ADVANCES IN DESIGN AUTOMATION
ASME, September 1993

A ROBUST OPTIMIZATION PROCEDURE WITH VARIATIONS ON DESIGN VARIABLES AND CONSTRAINTS

Sivakumar Sundaresan, Kosuke Ishii, and Donald R. Houser**
Department of Mechanical Engineering
The Ohio State University
206 W 18th Ave.
Columbus, OH 43210-1107

** Currently at Eastman Kodak, Rochester, NY

ABSTRACT

This paper describes a procedure that incorporates manufacturing and operational variances to achieve designs with robust and optimal performance. The procedure optimizes the expected value of a performance characteristic subject to a set of constraints. It uses concepts from statistical design of experiments to approximate the expected value of a performance characteristic. The procedure incorporates uncertainties in design variables and variations in constraints due to uncertainty in design variables. This paper discusses the following three methods to incorporate variations in constraints: 1) A method using heuristics that evaluates constraints at the worst combinations of design variables, 2) A method with built-in constraint variation that models constraints using first order Taylor expansion, and 3) A method based on differentiating KKT optimality conditions. The design of spur and helical gears with minimum transmission error serves as the target application. The key gear design research issue is to determine the optimal combination of geometric design variables like number of teeth, pressure angle that minimizes transmission error subject to constraints like minimum number of teeth to avoid undercut and maximum bending stress.

NOMENCLATURE

F	Statistical objective function
f	Performance function
f_c	Performance at the target design
f_i	Performance at the worst combination of design variables
g	Inequality constraint
h	Equality constraint
L	Lagrangian
n	Number of design variables
N_p	Number of teeth in the pinion
p(x)	Probability density function of variable x
P_{nd}	Normal generating diametral pitch
S	Solution space
T	Tolerance space
W	Corner space
x	Design variable
$\mathbf{x}$	Vector of n design variables
x^+	Upper bound for x
x^-	Lower bound for x
$\mathbf{x}_t$	Target design vector
y_i	Performance values
α	Weighting factor
Δx	Variation in design variable x
$\Delta\mathbf{x}$	Vector of variations in all variables
ϕ_n	Normal generating pressure angle
λ	Lagrange multiplier for inequality constraints
λ^*	Optimum Lagrange multiplier
#FE	Number of function evaluations
#GE	Number of constraint evaluations
LDP	Load Distribution Program
OA	Orthogonal Array
PPTE	Peak to Peak Transmission Error
SI	Sensitivity Index

1. INTRODUCTION

One of the important tasks in engineering design is to account for manufacturing errors and operational variances. Traditionally, engineers conduct sensitivity analysis after design optimization. Taguchi (1978 and 1987) introduced the concept of *parameter design* to improve the quality of a product whose manufacturing process involves significant variability or "noise."

The parameter design concept "reduces variation in performance by reducing the sensitivity of an engineering design to sources of variation rather than controlling the sources."

Taguchi's techniques were based on direct experimentation. However, designers often use computer programs to evaluate a performance function instead of actual experiments. Ragsdell and d'Entremont (1988) developed a non-linear code that applies Taguchi's concepts to design optimization. In some cases, optimal design is the least robust, and designers have to make a trade-off between target performance and robustness. Ideally, one should optimize the expected value of a performance function when there are variations/uncertainties in design variables.

One of the earliest forms of sensitivity analysis was performed using Lagrange multipliers. Lagrange multipliers give a unit change in the optimal function for an unit change in the constraints. Fiacco (1968) and Sobieski et al. (1982) provide methods for evaluating sensitivities of the optimal objective function and optimal design to changes in design variables and parameters that were kept constant during optimization. These methods were based on differentiating Karush-Kuhn-Tucker (KKT) conditions of optimality with respect to the fixed design parameters and design variables. Vanderplaats (1985) also provided a method to estimate the sensitivity of the optimum design to changes in fixed parameters based on the method of feasible directions. In this method, the fixed parameter is modeled as an additional design variable, and hence this method is also called Extended Design Space (EDS) method. One advantage of this method over methods based on KKT optimality conditions is that this method does not require second derivatives of the objective function and constraints.

Beltracchi and Gabriele (1988a and b) presented a method based on Recursive Quadratic Programming (RQP) to estimate sensitivity without having to evaluate the second derivatives of the objective function. In their method, for each parameter (p), they performed two optimizations using RQP method at the upper (p^+) and the lower (p^-) variations of the design variables to evaluate the optimum functions f^+ and f^- respectively and evaluated the sensitivity derivative using a central difference scheme. Beltracchi and Gabriele (1988b) briefly discussed the merits and demerits of each of these methods. However, in statistical optimization, we are interested not only in sensitivity derivatives but also in finding the feasible combination of design variables that minimize objective function and its sensitivity to design parameters. There is a need to incorporate the calculation of sensitivity derivatives during the optimization procedure rather than *after* optimization.

