\tilde{\mathbf{b}}'_i \Sigma_w \tilde{\gamma} | \mathbf{Y}) = \text{trace}(\Sigma_{\mathbf{b}_i \gamma} \Sigma_w)$, where $\Sigma_{\mathbf{b}_i \mathbf{b}_j} = E_{\theta, \sigma}(\tilde{\mathbf{b}}_i \tilde{\mathbf{b}}'_j | \mathbf{Y})$ and $\Sigma_{\mathbf{b}_i \gamma} = E_{\theta, \sigma}(\tilde{\mathbf{b}}_i \tilde{\gamma}' | \mathbf{Y})$. Notice that Σ_α , Σ_β , $\Sigma_{\beta\alpha}$, Σ_γ , $\Sigma_{\mathbf{b}_i \mathbf{b}_j}$, and $\Sigma_{\mathbf{b}_i \gamma}$ are all directly available as submatrices of Σ_θ . Consequently, given $\mathbf{g}(\mathbf{x})$, $\hat{\theta}$, $\hat{\sigma}^2$, Σ_θ , Σ_w , and T , Equation (9) can be easily evaluated analytically, without the need for Monte Carlo simulation.

Equation (9) has a revealing interpretation in terms of the effects of parameter uncertainty on the posterior MSE. Since $E_{\varepsilon, \mathbf{w}, \theta, \sigma}(y | \mathbf{Y}) = \hat{\alpha} + \hat{\beta}' \mathbf{g}(\mathbf{x})$, the first term in Equation (9) is the component of the MSE due to differences between the posterior mean of y and the target. The second term represents the variance of y that is due to the random noise variables \mathbf{w} , under the assumption that the true parameters are equal to their estimates. This assumption is often referred to as *Certainty Equivalence* (CE) in the adaptive control literature. Borrowing this terminology, the analogous CE objective function is the familiar standard robust design expression:

$$J_{CE}(\mathbf{x}) = E_{\varepsilon, \mathbf{w}}[(y - T)^2 | \theta = \hat{\theta}, \sigma = \hat{\sigma}] = (\hat{\alpha} + \hat{\beta}' \mathbf{g}(\mathbf{x}) - T)^2 + (\hat{\gamma} + \hat{\mathbf{B}}' \mathbf{x})' \Sigma_w (\hat{\gamma} + \hat{\mathbf{B}}' \mathbf{x}) + \hat{\sigma}^2. \tag{10}$$

Notice that we can write Equation (9) as $J_{CRD}(\mathbf{x}) = J_{CE}(\mathbf{x}) + J_\theta(\mathbf{x})$, where

$$J_\theta(\mathbf{x}) = \Sigma_\alpha + \mathbf{g}'(\mathbf{x}) \Sigma_\beta \mathbf{g}(\mathbf{x}) + \mathbf{x}' \mathbf{A} \mathbf{x} + 2\mathbf{g}'(\mathbf{x}) \Sigma_{\beta\alpha} + 2\mathbf{x}' \mathbf{a} + d \tag{11}$$

represents the additional MSE due to parameter uncertainty. As parameter uncertainty decreases (i.e., as $\Sigma_\theta \rightarrow \mathbf{0}_k$), all of the terms in Equation (11) disappear. As parameter uncertainty increases, $J_\theta(\mathbf{x})$ increases, because each term in Equation (11) is proportional to elements of the posterior covariance Σ_θ .

5. A closed-form solution for \mathbf{x}_{CRD} with only linear effects

When $\mathbf{g}(\mathbf{x}) = \mathbf{x}$, the model (1) reduces to the linear effects model $y = \alpha + \beta' \mathbf{x} + \gamma' \mathbf{w} + \mathbf{w}' \mathbf{B} \mathbf{x} + \varepsilon$, in which case we can obtain a closed-form solution for the optimal CRD settings when there are no constraints on \mathbf{x} (constraints are discussed in Section 9). Although the linear model is not as broadly applicable as the model with quadratic $\mathbf{g}(\mathbf{x})$, the closed-form solution for the linear case provides insight into the nature of CRD.

Substituting $\mathbf{g}(\mathbf{x}) = \mathbf{x}$ in Equation (9) and setting the partial derivative equal to zero gives the optimal CRD input settings:

$$\mathbf{x}_{CRD} = [\hat{\beta} \hat{\beta}' + \hat{\mathbf{B}}' \Sigma_w \hat{\mathbf{B}} + \Sigma_\beta + \mathbf{A}]^{-1} \{ (T - \hat{\alpha}) \hat{\beta} - \hat{\mathbf{B}}' \Sigma_w \hat{\gamma} - \Sigma_{\beta\alpha} - \mathbf{a} \}. \tag{12}$$

In contrast, the input settings that minimize the analogous CE objective function (10) are

$$\mathbf{x}_{CE} = [\hat{\beta} \hat{\beta}' + \hat{\mathbf{B}}' \Sigma_w \hat{\mathbf{B}}]^{-1} \{ (T - \hat{\alpha}) \hat{\beta} - \hat{\mathbf{B}}' \Sigma_w \hat{\gamma} \}, \tag{13}$$

which follows from Equation (12) with all parameter covariance terms set equal to zero. Notice that if $p > m + 1$, the matrix $\hat{\beta} \hat{\beta}' + \hat{\mathbf{B}}' \Sigma_w \hat{\mathbf{B}}$ in Equation (13) is not invertible, and the inputs that minimize J_{CE} are not unique. In this case, replacing the inverse of $\hat{\beta} \hat{\beta}' + \hat{\mathbf{B}}' \Sigma_w \hat{\mathbf{B}}$ by its singular value decomposition pseudoinverse corresponds to taking \mathbf{x}_{CE} to be the minimum-norm solution. This minimum-norm solution is used for all of the CE examples in this article.

Comparing Equations (12) and (13), the origin of the term *cautious* in CRD becomes more apparent. Larger parameter uncertainty (as measured by, say, the eigenvalues of the positive semi-definite matrices Σ_β and \mathbf{A}) results in the inverse of the matrix in brackets in Equation (12), and thus \mathbf{x}_{CRD} , being smaller than if parameter uncertainty were neglected. Strictly speaking, larger parameter uncertainty causes \mathbf{x}_{CRD} to be closer to the center of the experimental design region, but this translates to \mathbf{x}_{CRD} being smaller if \mathbf{x} is coded so that the zero vector represents the center. In this sense, the optimal settings for \mathbf{x} are chosen more cautiously in CRD than if parameter uncertainty is neglected.