To minimize a function with variations in constraints, Parkinson et. al. (1990) advocate a two-step solution method. The first step addresses the optimization problem with only the nominal constraints and variation (Æb) and does not include the function variations. The optimum achieved at the first step is called nominal optimum. The function variations are evaluated at the nominal optimum. The statistical optimum is assumed to be close to the nominal optimum and hence they assumed the constraint variations to be a constant in the domain containing both the peak and statistical optimum. The second step involves optimizing the problem with constant constraint variation built into the model. A major limitation in this two-step procedure is

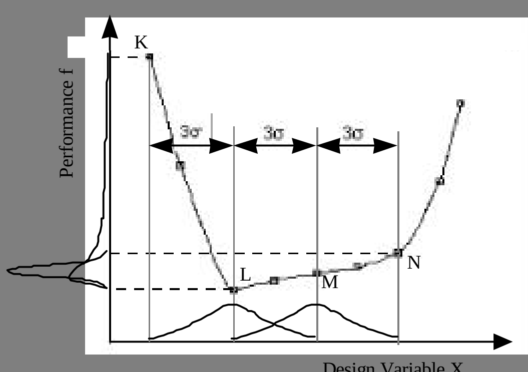

FIG. 1 CONCEPT OF STATISTICAL OPTIMUM

Points L and M are the peak and statistical optimum respectively. If the target design is point L, a deviation 3σ in the X could change function f to a value corresponding to point K. However, if point M is the target value, the change in function f is relatively small as function values at points L, M and N are nearly equal. Hence, the peak optimum may not be the best when one considers deviations of 3σ which results in drastic deterioration of function f. When functions of two variables are considered, designers should be interested in flat surfaces near the peak optimum.

The goal in statistical optimization is to minimize the expected value of a function. Let $f(x_1, x_2, \ldots x_n)$ be a function of n variables $x_1, x_2, \ldots x_n$ that needs to be optimized. Let $p(x_1, x_2, \ldots x_n)$ be a joint probability density function associated with **n** random variables. The expected value of a function $f(x_1, x_2, \ldots x_n)$ is given by

$$E[f(x_1, x_2, \ldots x_n)] = \int_{x_1^-}^{x_1^+} \int_{x_2^-}^{x_2^+} \ldots \int_{x_n^-}^{x_n^+} f(x_1, x_2, \ldots x_n)\, p(x_1, x_2, \ldots x_n)\, dx_1\, dx_2 \ldots dx_n \qquad (1)$$

where x_i^- and x_i^+ represent the lower and upper bounds for the ith random variable.
If n random variables are independent,

$$E[f(x_1, x_2, \ldots x_n)] = \int_{x_1^-}^{x_1^+} \int_{x_2^-}^{x_2^+} \ldots \int_{x_n^-}^{x_n^+} f(x_1, x_2, \ldots x_n)\, p_1(x_1)\, p_2(x_2) \ldots p_n(x_n)\, dx_1\, dx_2 \ldots dx_n \qquad (2)$$

In discrete form, equation (2) can be written as

$$E[f(x_1, x_2, \ldots x_n)] = \sum_{x_1^-}^{x_1^+} \sum_{x_2^-}^{x_2^+} \ldots \sum_{x_n^-}^{x_n^+} f(x_1, x_2, \ldots x_n)\, p_1(x_1)\, p_2(x_2) \ldots p_n(x_n) \qquad (3)$$

Equation (3) gives the weighting factor used based on probability for each performance value $f(x_1, x_2, \ldots x_n)$ to calculate the expected value of a function. In most engineering problems, the probability distribution function for each variable is unknown. In cases where function $f(x_1, x_2, \ldots x_n)$ is evaluated using computer programs, using equation (3) would be computationally time consuming. One needs a method to approximate equation (3) by an expression that will be easier to evaluate when using performance simulation programs. Sundaresan et al. (1991a) evaluated the function at all the worst combinations of design variables and defined a root mean square value of the difference between the function at the worst cases and the function value at the target as a measure of the sensitivity. Equation (4) gives the sensitivity index (SI) as the measure of sensitivity.

$$SI = \sqrt{\frac{1}{4} \sum_{i=1}^{4} (f_i - f_c)^2} \qquad (4)$$

The objective function (F) during statistical optimization is a linear weighted function of the targeted performance(f_c) and the sensitivity index. Equation (5) gives the objective function.

$$F = \alpha f_c + (1-\alpha)\, SI \qquad 0 \le \alpha \le 1 \qquad (5)$$

In this technique, the number of function evaluations increases exponentially as the number of variables increases.

This deficiency can be overcome by applying concepts from statistical design of experiments in defining the sensitivity index of a design. Since equation (4) considers all the worst combinations of design variables, we refer to it as a full factorial design. The use of fractional factorials instead of full factorials reduces the number of function evaluations in computing the sensitivity index.

The orthogonal arrays proposed by Taguchi help in designing our fractional factorials and assume negligible interactions between some design parameters. The L4 orthogonal arrays require 4 function evaluations and accommodate up to 3 design parameters. Equation (4) corresponds to a L4 array applied to 2 design parameters. Using the L4 array for three parameters saves 4 function evaluations compared to the full factorial design. For problems with more than 3 but less than 8 design parameters, the L8 orthogonal array is suitable.