The effects of parameter uncertainty on \mathbf{x}_{CRD} are even more apparent in the special case that a fractional factorial design (with no aliasing of terms that are included in the model) is used and \mathbf{x} and \mathbf{w} are transformed to a scale for which they take on values of ± 1 over the experiment. With an orthogonal design matrix \mathbf{Z} , and assuming a non-informative prior for θ , the posterior covariance becomes $\Sigma_\theta = \hat{\sigma}^2 n^{-1} \mathbf{I}_k$. Thus, $\Sigma_{\beta\alpha}$ and \mathbf{a} are zero, $\Sigma_\beta = \hat{\sigma}^2 n^{-1} \mathbf{I}_p$ and $\mathbf{A} = \hat{\sigma}^2 n^{-1} \text{trace}(\Sigma_w) \mathbf{I}_p$, which, when substituted into Equation (12), yields:

$$\mathbf{x}_{CRD} = [\hat{\beta} \hat{\beta}' + \hat{\mathbf{B}}' \Sigma_w \hat{\mathbf{B}} + \hat{\sigma}^2 n^{-1} \{1 + \text{trace}(\Sigma_w)\} \mathbf{I}_p]^{-1} \times \{ (T - \hat{\alpha}) \hat{\beta} - \hat{\mathbf{B}}' \Sigma_w \hat{\gamma} \}.$$

When parameter uncertainty (as measured by $\hat{\sigma}^2 n^{-1}$, the posterior variance of the estimated parameters) is zero, \mathbf{x}_{CRD} coincides with \mathbf{x}_{CE} . As parameter uncertainty increases, the \mathbf{x}_{CRD} settings shrink monotonically toward $\mathbf{0}$, the center of the experimental design region.

Table 1. Description of variables in the leaf spring example

Variable	Represents	Level	
		Low	High
x_1	Temperature	1840	1880
x_2	Heating time	25	23
x_3	Hold down time	2	3
x_4	Transfer time	12	10
w_1	Quench temperature	130–150	150–170

6. An example

The use of the proposed approach is illustrated with data from an experiment involving the manufacture of truck leaf springs, originally analyzed in Pignatiello and Ramberg (1985) and later in Chipman (1998). There are four controllable variables and one noise variable, whose descriptions are given in Table 1. All temperatures are in degrees Fahrenheit, and all times are in seconds. The high and low values in Table 1 correspond to ± 1 values for \mathbf{x} and \mathbf{w} , which are expressed in coded units for the analyses in this article. The output variable y is the free height of the leaf spring, for which the target is $T = 8$ inches. The experiment was three replicates of a crossed-array design with a two-level fractional factorial in \mathbf{x} , the data for which are shown in Table 2. No quadratic effects can be estimated with this experiment, and, as shown in Chipman (1998), all \mathbf{x} -by- \mathbf{x} interactions appear insignificant. Hence, throughout the analysis, the linear effects model $y = \alpha + \beta' \mathbf{x} + \gamma w_1 + w_1 \mathbf{B} \mathbf{x} + \varepsilon$ discussed in Section 5 is used (i.e., Equation (1) with $\mathbf{g}(\mathbf{x}) = \mathbf{x}$ and $m = 1$).

Let $\phi_i \rightarrow \infty$ in order to represent minimal prior knowledge of θ , in which case Equations (5) and (6) yield the point estimates $\hat{\alpha} = 7.636$, $\hat{\beta} = [0.111, -0.088, -0.014, 0.052]'$, $\hat{\gamma} = -0.062$, $\hat{\mathbf{B}} = [0.016, 0.037, 0.005, -0.018]$, and $\hat{\sigma}^2 = (0.186)^2$. Since the design matrix was orthogonal, $\mathbf{Z}'\mathbf{Z} = n\mathbf{I}_{10} = 48\mathbf{I}_{10}$, and the parameter covariance becomes $\Sigma_\theta = \hat{\sigma}^2[\Phi^{-1} + \mathbf{Z}'\mathbf{Z}]^{-1} = n^{-1}\hat{\sigma}^2\mathbf{I}_{10} = (0.0268)^2\mathbf{I}_{10}$. The noise variance was assumed to be $\Sigma_w = 1$. Neglecting parameter uncertainty, the CE input settings from Equa-

Table 2. Response data for the leaf spring example

x_1	x_2	x_3	x_4	w_1					
				-1			1		
-1	-1	-1	-1	7.78	7.5	7.78	7.25	7.81	7.12
1	-1	-1	1	8.15	7.88	8.18	7.88	7.88	7.44
-1	1	-1	1	7.5	7.5	7.56	7.56	7.5	7.5
1	1	-1	-1	7.59	7.63	7.56	7.75	7.75	7.56
-1	-1	1	1	7.94	7.32	8	7.44	7.88	7.44
1	-1	1	-1	7.69	7.56	8.09	7.69	8.06	7.62
-1	1	1	-1	7.56	7.18	7.62	7.18	7.44	7.25
1	1	1	1	7.56	7.81	7.81	7.5	7.69	7.59

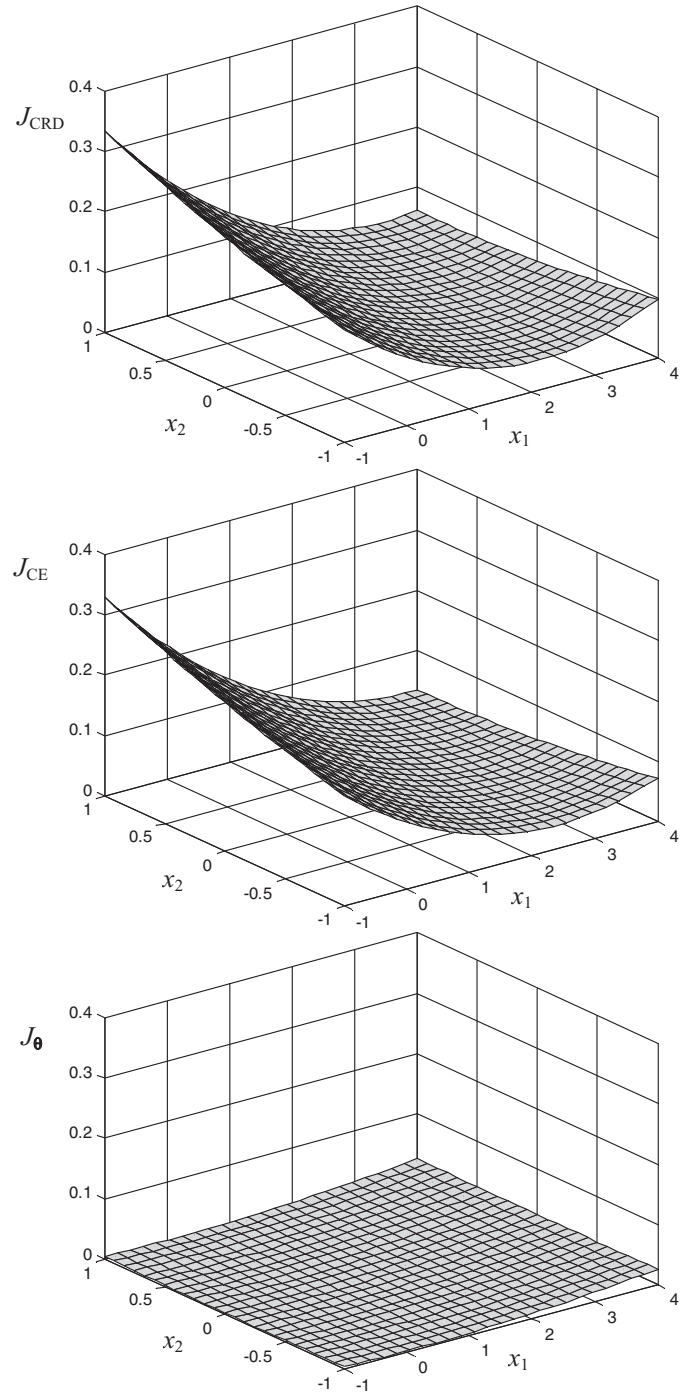


Fig. 1. Plots of J_{CRD} , J_{CE} , and J_θ versus the first two inputs for the leaf spring example. J_{CRD} results in a smaller optimal value for x_1 , for which the adverse effects of parameter uncertainty are lessened.

tion (13) are $\mathbf{x}_{CE} = [3.43, 0.24, -0.01, 0.09]'$. In comparison, the CRD input settings from Equation (12) are $\mathbf{x}_{CRD} = [2.51, -0.45, -0.10, 0.38]'$.