After evaluating the function at n worst combinations of design variables, equation (6) yields the sensitivity index. Here, we normally set k=2. Equation (4) defines SI using f_i and f_c values while equation (6) uses only f_i values. Since the value of SI is independent of the value f_c, the weight factor α can take values between 0 and 1. The function F defined by equation (5) serves as the objective function during optimization. Higher values of k penalize larger values of fi.

$$SI = \left(\frac{1}{n} \sum_{i=1}^{n} f_i^k \right)^{\frac{1}{k}} \tag{6}$$

3. CONSTRAINED STATISTICAL OPTIMIZATION

3.1 Background

The technique presented in section 2 evaluated the expected value of the objective function by applying statistical design of experiments concepts. One could also use the same concepts to evaluate the expected value of a constraint. A general non-linear statistical optimization problem may be stated as

Minimize $E[F(\mathbf{x})]$

subject to $E[g_j(\mathbf{x}, b_j)]$ ² 0.0 j=1,2......m

$E[h_k(\mathbf{x}, b_k)] = 0.0$ k=1,2......l

$\mathbf{x} \in S$

where $S = \{ \mathbf{x} : \mathbf{x} \in R^n \mid \mathbf{x}^l \le \mathbf{x} \le \mathbf{x}^u \}$

$\mathbf{x} = [x_1, x_2, \ldots x_n]^T$

$\mathbf{x}^l = [x_1^l, x_2^l, \ldots x_n^l]^T \quad \mathbf{x}^u = [x_1^u, x_2^u, \ldots x_n^u]^T$

with variations Δx_i and Δb_j in x_i, (i=1, 2,..n) and b_j respectively. (7)

The use of *expected values* for a design constraint implies that the constraints are satisfied in a probabilistic sense. In a strict sense, the constraints are violated by some worst combinations of the design variable. For an inequality constraint, $E[g_j(\mathbf{x}, bj)]$ ² 0.0 would mean that approximately 50 % of the time, the design constraint would be violated. Instead, one prefers the worst combinations of design variables to satisfy the design constraints. Hence, there is a need to define when a constraint is active or violated for statistical optimization. This leads us to the following definitions of design spaces and criticality of a constraint:

Definition 1: Tolerance Space/Corner Space

Tolerance space (T) is a set of points close to the target design point where each point represents a possible combination of design variables due to uncertainties in each design variable. The shaded rectangle in Fig. 2 represents the tolerance space for 2-D problems. The target design point is represented by a small rectangle in the center of the shaded rectangle. Each target design point $\mathbf{x}_t = [x_{1t}, x_{2t}, \ldots x_{nt}]^T \in S$ has a tolerance space associated with it.

$$T(\mathbf{x}_t) = \{ \mathbf{x} : \mathbf{x} \in S \mid |\mathbf{x} - \mathbf{x}_t| \le \mathbf{Dx} \}$$

where $\mathbf{Dx} = [\Delta x_1, \Delta x_2, \ldots \Delta x_n]^T$ (8)

The corner space (W) consists only of corner vertices of the tolerance space. In Fig. 2, the 4 corner vertices of the shaded rectangle are the elements of the corner space.

$$W(\mathbf{x}_t) = \{ \mathbf{x} : \mathbf{x} \in S \mid |\mathbf{x} - \mathbf{x}_t| = \mathbf{Dx} \} \tag{9}$$

Theorem 1

If a constraint is monotonic with respect to all design variables in the tolerance space, then the maximum constraint value will occur at one of the corner points.

Proof:

Consider a vector $\mathbf{x}$ of two design variables (x_1, x_2) with target values $\mathbf{x}_t$=(x_{1t}, x_{2t}). Assume that constraint $g \cong g(x_1, x_2)$ is monotonically increasing in x_1 and monotonically decreasing in x_2. Since g is monotonically increasing in x_1 and monotonically decreasing in x_2, the following equations are satisfied:

$$g(x_1^+, x_2) \ \ge \ g(x_1, x_2) \quad \forall (x_1, x_2) \in T(x_{1t}, x_{2t}) \tag{11a}$$

$$g(x_1, x_2^-) \ \ge \ g(x_1, x_2) \quad \forall (x_1, x_2) \in T(x_{1t}, x_{2t}) \tag{11b}$$

Substituting for $x_2 = x_2^-$ in eq.(11a) and using eq.(11b) we get

$$g(x_1^+, x_2^-) \ \ge \ g(x_1, x_2^-) \ \ge \ g(x_1, x_2) \quad \forall (x_1, x_2) \in T(x_{1t}, x_{2t}) \tag{12}$$

which proves that the maximum occurs at one of the vertices. The proof of this theorem can be extended to n design variables easily.