Figure 1 plots $J_{CRD}(\mathbf{x})$, $J_{CE}(\mathbf{x})$, and $J_\theta(\mathbf{x})$ versus x_1 and x_2 , with x_3 and x_4 held fixed at -0.10 and 0.38 , respectively (their optimal CRD values). Notice that as x_1 increases,

J_{CRD} increases more so than J_{CE} , because $J_{\theta}(\mathbf{x})$ increases as one attempts to extrapolate to values of x_1 beyond the experimental region. This is the primary net effect of parameter uncertainty in this example, and it has the beneficial consequence of helping to ensure that the CRD x_1 setting is closer to the center of the experimental region without adding an explicit constraint (explicit constraints will be discussed in Section 9). Recall that the CE setting for x_1 is 3.43, which is far outside the experimental region. The CRD setting for x_1 is 2.51, which, while still far enough outside the experimental region to cause concern, is substantially smaller than the CE setting. The posterior MSEs corresponding to the optimal CE and CRD input settings are $J_{CRD}(\mathbf{x}_{CE}) = 0.053$ and $J_{CRD}(\mathbf{x}_{CRD}) = 0.048$. Hence, a modest 10% improvement in the posterior MSE is achieved by taking into account parameter uncertainty when selecting the optimal input settings. For this example, there was a relatively large number of experimental runs ($n = 48$) and, correspondingly, a relatively small level of parameter uncertainty. In the next section it will be demonstrated that the differences between the CRD and CE are much more pronounced for larger parameter uncertainty.

7. Assessing the impact of parameter uncertainty and the need for further experimentation

For higher levels of parameter uncertainty than in the preceding example, the posterior MSE when using \mathbf{x}_{CRD} may be much lower than when using \mathbf{x}_{CE} , as will be demonstrated shortly. However, even if the CRD settings are used, one may find that the inflation of the MSE due to parameter uncertainty is still unacceptably large. The decomposition of the CRD objective function (9) into $J_{CRD}(\mathbf{x}) = J_{CE}(\mathbf{x}) + J_{\theta}(\mathbf{x})$, where $J_{CE}(\mathbf{x})$ and $J_{\theta}(\mathbf{x})$ are given by Equations (10) and (11), respectively, can aid in assessing whether this is the case. Recall that $J_{\theta}(\mathbf{x})$ represents the additional MSE due to parameter uncertainty and that when parameter uncertainty disappears, $J_{\theta}(\mathbf{x})$ reduces to zero, and $J_{CRD}(\mathbf{x})$ reduces to $J_{CE}(\mathbf{x})$. A simple plot of $J_{\theta}(\mathbf{x})$ versus \mathbf{x} , as in the bottom panel of Fig. 1, can be used to assess the net effect of parameter uncertainty in terms of its direct impact on the robust design objective. This, in turn, provides insight into whether additional experimentation is necessary to reduce parameter uncertainty, which is illustrated with a continuation of the leaf spring example.

Example continued: The $n = 48$ response observations in Table 2 represent three replicates of a 2^{5-1} fractional factorial design in \mathbf{x} and \mathbf{w} . Suppose, instead, that only a single replicate was conducted, resulting in only $n = 16$ runs, and that $\hat{\sigma}$ increased from 0.186 to 0.372 (i.e., doubled). Rather than arbitrarily choose one of the three replicates from Table 2 to retain, it will be simply assumed that the point estimates from the previous section ($\hat{\alpha} = 7.636$, $\hat{\beta} = [0.111, -0.088, -0.014, 0.052]'$, $\hat{\gamma} = -0.062$, $\hat{\mathbf{B}} = [0.016, 0.037, 0.005, -0.018]$ and $\hat{\sigma}^2 =$

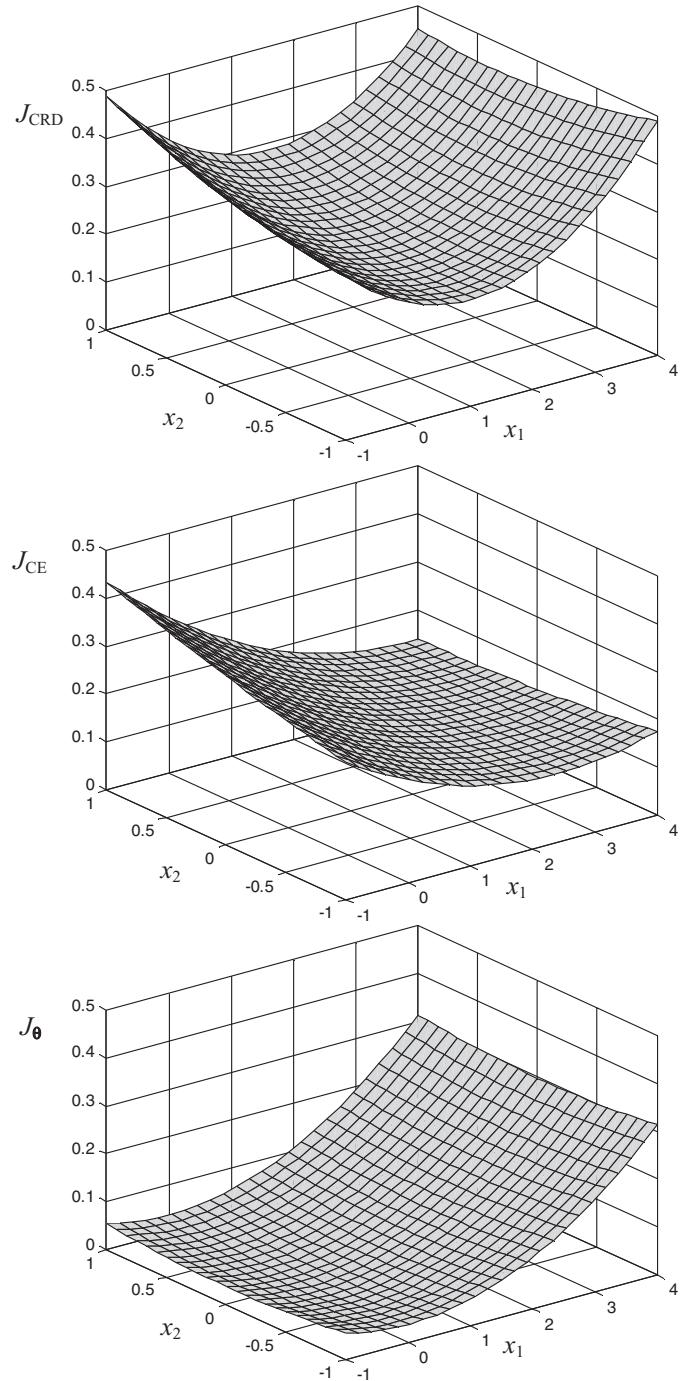


Fig. 2. Plots of J_{CRD} , J_{CE} , and J_{θ} versus the first two inputs for the modified leaf spring example with $n = 16$ and $\hat{\sigma} = 0.372$, instead of $n = 48$ and $\hat{\sigma} = 0.186$. The effects of parameter uncertainty now dominate, resulting in a more cautious CRD setting for x_1 .

$(0.372)^2$) came from a single replicate. In terms of the analysis for this case, the net effect is that the parameter covariance matrix increases by a factor of 12: From $\Sigma_{\theta} = (0.0268)^2 \mathbf{I}_{10}$ to $\Sigma_{\theta} = \hat{\sigma}^2 [\Phi^{-1} + \mathbf{Z}'\mathbf{Z}]^{-1} = n^{-1} \hat{\sigma}^2 \mathbf{I}_{10} = (0.0928)^2 \mathbf{I}_{10}$. The CE input settings, which neglect parameter uncertainty, remain unchanged from their

Downloaded By: [apley@northwestern.edu] At: 16:21 1 April 2011

earlier values: $\mathbf{x}_{CE} = [3.43, 0.24, -0.01, 0.09]'$. In contrast, the CRD input settings change from $\mathbf{x}_{CRD} = [2.51, -0.45, -0.10, 0.38]'$ (for the original Fig. 1 example) to $\mathbf{x}_{CRD} = [1.10, -0.66, -0.11, 0.40]'$ (for the example with smaller n and larger $\hat{\sigma}$). This has the benefit of further reducing the magnitude of the largest input x_1 to the point that it is almost within the experimental region.

Figure 2 plots $J_{CRD}(\mathbf{x})$ and its two components, $J_{CE}(\mathbf{x})$ and $J_{\theta}(\mathbf{x})$, versus x_1 and x_2 . The other two inputs, x_3 and x_4 , were held fixed at their optimal CRD values of -0.11 and 0.40 , respectively. A few points are worth noting. $J_{CRD}(\mathbf{x})$ increases dramatically as x_1 increases, because of the very pronounced component due to parameter uncertainty (J_{θ} , plotted in the bottom panel of Fig. 2). The contribution of $J_{\theta}(\mathbf{x})$ to the MSE is much larger for the higher levels of parameter uncertainty considered in this example, especially for large values of x_1 . Consequently, $J_{CRD}(\mathbf{x})$ penalizes large values of x_1 more than was the case for the Fig. 1 example, which explains why the optimal x_1 setting is substantially smaller in this case. The posterior MSEs corresponding to the optimal CE and CRD input settings are $J_{CRD}(\mathbf{x}_{CE}) = 0.360$ and $J_{CRD}(\mathbf{x}_{CRD}) = 0.219$. Thus, with the larger parameter uncertainty in this example, neglecting parameter uncertainty when selecting the inputs results in a 64% larger posterior MSE than when CRD is used.

At the optimal CRD inputs, $J_{CRD}(\mathbf{x}_{CRD}) = 0.219$ decomposes into its two components: $J_{CE}(\mathbf{x}_{CRD}) = 0.170$, and $J_{\theta}(\mathbf{x}_{CRD}) = 0.049$. The contribution of parameter uncertainty to the posterior MSE is roughly 29% of the CE contribution. Based on this, one might decide that further experimentation is required to reduce the parameter uncertainty. When assessing whether further experimentation is needed, one should also consider the relative contributions of the two components of J_{CRD} at the optimal CE inputs \mathbf{x}_{CE} . These are the input settings that would be used if there were no parameter uncertainty and the true parameters coincided with the point estimates. Because the point estimates are the posterior mean of θ , they do, after all, represent one's best guess at the true parameters. Following this line of reasoning, one might be interested in the following questions.

1. What benchmark MSE could be achieved in the hypothetical scenario that we know the true parameters and they happen to coincide with their point estimates?
2. How much will the reality of parameter uncertainty add to the MSE if we use the inputs optimized under the hypothetical benchmark scenario?

The answers to these two questions are precisely $J_{CE}(\mathbf{x}_{CE})$ and $J_{\theta}(\mathbf{x}_{CE})$, respectively. For the results shown in Fig. 2, we have $J_{CE}(\mathbf{x}_{CE}) = 0.138$ and $J_{\theta}(\mathbf{x}_{CE}) = 0.221$. This hypothetical benchmark MSE of 0.138 is better than $J_{CE}(\mathbf{x}_{CRD}) = 0.170$, the analogous value using the cautious inputs \mathbf{x}_{CRD} . Based on this, one might hesitate to so quickly rule out using the potentially very good \mathbf{x}_{CE} settings in favor of

the more-robust-to-parameter-uncertainty \mathbf{x}_{CRD} settings. However, neither should one just go ahead and use \mathbf{x}_{CE} , considering that parameter uncertainty at \mathbf{x}_{CE} could be expected to almost triple the MSE (from 0.138 to $0.138 + 0.221 = 0.360$). One might conclude from this analysis that further experimentation is necessary to reduce parameter uncertainty to levels at which one can use input settings with greater confidence.

Further experimentation also encompasses confirmation experiments, which are generally considered sound practice in any response surface optimization. In the preceding example, it had appeared that \mathbf{x}_{CE} may result in a lower MSE than \mathbf{x}_{CRD} if one neglects parameter uncertainty ($J_{CE}(\mathbf{x}_{CE}) = 0.138$ versus $J_{CE}(\mathbf{x}_{CRD}) = 0.170$). However, after performing the analyses described in the preceding paragraph, it is also clear that the MSE could in fact be substantially higher (e.g., 0.360 versus 0.138) because of parameter uncertainty if one uses \mathbf{x}_{CE} . Consequently, if one wished to entertain the notion of using \mathbf{x}_{CE} in hopes of achieving a lower MSE, then at the very least one should run a confirmation experiment at \mathbf{x}_{CE} . After running a confirmation experiment, the entire analysis should be repeated to update all relevant posterior distributions.

8. Parameter versus noise uncertainty

In the CRD paradigm, when formulating the objective function and solving for the optimal inputs, no distinction has been made between parameter uncertainty and noise uncertainty. However, the two forms of uncertainty are of course very different: The noise variables \mathbf{w} are true random variables and will vary from part-to-part or batch-to-batch. In contrast, the parameters θ are fixed (but unknown) variables. They have been assigned a probability distribution only as a convenient means of quantifying their uncertainty. Instead of considering the uncertainties in the noise and parameters on par in the CRD criterion, it is straightforward to keep them distinct by decomposing

$$J_{CRD}(\mathbf{x}) = J_{CE}(\mathbf{x}) + J_{\theta}(\mathbf{x}) = (\hat{\alpha} + \hat{\beta}'\mathbf{g}(\mathbf{x}) - T)^2 + (\hat{\gamma} + \hat{\mathbf{B}}\mathbf{x})'\Sigma_w(\hat{\gamma} + \hat{\mathbf{B}}\mathbf{x}) + J_{\theta}(\mathbf{x}) + \hat{\sigma}^2,$$

similar to what was done when plotting the individual components $J_{CE}(\mathbf{x})$ and $J_{\theta}(\mathbf{x})$ in Figs. 1 and 2.