Definition 2: Statistically Active Constraint

An i^{th} inequality constraint is considered active when the value of the constraint is zero at some of the worst combinations of design variables and negative at the target design and at the remaining worst combinations of design variables.

$$\text{For } \mathbf{x}_t \in S,\quad g_i \text{ is statistically active if}$$
$$1)\quad g_i(\mathbf{x}_t) < 0.0$$
$$2)\quad \exists\ \mathbf{x} \in W(\mathbf{x}_t) \text{ s.t. } g_i(\mathbf{x}) = 0.0$$
$$3)\quad \forall\ \mathbf{x} \in T(\mathbf{x}_t)\quad g_i(\mathbf{x}) \leq 0.0 \qquad (13)$$

Definition 3: Quasi-Active Constraint:

An i^{th} inequality constraint is considered quasi-active when value of the constraint is less than zero at the target design point and positive for some worst combinations of design variables.

$$\text{For } \mathbf{x}_t \in S,\quad g_i \text{ is quasi-active if}$$
$$1)\quad g_i(\mathbf{x}_t) < 0.0$$
$$2)\quad \exists\ \mathbf{x} \in T(\mathbf{x}_t) \text{ s.t. } g_i(\mathbf{x}) > 0.0 \qquad (14)$$

Definition 4: Peak Active constraint

An i^{th} inequality constraint is considered peak-active when the value of a constraint is greater than zero for some of the worst combinations of design variables and zero for the target design point. This would be an outcome in an optimization process if we did not model variations in constraints due to the deviations in design variables.

$$\text{For } \mathbf{x}_t \in S,\quad g_i \text{ is peak-active if} \qquad (15)$$
$$1)\quad g_i(\mathbf{x}_t) = 0.0$$

Definition 5: Quasi- Violated Constraint

An i^{th} inequality constraint is considered quasi-violated when the target design point violates the constraint and some of the worst combinations of design variables do not.

$$\text{For } \mathbf{x}_t \in S,\quad g_i \text{ is quasi-violated if}$$
$$1)\quad g_i(\mathbf{x}_t) > 0.0$$
$$2)\quad \exists\ \mathbf{x} \in T(\mathbf{x}_t) \text{ s.t. } g_i(\mathbf{x}) \leq 0.0 \qquad (16)$$

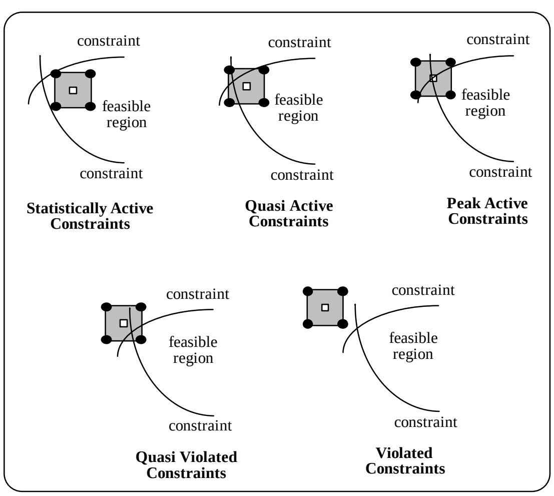

FIG. 2 CRITICALITY OF CONSTRAINTS

Definition 6: Violated Constraint

An i^{th} inequality constraint is considered violated when all the points in the tolerance space violate the constraint.

$$\text{For } \mathbf{x}_t \in S,\quad g_i \text{ is violated if} \qquad (17)$$
$$g_i(\mathbf{x}) \geq 0.0\ \ \forall\ \mathbf{x} \in T(\mathbf{x}_t)$$

Fig. 2 shows the various cases of criticality for inequality constraints. For an equality constraint, variation in the design variable would cause a violation of the constraint. Hence we do not describe any special definitions for activity of an equality constraint. For an equality constraint, the procedure will assure that the target design point *satisfies* the equality constraint. The next section presents three techniques for statistical optimization to compute a **feasible**, **robust** and **optimal** combination of design variables.

3.3 Procedure

3.3.1 Method 1: Using heuristics. This method is a single step optimization procedure used to achieve feasible robust optimum. This method evaluates the objective function and constraints at the worst combinations of design variables. A constraint ($g_j(\mathbf{x})$ - b_j ² 0.0) during peak optimization would be modeled during statistical optimization as

$$\text{Max}\left\{\left[g_j(\mathbf{y}) - b_j\right],\ \forall\, \mathbf{y} \in W(\mathbf{x})\right\} \leq 0.0 \qquad (18)$$

The expected value of objective function would be evaluated using Taguchi's orthogonal arrays as discussed in section 2. The problems solved during statistical optimization can be stated as

$$\begin{aligned} &\text{Minimize} && E\,[\,F(\mathbf{x})\,] \\ &\text{Subject to} && \text{Max}\left\{\left[g_j(\mathbf{y}) - b_j\right],\ \forall\, \mathbf{y} \in W(\mathbf{x})\right\} \leq 0.0 \\ & && \mathbf{x} \in S \\ & && \text{where } S = \left\{\mathbf{x} : \mathbf{x} \in R^n \ \middle|\ \mathbf{x}^l \leq \mathbf{x} \leq \mathbf{x}^u\right\} \\ & && \mathbf{x} = [x_1, x_2, \ldots x_n]^T \\ & && \mathbf{x}^l = [x_1^l, x_2^l, \ldots x_n^l]^T \quad \mathbf{x}^u = [x_1^u, x_2^u, \ldots x_n^u]^T \end{aligned} \qquad (19)$$

with variations Δx_i and Δb_j in x_i, (i=1, 2,..n) and b_j respectively.