The term $(\hat{\gamma} + \hat{\mathbf{B}}\mathbf{x})'\Sigma_w(\hat{\gamma} + \hat{\mathbf{B}}\mathbf{x})$ represents the contribution of noise variability to the MSE, whereas the term $J_{\theta}(\mathbf{x})$ represents the contribution of parameter uncertainty. The term $(\hat{\alpha} + \hat{\beta}'\mathbf{g}(\mathbf{x}) - T)^2$ represents the contribution of an off-target mean. These terms could all be plotted individually to understand their relative contributions to the MSE, keeping the effects of noise variability distinct from parameter uncertainty. Considering them together by minimizing $J_{CRD}(\mathbf{x})$ is simply a convenient means to mitigating the adverse impact of parameter uncertainty.

9. Input constraints and extrapolation beyond the experimental region

In robust design optimization, it is common to incorporate constraints on the inputs. A constraint can result if the inputs are physically or economically constrained to lie in a region; or they can result simply because the experimental region is finite, and extrapolation beyond the experimental region is viewed as risky. Regarding the latter, CRD inherently penalizes extrapolation beyond the experimental region, where the effects of parameter uncertainty are greater. However, constraining the inputs to be near the experimental region while optimizing the CE objective function may also possess an inherent form of caution.

To illustrate this, let $\mathbf{x}_{CE,con}$ denote the input settings that minimize $J_{CE}(\mathbf{x})$ under the constraint that \mathbf{x} lies within the experimental region (e.g., that each element of \mathbf{x} lies between -1 and 1). For the example of Fig. 2, numerical optimization reveals that $\mathbf{x}_{CE,con} = [1, -1, -1, 1]'$, for which the posterior MSE is $J_{CRD}(\mathbf{x}_{CE,con}) = 0.246$. This is only a modest 12% larger than the posterior MSE for the optimal CRD inputs [$J_{CRD}(\mathbf{x}_{CRD}) = 0.219$].

The examples of Figs. 1 and 2 are regular orthogonal designs for which $\Sigma_{\beta\alpha}$ and \mathbf{a} are zero and, hence, the effect of parameter uncertainty is the lowest at the origin $\mathbf{x} = \mathbf{0}$ (see Equation (11)). In this case, minimizing J_{CE} while constraining \mathbf{x} to be closer to the origin inherently results in more cautious input settings than if J_{CE} is minimized without constraints. Furthermore, suppose one conducted a ridge analysis in which J_{CE} is optimized under the constraint that $J_{\theta} = \lambda$ for a range of values $\lambda > 0$. Because the CRD solution lies somewhere along the ridge path, there exists some $\lambda > 0$ for which the constrained CE optimization solution coincides with the CRD solution.

The situation becomes more nuanced for non-orthogonal designs. Consider a further modification of the leaf spring example with everything the same as in the example in Fig. 2, except that only $n = 13$ runs are conducted. Suppose that the three omitted runs (relative to the example in Fig. 2, for which $n = 16$) are three of the four runs at the $\{x_1, x_2\} = \{1, -1\}$ corner. In particular, suppose the omitted runs are $\{x_1, x_2, x_3, x_4, w\} = \{1, -1, -1, 1, -1\}$, $\{1, -1, 1, -1, -1\}$, and $\{1, -1, -1, 1, 1\}$. In this case, the optimal CRD input settings are $\mathbf{x}_{CRD} = [0.62, -0.08, 0.17, 0.38]'$, for which $J_{CRD}(\mathbf{x}_{CRD}) = 0.278$. Figure 3 plots $J_{CRD}(\mathbf{x})$, $J_{CE}(\mathbf{x})$, and $J_{\theta}(\mathbf{x})$, versus x_1 and x_2 , with x_3 and x_4 held fixed at their optimal CRD values of 0.17 and 0.38, respectively. Notice that the effects of parameter uncertainty are much higher at the $x_1, x_2 = 1, -1$ corner in the bottom panel of Fig. 3.

In comparison, the optimal constrained CE inputs are $\mathbf{x}_{CE,con} = [1, -1, -1, 1]'$ (from numerical optimization), for which the posterior MSE is $J_{CRD}(\mathbf{x}_{CE,con}) = 0.413$. This is almost 50% larger than the posterior MSE for the optimal CRD inputs [$J_{CRD}(\mathbf{x}_{CRD}) = 0.278$]. The reason why simply

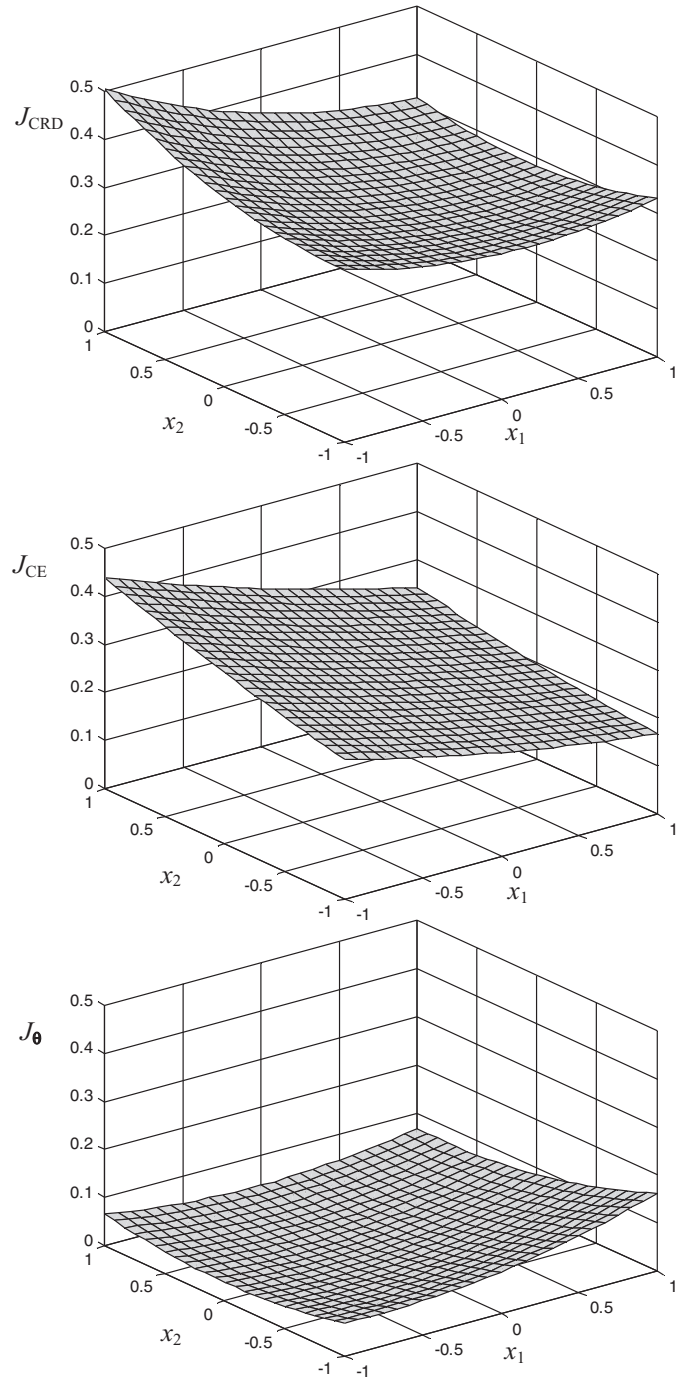


Fig. 3. Plots of J_{CRD} , J_{CE} , and J_{θ} versus the first two inputs for the modified leaf spring example with $n = 13$ and $\hat{\sigma} = 0.372$. Because there was only one run at $\{x_1, x_2\} = \{1, -1\}$, there is higher uncertainty at this corner.

constraining the CE inputs to the experimental region did not provide sufficient caution in this case is that the optimal constrained CE settings happened to fall in the corner of the experimental region for which the effects of parameter uncertainty were large (the three omitted runs were

deliberately chosen so as to create this scenario). In situations like this, in which the design matrix is not orthogonal, and the effects of parameter uncertainty are larger in certain directions of the input space, CRD automatically accounts for the nuanced characteristics of the parameter uncertainty.