This method, equivalent to brute force technique, is very simple to implement in a computer program because it does not require second derivatives of either the objective function or the constraints. One of the drawbacks of this method is the need to evaluate the constraints at all the worst combinations of design variables. This method is suitable for problems where evaluation of the constraints is not computationally time consuming. For problems with n variables, one would have to evaluate constraints at 2^n worst combinations of design variables.

In order to reduce the number of constraint evaluations during optimization for problems with a large number of design variables, one could use Taguchi's orthogonal arrays and evaluate constraints at only a fraction of the worst combinations of the design variables. After evaluating the constraints at the selected worst cases, we use a statistical technique (Phadke, 1991) to predict which other worst cases would show larger constraint values and evaluate the constraint only at those predicted worst cases. Theorem 1 shows that if one assumes monotonicity in the tolerance space, one of the worst combinations of design variables will exhibit the maximum constraint value. This technique is similar but superior to evaluating gradients of constraints using finite difference methods because the finite difference technique changes one variable at a time and thus lacks information on interactions between design variables.

In our applications, the evaluation of constraints (bending and contact stresses) takes less than 0.1% of the time it takes to evaluate a performance function. The solution of the optimization problem usually has 2 to 3 active constraints. Method 1 is suitable for such applications.

3.3.2 <u>Method 2: Constraints with Built-in Constraint Variation</u>. This method is an off-shoot of the brute force method. Instead of using orthogonal arrays to determine which worst combination of design variables has a larger constraint value, we use the gradient information to evaluate the worst possible value of the constraint. The maximum possible variation in constraint g due to deviation Δx_i in design variables x_i (i=1, 2, ...n) is approximated using equation (20).

$$\Delta g_j \quad = \quad \sum_{i=1}^{n} \left| \frac{\partial g_j}{\partial x_i} \Delta x_i \right| \tag{20}$$

In this method, we add this variation to the constraint formulation and solve the following optimization problem:

$$\begin{aligned} &\text{Minimize} \quad E\,[\,F(\mathbf{x})\,] \\ &\text{Subject to} \quad g_j\left(\mathbf{x}, b_j\right) + \sum_{i=1}^{n} \left| \frac{\partial g_j}{\partial x_i} \Delta x_i \right| \quad \le \quad 0.0 \\ &\mathbf{x} \in S \\ &\text{where} \quad S = \left\{ \mathbf{x} : \mathbf{x} \in R^n \;\middle|\; \mathbf{x}^l \le \mathbf{x} \le \mathbf{x}^u \right\} \\ &\mathbf{x} = \left[x_1, x_2, \ldots x_n \right]^T \\ &\mathbf{x}^l = \left[x_1^l, x_2^l, \ldots x_n^l \right]^T \quad \mathbf{x}^u = \left[x_1^u, x_2^u, \ldots x_n^u \right]^T \end{aligned} \tag{21}$$

One could also replace the constraint in the above equation by

$$g_j\left(\mathbf{x}, b_j\right) + \sqrt{\sum_{i=1}^{n} \left(\frac{\partial g_j}{\partial x_i} \Delta x_i \right)^2} \quad \le \quad 0.0 \tag{22}$$

3.3.3 <u>Method 3: Approach using KKT Condi-tions</u>. This method is based on differentiating KKT conditions of optimality with respect to design variables. In this method, the statistical optimization procedure consists of two steps. The first step involves finding the robust optimum without considering any variation in the constraints. However, due to inherent variations in the constraints, the optimum achieved at this stage is infeasible. The second step involves finding the change in the optimum design variables for a change in the constraints.

The first step solves a constrained optimization problem with objective function E[F($\mathbf{x}$)], where E[F($\mathbf{x}$)] is the expected value of a performance characteristic. The procedure uses Taguchi orthogonal arrays to approximate the expected value of a performance characteristic. Algorithms like Augmented Lagrange Multiplier (ALM) methods, along with BFGS (Vanderplaats, 1984) for search directions, can be used to find the constrained optimum. After the first step, optimum design variables and Lagrange multipliers associated with active constraints are available. However, if the first step does not yield Lagrange multipliers, one can estimate the Language multipliers by

$$\Lambda = -\left[A^T A \right]^{-1} A^T \nabla_x F \tag{23}$$

Where Λ is a vector of Lagrange multipliers and matrix A contains the gradient information of the active constraints. Element A_{ij} of matrix A is given by

$$A_{ij} = \frac{\partial g_i}{\partial x_j} \tag{24}$$

The second step involves calculating the following:

$$\frac{\partial \mathbf{x}^*}{\partial g_j} \quad \text{and} \quad \frac{\partial \lambda^*}{\partial g_j} \qquad j=1,2,.....m \tag{25}$$

Since for active constraints $g_j = b_j$, derivatives in eq. (25) are equal to

$$\frac{\partial \mathbf{x}^*}{\partial b_j} \quad \text{and} \quad \frac{\partial \lambda^*}{\partial b_j} \qquad j=1,2,.....m \tag{26}$$

In equations (25) and (26), λ^* and $\mathbf{x}^*$ are the Lagrange multipliers and the constrained optimum based on the Lagrangian is given by equation (27).

$$L(\mathbf{x}, \lambda, b) = F(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j \, g_j\left(\mathbf{x}, b_j\right) \tag{27}$$

To calculate the partial derivatives $\nabla_b \mathbf{x}^*$ and $\nabla_b \lambda^*$ in equation (26), this method differentiates the KKT conditions with respect to b_j. This ensures that KKT conditions are satisfied for various values of b_j. The KKT conditions are:

$$\nabla_x L = \nabla_x F + \sum_{j=1}^{m} \lambda_j \nabla_x g_j\left(\mathbf{x}, b_j\right) = 0 \tag{28}$$

$$\lambda_j \, g_j\left(\mathbf{x}, b_j\right) = 0.0 \qquad \text{For } j=1,2,.....\ m \tag{29}$$

Differentiating Kuhn-Tucker conditions with respect to b,

$$\begin{bmatrix} \nabla_x^2 L & \nabla_x g_1^T & \nabla_x g_2^T & -- & \nabla_x g_m^T \\ l_1 \nabla_x g_1 & g_1 & 0 & -- & 0 \\ l_2 \nabla_x g_2 & 0 & g_2 & -- & 0 \\ -- & -- & -- & -- & -- \\ l_m \nabla_x g_m & 0 & -- & -- & g_m \end{bmatrix} \begin{bmatrix} \nabla_b \mathbf{x}^T \\ \nabla_b l_1^T \\ \nabla_b l_2^T \\ -- \\ \nabla_b l_m^T \end{bmatrix} = - \begin{bmatrix} \nabla_{xb}^2 L \\ l_1 \nabla_b g_1^T \\ l_2 \nabla_b g_2^T \\ -- \\ l_m \nabla_b g_m^T \end{bmatrix} \tag{30}$$

Since $\nabla_x L$ is not an explicit function of b_j, $\nabla^2_{xb} L$ is a zero matrix. Since only active constraints are considered, $g_j = g_j(\mathbf{x}) - b_j = 0.0$

$$\begin{bmatrix} \nabla^2_x L & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \nabla_b \mathbf{x}^T \\ \nabla_b \lambda^T \end{bmatrix} = \begin{bmatrix} 0 \\ I \end{bmatrix} \quad (31)$$

where A is a mxn matrix with each A_{ij} term given by equation (24), I is an identity matrix of size mxm and Hessian $\nabla^2_x L$ is given by

$$\nabla^2_x L = \frac{\partial^2 L}{\partial x_i \partial x_j} \quad (32)$$

After solving for $\nabla_b \mathbf{x}$ and $\nabla_b \lambda$, one can evaluate the change in $\mathbf{x}^*$ and λ^* for changes in constraint g using

$$\Delta\lambda_j = -\sum_{k=1}^{m} \frac{\partial \lambda_j^*}{\partial b_k} \Delta g_k \qquad \Delta x_j = -\sum_{k=1}^{m} \frac{\partial x_j^*}{\partial b_k} \Delta g_k \quad (33)$$

where the variation in the constraint Æg_i is given by equation (20).

The new optimum is given by

$$\mathbf{x}^*_{new} = \mathbf{x}^*_{old} + \mathbf{Dx} \quad (34)$$

This approach reduces the number of constraint evaluations during optimization, because it does not consider the variation in the constraints in the first step. However, this technique fails if one of the inactive constraints is violated in the new optimum.

3.3.4 <u>Discussion</u>. The first method based on brute force with heuristics and the second method based on first order approximation of the constraint variation are ideal for problems where evaluation of the constraints is easy and not time-consuming. The third method, based on differentiating KKT conditions, is suitable for problems where constraint evaluation is computationally expensive. However, the third method is a two-step procedure where the method assumes that there is no change in the set of active constraints during the second step. This problem, however, does not occur in the first two methods. In cases where the evaluation of constraints involve analysis procedures (like FEM) that are time consuming, it is recommended that one should come up with simple algebraic equations by using regression analysis (Bhakuni, 1991). Using curve-fitted equations significantly reduces computing time.

4. SPUR GEAR DESIGN EXAMPLE

This section applies the methods discussed in section 3 to design spur gear meshes with minimum transmission error. In recent years, researchers have established a relationship between static transmission error and gear noise. Studies typically employ analytical tools to predict static transmission error. This example uses one such tool called the Load Distribution Program which evaluates transmission error as a function of the elastic properties of the gear mesh and errors in the gear tooth profile. Welbourn (1979) defined transmission error as "the difference between the actual position of the output gear and the position it would occupy if the gear drive is perfect (infinite stiffness and conjugate teeth)." This example characterizes transmission error by its peak to peak value (PPTE). In this example, the aim is to design a spur gear mesh with minimum transmission error and insensitivity to errors in pressure angle and uncertainty in number of teeth due to its discrete values.