Inclusion of model structure uncertainty within the CRD framework, as discussed in the following section, would tend to further penalize extrapolation outside the experimental region. More generally, it would penalize choosing input settings that are far from design points. Consider the two-level fractional factorial design with no center points that was used in the example of Fig. 2. Under the linear model assumption, the smallest uncertainty is at the origin (see Equation (11) and the bottom panel of Fig. 2), even though there were no design points there. However, if one considers the possible presence of a quadratic term, then the uncertainty would be much larger at the origin. Similarly, the possible presence of quadratic terms would greatly inflate the uncertainty as one extrapolates outside the range of the experimental region.

10. Consideration of model structure uncertainty

Following the approach of Chipman (1998), which was also adopted by Apley and Kim (2002), Kim (2002), Rajagopal and Del Castillo (2005), Rajagopal *et al.* (2005), and Ng (2010) when considering model structure uncertainty in robust design, the CRD results can be extended to account for model structural uncertainty. Let $\{S_1, S_2, \dots, S_q\}$ denote the set of all candidate model structures, where each model structure consists of subsets of the $g_i(\mathbf{x})$, w_j , and $x_i w_j$ interaction terms in Equation (1). Under certain assumptions on the prior probabilities, the posterior probabilities $\{\pi_1, \pi_2, \dots, \pi_q\}$ that each model structure holds can be calculated in much the same manner as the posterior distributions for the parameters (refer to Box and Meyer (1993) or Chipman (1998) for details). The CRD strategy would be to select the \mathbf{x} settings to minimize the weighted sum:

$$J(\mathbf{x}) = \sum_{i=1}^q \pi_i J_i(\mathbf{x}),$$

where $J_i(\mathbf{x}) = E_{\varepsilon, \mathbf{w}, \theta, \sigma}[(y - T)^2 | S_i]$ is the MSE from Equation (5) under the assumption that the model structure S_i holds. Since some of the models will exclude subsets of the controllable inputs, uncontrollable noise, and/or their interactions, one must be careful in summing the $J_i(\mathbf{x})$ across the different models in any optimization algorithm. The most straightforward way to do this is to include all of the input and noise variables in each $J_i(\mathbf{x})$. If S_i excludes a particular main effect or interaction term, the corresponding element of $\hat{\theta}$ (and the variance/covariance for that parameter) would be set equal to zero when forming $J_i(\mathbf{x})$.

11. A dual-response version of CRD

The CRD problem has been formulated as minimizing the single MSE criterion $E_{\varepsilon, \mathbf{w}, \theta, \sigma}[(y - T)^2 | \mathbf{Y}]$. It is straightforward to extend the CRD approach to the analogous dual response criteria in which $\text{Var}_{\varepsilon, \mathbf{w}, \theta, \sigma}[y | \mathbf{Y}]$ is minimized subject to the constraint $E_{\varepsilon, \mathbf{w}, \theta, \sigma}[y | \mathbf{Y}] = T$. Then, it can be written that

$$E_{\varepsilon, \mathbf{w}, \theta, \sigma}[y | \mathbf{Y}] = \hat{\alpha} + \hat{\beta}' \mathbf{g}(\mathbf{x}),$$

and

$$\begin{aligned} \text{Var}_{\varepsilon, \mathbf{w}, \theta, \sigma}[y | \mathbf{Y}] &= E_{\varepsilon, \mathbf{w}, \theta, \sigma}[(y - T)^2 | \mathbf{Y}] - (E_{\varepsilon, \mathbf{w}, \theta, \sigma}[y | \mathbf{Y}] - T)^2 \\ &= J_{\text{CRD}}(\mathbf{x}) - (\hat{\alpha} + \hat{\beta}' \mathbf{g}(\mathbf{x}) - T)^2 \\ &= (\hat{\gamma} + \hat{\mathbf{B}}\mathbf{x})' \Sigma_{\mathbf{w}} (\hat{\gamma} + \hat{\mathbf{B}}\mathbf{x}) + \Sigma_{\alpha} + \mathbf{g}'(\mathbf{x}) \Sigma_{\beta} \mathbf{g}(\mathbf{x}) \\ &\quad + \mathbf{x}' \mathbf{A} \mathbf{x} + 2\mathbf{g}'(\mathbf{x}) \Sigma_{\beta\alpha} + 2\mathbf{x}' \mathbf{a} + d + \hat{\sigma}^2, \end{aligned} \quad (14)$$

where Equation (9) has been used for $J_{\text{CRD}}(\mathbf{x})$. For this formulation, the use of Lagrange multipliers when performing the constrained optimization may be helpful. For the special case that $\mathbf{g}(\mathbf{x}) = \mathbf{x}$, the variance expression is quadratic in \mathbf{x} , and the constraint is linear. In this case, using Lagrange multipliers, it is straightforward to show (see Kim (2002)) that the closed-form solution is

$$\mathbf{x}_{\text{CRD, dual}} = \frac{\mathbf{D}^{-1} \hat{\beta} (T - \hat{\alpha} + \mathbf{z}' \mathbf{D}^{-1} \hat{\beta})}{\hat{\beta}' \mathbf{D}^{-1} \hat{\beta}} - \mathbf{D}^{-1} \mathbf{z}, \quad (15)$$

where $\mathbf{D} = \hat{\mathbf{B}}' \Sigma_{\mathbf{w}} \hat{\mathbf{B}} + \Sigma_{\beta} + \mathbf{A}$, and $\mathbf{z} = \mathbf{a} + \Sigma_{\beta\alpha} + \hat{\mathbf{B}}' \Sigma_{\mathbf{w}} \hat{\gamma}$.

The dual-response CRD approach bears some resemblance to the two frequentist approaches mentioned in Section 2 for taking parameter uncertainty into account in dual-response robust design. As a dual-response objective function, Miró-Quesada and Del Castillo (2004) proposed using an unbiased estimate of $\text{Var}_{\hat{\theta}, \mathbf{w}}(\hat{y}(\mathbf{w}) | \theta, \sigma)$, where $\hat{y}(\mathbf{w}) = \hat{\alpha} + \hat{\beta}' \mathbf{g}(\mathbf{x}) + \hat{\gamma}' \mathbf{w} + \mathbf{w}' \hat{\mathbf{B}} \mathbf{x}$. Their objective reduces to minimizing $(\hat{\gamma} + \hat{\mathbf{B}}\mathbf{x})' \Sigma_{\mathbf{w}} (\hat{\gamma} + \hat{\mathbf{B}}\mathbf{x}) + \mathbf{g}'(\mathbf{x}) \Sigma_{\hat{\beta}} \mathbf{g}(\mathbf{x})$, where the parameter estimates and covariance matrix $\Sigma_{\hat{\beta}}$ are from standard least squares. For comparison purposes, suppose a non-informative prior is assumed in the CRD approach, so that the posterior parameter estimates and covariance matrix coincide with their least squares counterparts. The objective function of Miró-Quesada and Del Castillo (2004) is missing a number of terms related to parameter uncertainty that are present in the CRD objective function (14). In particular, the quantity $\mathbf{x}' \mathbf{A} \mathbf{x} + 2\mathbf{x}' \mathbf{a} + d$ is missing. Consequently, although their approach considers uncertainty in β in the same manner as does CRD, it does not take into account uncertainty in \mathbf{B} . It is worth noting that the quantity $\mathbf{x}' \mathbf{A} \mathbf{x} + 2\mathbf{x}' \mathbf{a} + d$ appears at one point in the derivations of Miró-Quesada and Del Castillo (2004), but it is later subtracted out as a result of the fact that it is precisely the bias correction suggested by Myers and

Montgomery (2002, p. 576) when estimating the response variance.

To contrast the two approaches, reconsider the example of Fig. 3, discussed in Section 9. From Equation (15), the CRD inputs are $\mathbf{x}_{\text{CRD,dual}} = [2.10, -0.86, 0.40, 1.18]'$, for which the posterior MSE is $J_{\text{CRD}}(\mathbf{x}_{\text{CRD,dual}}) = 0.480$. In comparison, the optimal inputs using the criterion of Miró-Quesada and Del Castillo (2004), which are given by Equation (15) with \mathbf{A} , \mathbf{a} , and $\Sigma_{\beta\alpha}$ all set equal to zero, are $\mathbf{x}_{\text{MdC}} = [2.17, -0.85, 0.31, 1.00]'$, for which the posterior MSE is $J_{\text{CRD}}(\mathbf{x}_{\text{MdC}}) = 0.481$. Two things are apparent. First, the dual-response solution of Miró-Quesada and Del Castillo (2004) is very similar to the dual-response CRD solution for this example. Second, both of the dual-response solutions call for much larger input settings and incur substantially higher posterior MSE than the non-dual CRD solution. Recall that $\mathbf{x}_{\text{CRD}} = [0.62, -0.08, 0.17, 0.38]'$, for which $J_{\text{CRD}}(\mathbf{x}_{\text{CRD}}) = 0.278$. Evidently, even though parameter uncertainty is incorporated into the objective function (14), the hard constraint that the posterior response mean $\hat{\alpha} + \hat{\beta}'\mathbf{g}(\mathbf{x})$ equals the target can outweigh the penalty that Equation (14) places on using input settings for which the effects of parameter uncertainty are large.

Myers and Montgomery (2002, p. 576) recommend minimizing $(\hat{\mathbf{Y}} + \hat{\mathbf{B}}\mathbf{x})'\Sigma_w(\hat{\mathbf{Y}} + \hat{\mathbf{B}}\mathbf{x}) - \mathbf{x}'\mathbf{A}\mathbf{x} - 2\mathbf{x}'\mathbf{a} - d + \hat{\sigma}^2$, which they derive as an unbiased estimate of $\text{Var}_{\varepsilon,w}(y|\theta, \sigma)$. Notice the minus sign on the three bias correction terms. From a frequentist perspective, the minus signs are reasonable if the objective is purely to obtain an unbiased estimate of the response variance. From a Bayesian perspective, however, it is counterproductive as a strategy for minimizing the response variance in a manner that takes into account parameter uncertainty. It would tend to call for input settings that make $\mathbf{x}'\mathbf{A}\mathbf{x} + 2\mathbf{x}'\mathbf{a}$ larger, which, from Equation (14), serves to increase the posterior response variance, rather than decrease it. Even from a frequentist perspective, it has some undesirable properties as an objective function. Because \mathbf{A} is positive semi-definite, including the $-\mathbf{x}'\mathbf{A}\mathbf{x}$ term in the objective function will tend to call for *larger* \mathbf{x} settings than if the standard variance objective function $(\hat{\mathbf{Y}} + \hat{\mathbf{B}}\mathbf{x})'\Sigma_w(\hat{\mathbf{Y}} + \hat{\mathbf{B}}\mathbf{x}) + \hat{\sigma}^2$ is used. In fact, if parameter uncertainty in \mathbf{B} is large enough that $\hat{\mathbf{B}}'\Sigma_w\hat{\mathbf{B}} - \mathbf{A}$ has a negative eigenvalue, their objective function to be minimized (the unbiased expression for the variance) is unbounded from below, which will call for infinitely large \mathbf{x} settings. The cautious approach, on the other hand, tends to call for *smaller* \mathbf{x} settings as parameter uncertainty increases, which seems a more intuitively appealing way to mitigate the adverse effects of parameter uncertainty.

12. Conclusions

This article has investigated a Bayesian MSE objective function as a means of taking parameter uncertainty into

account in robust design optimization. A key feature of this approach is that a tractable, closed-form expression for the CRD objective function is obtained as a function of the input settings. The only posterior information needed to calculate the CRD objective function is the posterior mean (i.e., point estimates) and posterior covariance matrix of the parameters, for which simple expressions have been provided in Section 3. The resulting CRD objective function (9) is a quadratic function of $\mathbf{g}(\mathbf{x})$ and \mathbf{x} and, hence, is a well-behaved one for numerical optimization, at least for typical $\mathbf{g}(\mathbf{x})$ considered in robust design studies.

The term cautious robust design is fitting. As has been demonstrated, it tends to call for more cautious \mathbf{x} settings than if parameter uncertainty is neglected. By *cautious* it is meant \mathbf{x} settings at which the effects of parameter uncertainty are mitigated. For rotatable, orthogonal designs, for which parameter uncertainty is the same in all directions of the parameter space, the cautious settings amount to ones that are closer to the center of the experimental design region. On the other hand, for non-orthogonal experimental designs such as in the example of Fig. 3, the cautious nature of the CRD solution is more nuanced. The effect of parameter uncertainty is larger in certain directions or regions of the input space, and CRD automatically takes this into account by selecting \mathbf{x} settings that avoid this.

It has been shown that the CRD objective function, which is a Bayesian MSE, decomposes naturally into three components: the first component is the square of the response mean deviation from target; the second component is the certainty equivalence variance—i.e., what would result due to noise and random error variability if there were no parameter uncertainty; and the third component is the additional variance due to the uncertainty in the parameters, characterized by their posterior distribution. It has been demonstrated that this decomposition is particularly useful in determining whether further experimentation is necessary to reduce parameter uncertainty. The assessment of parameter uncertainty is entirely objective oriented, in the sense that the third component quantifies the extent to which parameter uncertainty inflates the MSE objective function. The current authors believe that this is a very direct and appropriate way of quantifying parameter uncertainty and facilitating efficient experimentation.