The problem specifications used are:
1) Transmitted torque of 1250 in-lbs
2) Center distance of 3.5 inches
3) Gear ratio of unity

During optimization, the following variables were varied:
1) **Number of teeth** (N_p)was continuously varied between 20 and 80.
2) **Generating pressure angle** (ϕ_n) was varied between 10 and 25 degrees.

The following parameters were assumed during optimization:
1) The face width equals 1.0 for both the pinion and the gear.
2) The tooth thicknesses of the pinion and the gear were reduced by $0.0025/P_{nd}$ at the standard pitch diameter to account for positive backlash.
3) This example considered only standard gears with the operating pitch diameter equal to the standard pitch diameter. The addendum coefficient of the pinion/gear was 1.0.
4) This example used standard hobs with hob addendum coefficient of 1.25 and hob tip radius coefficient of 0.157.
5) The manufacturing variation in pressure angle is 0.3 degrees, and uncertainty in the number of teeth is 0.5

An unique combination of number of teeth and pressure angle represents a target gear design. The procedure evaluated PPTE for the target gear mesh. The procedure computed sensitivity index using eq. (6) by evaluating PPTE at the 4 worst cases (due to variations in pressure angle and uncertainty in number of teeth) given by the L4 orthogonal array. We assumed α to be 0.5 and used eq. (5) to evaluate the objective function.

The optimization procedure considered the following constraints:
1) Minimum number of teeth to avoid undercut. This is a function of pressure angle, hob shift to account for backlash, hob addendum and hob tip radius.
2) Minimum top land thickness for the pinion of $0.3/P_{nd}$, where P_{nd} is the normal generating diametral pitch.
3) Minimum top land thickness for the gear of $0.3/P_{nd}$.
4) Bending stress of the pinion should be less than 30 Ksi. The procedure evaluated the bending stress using the AGMA geometry factor method (Errichello, 1981).

Fig. 3 shows contour plots of equal objective function values for various combinations of number of teeth and pressure angle. The two dark lines represent the bending stress constraint and the constraint that represents the minimum number of teeth to avoid undercutting. The optimum point (P) lies at the intersection of these two constraints. The optimization process without modeling variations in the constraints, converged to point P (peak optimum).

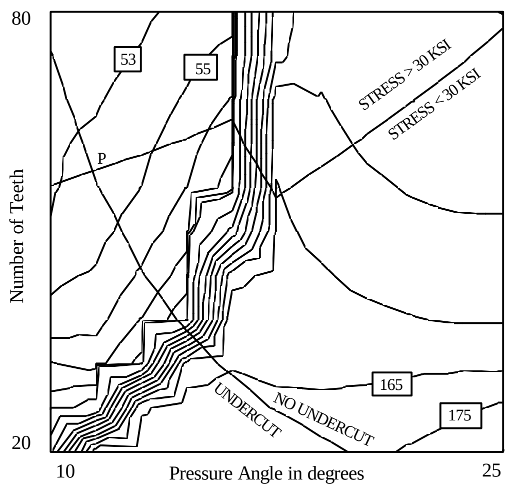

FIG. 3 CONSTRAINTS AND FUNCTION CONTOURS

Table 1 tabulates the results of the three methods discussed in Section 3 where #FE and #GE represent the number of function and constraint evaluations. In method 1, we evaluated at all of the 4 (2^2) worst combinations of design variables, and hence the statistical optimum achieved in method 1 is **the optimum** and constraints were **statistically active**. Two cases were run for method 2, one using eq. (21) and the other using eq. (22). Method 2, using eq. (21) converged to an optimum close to the one achieved in method 1. The results using method 2 with eq. (22) were in between the peak and the statistical optimum. The result using method 3 is very close to the optimum in method 1. However, this is coincidental and one cannot conclude that method 3 will always yield better results than method 2. The number of function evaluations (# F.E) is nearly the same for all the methods but method 3 requires less number of constraint evaluations (# G.E).

This example treated the number of teeth to be a continuous variable and the optimum value for number of teeth was 57.84. However, in practice, a designer will choose the number of teeth as 58. Since, the uncertainty in the number of teeth chosen was 0.5, 58 will lie within the tolerance space and constraints will not be violated.

TABLE 1 RESULTS OF VARIOUS METHODS

METHODS USED	FUNC-TION	N_p	ϕ_n	# FE	# GE
Peak Optimum (Point P)	53.416	58.21	11.36	188	188
Method 1 (Brute force)	53.735	57.84	11.75	193	772
Method 2 (using eq. 24)	53.726	57.82	11.74	193	579
Method 2 (using eq. 25)	53.662	58.00	11.68	196	579
Method 3 (Using KKT)	53.735	57.84	11.75	189	191

5. CONCLUSIONS

Previous work in this area addressed only problems with analytical objective functions in closed-form equations. Their approach was not readily applicable to problems involving objective functions that are expensive to evaluate. This paper proposed practical solutions to these problems by developing the following concepts:

1) **Use of design of experiments**: The statistical optimization technique developed in this research applies statistical design of experiments concepts in the computation of the expected value of a performance function. The developed technique is very effective when time-consuming computer models are used to measure performance for design problems with many variables. In Sundaresan et al. (1991 b), this method reduced the number of function evaluations by half compared to our previous technique by using an L8 orthogonal array to model variations in 4 design variables.
2) **Index of sensitivity**: The procedure defines a measure of sensitivity called Sensitivity Index(SI) which is the root mean square of the function values at the worst combinations of design variables. This concept is readily applicable to any problem that involves the measure of sensitivity. This concept models the effect of interactions between design variables, whereas concepts that use gradients to measure sensitivity do not.
3) **Constrained optimization**: This research proposed three techniques to solve constrained statistical optimiza-tion problems and defined various conditions of criticality of a constraint. The three statistical optimization techniques not only incorporate the effect of variations in design variables(ΔX_i), but also the uncertainty in the constraints(Δg_i) due to variations in design variable.

The developed statistical optimization methodology can be readily applied to other engineering design problems such as profile design for beverage can bottoms, etc. Potential applications also include design of rotational plastic parts and heat treated shaft elements.

ACKNOWLEDGMENT

The authors would like to acknowledge sponsors of this research: Ohio State University Gear Dynamics and Gear Noise Laboratory, National Science Foundation Division of Design and Manufacturing, Rockwell International, and Ohio State University Presidential Fellowship. Special Thanks to Mr. Carl Merhar for proof reading the manuscript.

REFERENCES

Adjunta, J. and Houser, D. R., 1992, "A study on the shrinkage characteristics and dimensional accuracy of cast gears," *Proceedings of the Power Transmissions and Gearing Conference* , Scottsdale, Arizona, ASME Pub. No. DE-Vol. 43-2, pp. 635-642.

AGMA *Standard for rating the Pitting Resistance and Bending Strength of Spur and Helical involute gear teeth* , American Gear Manufacturers Association Publication 218.01, Dec 1982.

Beltracchi, T. J. and Gabriele, G. A., 1988, "Observations on extrapolations using parameter sensitivity derivatives," *Proceedings of Design Automation Conference*, ASME Pub. No. DE-Vol. 14, pp. 165-173.

Beltracchi, T. J. and Gabriele, G. A., 1988, "A RQP based method for estimating parameter sensitivity derivatives," *Proceedings of Design Automation Conference*, ASME Pub. No. DE-Vol. 14, pp. 155-164.

Bhakuni N., 1991, "Structural Optimization Methods for Aluminum Beverage Can Bottoms," *Proceedings of ASME Advances in Design Automation Conference* , Publication No. DE-Vol. 32-1, pp. 265 - 271.

d'Entremont, K.L. and Ragsdell, K.M., 1988, "Design for latitude using TOPT," *Proceedings of ASME Advances in Design Automation*, DE-Vol.14, pp. 265-272.

Errichello, R., 1981, "An Efficient Algorithm for obtaining the gear strength geometry factor on a programmable calculator," American Gear Manufacturers Association Technical Paper P139.03.

Fiacco A.V., 1983, "Introduction to sensitivity and Stability analysis in non-linear programming," Academic Press, New York

Kliess, R. E., 1991, "The effect of thermal shrink and expansion on plastic gear geometry," *American Gear Manufacturers Association Fall Technical Meeting* , Detroit, Michigan.

Parkinson A, Sorensen C., Free J., and Canfield, B., 1990, "Tolerances and robustness in engineering design optimization," *Proceedings. of the 1990 ASME Design Automation Conference, Sept,. Chicago Vol 2, Pages 121-128.*

Phadke M.S., 1989, *Quality Engineering using robust design*, Prentice Hall, Englewood Cliffs, New Jersey.

Reklaitis, G.V., Ravindran, A. and Ragsdell, K.M., 1983, *Engineering Optimization methods and applications* , John Wiley and Sons.

Sobieski J., Barthelemy J., and Riley K. M., 1982, "Sensitivity of optimum solutions of problem parameters," *AIAA Journal* Vol 20, No.9, pp. 1291-99.

Sundaresan, S., Ishii, K., and Houser, D.R., 1991a, "A procedure using manufacturing variance to design gears with minimum transmission error," *ASME Journal of Mechanical Design* , Vol. 13, pp. 318-324.

Sundaresan, S., Ishii, K., and Houser, D.R., 1991b, "Design optimization for robustness using performance simulation programs," *Proceedings of ASME Advances in Design Automation Conference* , Publication No. DE-Vol. 32-1, pp. 249-256.

Taguchi, G., 1978, "Off-Line and On-Line quality control systems," *Proceedings of International Conference on Quality Control* , Tokyo, Japan.

Taguchi, G., 1987, *System of Experimental Design* , Edited by Don Clausing, American Supplier Institute, Dearborn, MI.

Vanderplaats, G.N., 1984, *Numerical optimization techniques for engineering design: with applications* , McGraw Hill, 1984.

Welbourn, D.B., 1979, "Fundamental knowledge of gear noise- A survey," *Proceedings of Noise and Vibration of Eng. and Trans. I. Mech. E.*, Cranfield, UK, July, pp. 9-14.