A dual-response version of CRD has also been investigated. In situations with no parameter uncertainty, a dual-response criterion may have conceptual advantage over a single MSE criterion, since the former constrains the response mean to be on target. In the presence of parameter uncertainty, however, the dual-response constraint that the posterior mean is on target is less meaningful: Enforcing the constraint $\hat{\alpha} + \hat{\beta}'\mathbf{g}(\mathbf{x}) = T$ will generally not ensure that the actual response mean $\alpha + \beta'\mathbf{g}(\mathbf{x})$ equals the target for a specific realization of α and β . Ironically, it could happen that the response mean using the CRD MSE criterion is closer to the target than when the constraint $\hat{\alpha} + \hat{\beta}'\mathbf{g}(\mathbf{x}) = T$ is

enforced. Furthermore, it was demonstrated in the example that enforcing the hard constraint that $\hat{\alpha} + \hat{\beta}'\mathbf{g}(\mathbf{x}) = T$ results in settings that are substantially less robust to parameter errors. For these reasons, the authors believe that the CRD MSE criterion of Equation (9) is preferable to the dual-response criterion when parameter uncertainty is large.

Acknowledgements

This work was supported in part by the National Science Foundation under grant CMMI-0758557. We also thank two anonymous referees and the Department Editor, Russell Barton, for their many helpful comments.

References

- Apley, D.W. (2004) A cautious minimum variance controller with ARIMA disturbances. *IIE Transactions*, **36**, 417–432.
- Apley, D.W. and Kim, J.B. (2002) A cautious approach to robust design with model uncertainty. In *Proceedings of the 2002 Industrial Engineering Research Conference*, IIE, Orlando, FL, paper 2164.
- Apley, D.W. and Kim, J.B. (2004) Cautious control of industrial process variability with uncertain input and disturbance model parameters. *Technometrics*, **46**(2), 188–199.
- Åström, K.J. and Wittenmark, B. (1995) *Adaptive Control*, second edition. Addison-Wesley, New York, NY.
- Box, G.E.P. and Hunter, J.S. (1954) A confidence region for the solution of a set of simultaneous equations with an application to experimental design. *Biometrika*, **41**, 190–199.
- Box, G.E.P. and Meyer, R.D. (1993) Finding the active factors in fractional screening experiments. *Journal of Quality Technology*, **25**, 94–105.
- Bunke, H. and Bunke, O. (1986) *Statistical Inference in Linear Models: Statistical Methods of Model Building, Volume I*, Wiley, New York, NY.
- Chipman, H. (1998) Handling uncertainty in analysis of robust design experiments. *Journal of Quality Technology*, **30**, 11–17.
- Joseph, V.R. (2006) A Bayesian approach to the design and analysis of fractionated experiments. *Technometrics*, **48**, 219–229.
- Joseph, V.R. and Delaney, J.D. (2007) Functionally induced priors for the analysis of experiments. *Technometrics*, **49**, 1–11.
- Khattree, R. (1996) Robust parameter design: a response surface approach. *Journal of Quality Technology*, **28**, 187–198.
- Kim, J.B. (2002) A cautious approach to minimizing industrial process variability. Ph.D. dissertation, Department of Industrial Engineering, Texas A&M University.
- Lucas, J.M. (1994) How to achieve a robust process using response surface methodology. *Journal of Quality Technology*, **26**, 248–260.
- Miró-Quesada, G. and Del Castillo, E. (2004) Two approaches for improving the dual response method in robust parameter design. *Journal of Quality Technology*, **36**, 15–168.
- Miró-Quesada, G., Del Castillo, E. and Peterson, J. (2004) A Bayesian Approach for multiple response surface optimization in the presence of noise variables. *Journal of Applied Statistics*, **31**, 251–270.
- Monroe, E.M., Pan, R., Anderson-Cook, C.M., Montgomery, D.C. and Borror, C.M. (2010) Sensitivity analysis of optimal designs for accelerated life testing. *Journal of Quality Technology*, **42**(2), 121–135.
- Montgomery, D.C. (2001) *Design and Analysis of Experiments*, fifth edition, John Wiley & Sons, New York, NY.
- Myers, R.H., Khuri, A.I. and Vining, G. (1992) Response surface alternatives to the Taguchi robust parameter design approach. *The American Statistician*, **46**, 131–139.
- Myers, R.H. and Montgomery, D. (2002) *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons, New York, NY.
- Nair, V.N. (1992) Taguchi's parameter design: a panel discussion. *Technometrics*, **34**, 128–161.
- Ng, S.H. (2010) A Bayesian model-averaging approach for multiple-response optimization. *Journal of Quality Technology*, **42**, 52–68.
- Peterson, J.J. (2004) A posterior predictive approach to multiple response surface optimization. *Journal of Quality Technology*, **36**, 139–153.
- Peterson, J.J., Cahya, S. and Del Castillo, E. (2002) A general approach to confidence regions for optimal factor levels of response surfaces. *Biometrics*, **58**, 422–431.
- Pignatiello, J.J. and Ramberg, J. S. (1985) Discussion of off-line quality control, parameter design, and the Taguchi method. *Journal of Quality Technology*, **17**, 198–206.
- Rajagopal, R. and Del Castillo, E. (2005) Model-robust process optimization using Bayesian model averaging. *Technometrics*, **47**, 152–163.
- Rajagopal, R., Del Castillo, E. and Peterson, J.J. (2005) Model and distribution-robust process optimization with noise factors. *Journal of Quality Technology*, **37**, 210–222.
- Sahni, N.S., Piepel, G.F. and Naes, T. (2009) Product and process improvement using mixture-process variable methods and robust optimization techniques. *Journal of Quality Technology*, **41**(2), 181–197.
- Shoemaker, A.C., Tsui, K. and Wu, C.F.J. (1991) Economical experimentation methods for robust design. *Technometrics*, **33**, 415–427.
- Taguchi, G. (1986) *Introduction to Quality Engineering*, UNIPUB/Kraus International, White Plains, NY.
- Vining, G.G. and Myers, R.H. (1990) Combining Taguchi and response surface philosophies: a dual response approach. *Journal of Quality Technology*, **22**, 38–45.
- Wu, C.F.J. and Hamada, M. (2000) *Experiments: Planning, Analysis, and Parameter Design Optimization*, John Wiley & Sons, New York, NY.

Biographies

Daniel W. Apley is an Associate Professor of Industrial Engineering and Management Sciences at Northwestern University, Evanston, IL. He obtained B.S., M.S., and Ph.D. degrees in Mechanical Engineering and an M.S. degree in Electrical Engineering from the University of Michigan. His research interests lie at the interface of engineering modeling, statistical analysis, and data mining, with particular emphasis on manufacturing variation reduction applications in which very large amounts of data are available. His research has been supported by numerous industries and government agencies. He received the NSF CAREER award in 2001, the IIE Transactions Best Paper Award in 2003, and the Wilcoxon Prize for best practical application paper appearing in *Technometrics* in 2008. He currently serves as Editor-in-Chief for the *Journal of Quality Technology* and has served as Chair of the Quality, Statistics & Reliability Section of INFORMS, Director of the Manufacturing and Design Engineering Program at Northwestern, and Associate Editor for *Technometrics*.

Jeongbae Kim is the Director of the Service Innovation Team at Korea Telecom Headquarters. He obtained his Ph.D. in Industrial Engineering from Texas A&M University.